# Protein dynamics as studied by single molecule Förster resonance energy transfer

Inaugural dissertation

For the attainment of the title of doctor

in the Faculty of Mathematics and Natural Sciences

at the Heinrich Heine University Düsseldorf

presented by

## Dmytro Rodnin

From Kiew, Ukraine

Düsseldorf, July 2015

From the institute for Physical Chemistry II

at the Heinrich Heine University Düsseldorf

Published by permission of the

Faculty of Mathematics and Natural Sciences

at the Heinrich Heine University Düsseldorf

Supervisor: Prof. Dr. Claus A. M. Seidel

Co-supervisor: Prof. Dr. Holger Gohlke

Date of the oral examination: 24.08.2015

**Summary**

The structure and dynamics of two proteins, phage T4 lysozyme (T4L) and initiation factor 3 (IF3) from *E. coli*, were studied in this work utilizing advanced single molecule fluorescence spectroscopy methods. The two proteins are examples for different dynamic regimes which are induced by the structure of the linker between their domains.

**The transient state of the T4L**

The T4L consists of two subdomains which are connected by a long and structured alpha helix C. The enzyme is a glycoside hydrolase, which cleaves the saccharides out of the bacterial cell wall in order to damage it. Over 500 structures of different variants of T4L are deposited in the PDB databank, nearly all of them show the same folding motive with a small re-orientations of the subdomains, whereby the structure 148L (Kuroki et al, 1993) represents the most closed and the structure 172L (Zhang et al, 1995) represents the most open conformation. The dynamics of the protein is known to include a hinge bending motion, which is thought to represent the catalysis of the reaction, as well as various normal modes and fast motions (Arnold & Ornstein, 1992; de Groot et al, 1998; Kuroki et al, 1993).

In this work three conformations of T4L were identified using network of 24 fluorescently double labeled variants of T4L. Two conformational states, $C_1$ and $C_2$, were assigned to two distinct structures from the PDB database. Additionally, a previously unknown state $C_3$ of T4L was identified. This hidden compact state is thought to be involved in the product release, which is the first time Michaelis-Menten model for non-ATP/GTP driven product release by a hydrolyzing enzyme was shown experimentally. Furthermore, the complete kinetic scheme of the T4L was solved for the functional cycle of the protein. Two relaxation times $t_{R1} = 4$ μs, and $t_{R2} = 230$ μs were identified, whereby $t_{R1}$ corresponds to the previously described hinge bending motion between the conformation $C_1$ and $C_2$, whereas $t_{R2}$ represents a previously unknown motion between conformations $C_2$ and $C_3$ which presumably leads to the release of the product. This step might imply an evolutionary advantage for the enzymes which require fast catalysis.

**Intrinsic flexibility of the IF3**

Like T4L, *E. coli* IF3 also consists of two sub-domains, which are connected by a long and predominantly unstructured linker, and is much less studied than T4L. IF3 plays an important role during the initiation of protein synthesis at the ribosome by binding to the 30S ribosomal subunit and controlling the composition of the initiation complex (IC), which is composed of 30S subunit, mRNA, initiator tRNA (fMet-tRNA$^{fMet}$ in bacteria), and three initiation factors (IFs), IF1, IF2 & IF3 (Allen & Frank, 2007; Gualerzi et al, 2001; Laursen et al, 2005; Milon & Rodnina, 2012; Myasnikov et al, 2009). Though the structures of individual domains of IF3 are known and the domains are assumed to move largely independent due to highly flexible linker, the full impact on the structure and dynamics of the IF3 is not clear. Furthermore, the conformation and the dynamics of the IF3 on the 30S subunit in combination with other components of IC remain unknown.

In this work IF3 in solution was identified as highly flexible protein with at least eight distinct interconverting states and three relaxation rates ranging from nano- to milliseconds. At the extremes the distance between the FRET dyes positioned in the N-terminal domain and C-terminal domain range from 30 Å to 69 Å. The relaxation rates of the IF3 correspond to (1) the reversible domain association, (2) ordered-to-disordered transition of the linker region and (3) the chain dynamics of the unstructured linker. Upon the binding of IF3 to the 30S ribosomal subunit the majority of the IF3 molecules adopt one of the two alternative static conformations depending on the nature of the mRNA, the start codon selection and the strength of the Shine-Dalgarno interactions. Importantly, the flexibility of IF3 enables it to rapidly adopt diverse conformations on 30S subunit on a timescale which does not interfere with the translation.

**Unfolding and surfactant effects on the denaturation of T4L**

Correct folding of the proteins is the pre-requisite for their correct function and the inability to fold properly might be a cause for a debilitating or even deadly disease like Alzheimer, Huntington or Parkinson (Uversky, 2009). Therefore, the understanding of the folding and unfolding protein pathways is extremely important.

Using T4L as a model protein, its unfolding behavior was studied in unfolding conditions of acidic pH or urea. Structurally, urea unfolding of T4L proceeds through a fully extended state and concludes into a molten globule state, possibly cross-linked by the urea molecules (Sagle et al, 2009). The unfolding of T4L by protonation results in a polymer-like extended random-coil

structure. The T4L variant unfolded by urea showed at least one intermediate state, while for unfolding by protonation the intermediate states were indistinguishable from the unfolded states. Furthermore, their kinetic behavior were different as well, with protonation dynamics spanning much broader range from μs to ms, while the kinetics of the urea denaturation was in the range of sub-ms to ms. Therefore, the unfolding pathways and the mechanism of denaturation by urea and by protonation of T4L are significantly different.

For the unfolding experiments a non-ionic surfactant Tween 20 is often used to prevent adsorption artefacts in the single molecule experiments (Borgia et al, 2012; Choi & Chae, 2010; Hofmann et al, 2010; Pfeil et al, 2009). It was assumed, that non-ionic surfactant should not influence the folding/unfolding behavior, but no measurements were performed to show this.

In this work unfolding of T4L in urea and acidic pH was tested in presence and absence of Tween 20 and the resulting structural and kinetic data were compared. A strong influence of Tween 20 on the folding landscape of the T4L was detected. Even at the Tween 20 concentrations of $10^{-5}$ % w/v, which is lower than critical micelle concentration (CMC, $\sim 0.7*10^{-2}$ % w/v) of Tween 20, significant effects were observed on the T4L unfolding induced by urea. The surfactant and urea show cooperative behavior, whereby urea dissolves the structural elements of the protein and binds to the exposed protein backbone while Tween 20 stabilizes the protein in the unfolded state by binding to it. For the unfolding by protonation the effective Tween 20 concentrations is higher at $\sim 10^{-3}$ % w/v. For this type of unfolding, non-covalent and non-specific binding of Tween 20 to the hydrophobic areas of T4L exposed during unfolding leads to the stabilization of the protein in the preferred excited state. It is also feasible, that this non-covalent/non-specific binding sterically hinders the formation of compact states for T4L as for both denaturation types these states are strongly disfavored. It is clear, that Tween 20 is an undesirable addition to the unfolding experiments and alternative methods to prevent adsorption of proteins should be explored.

**Zusammenfassung**

Die Struktur und die Dynamik zweier Proteine, Phagen T4 lysozyme (T4L) und Initiation Faktor 3 (IF3) vom *E.coli*, wurden mit Hilfe von fortschrittlichen spektroskopischen Einzelmolekülmethoden untersucht. Diese zwei Proteine sind jeweils Beispiele für zwei dynamische Systeme, deren Verhalten von der Struktur des Linkers ihrer Domäne abhängig ist.

**Der Übergangszustand von T4L**

Protein T4L besteht aus zwei Unterdomänen, welche durch eine lange und strukturierte Alphahelix C verbunden sind. Dieser Enzym gehört zu der Klasse von Glucosidasen, welche Saccharide der bakteriellen Zellwand aufspaltet und damit die Zellwand zerstört. In der PDB Datenbank sind über 500 T4L Strukturen hinterlegt, die alle sehr ähnliches Faltmotiv besitzen, welches sich durch leichte Orientierungsänderung der Unterdomänen unterscheidet. Die Struktur 148L stellt die geschlossene Struktur dar, während die Struktur 172L die am weitesten geöffnete repräsentiert. Die Dynamik des Proteins besteht aus einer bekannten „hinge bending" Bewegung, welche der Katalyse entspricht, und mehreren schnellen Bewegungen und Normalmoden.

In dieser Arbeit wurden drei Konformationen von T4L mit Hilfe von 24 doppelt fluoreszent markierten Varianten von T4L identifiziert. Zwei dieser drei Konformationen, $C_1$ and $C_2$, konnten zu zwei bereits bekannten Strukturen aus der PDB Datenbank zugeordnet werden. Die dritte Konformation $C_3$ ist ein unbekannter kompakter Zustand, welcher vermutlich bei der Produktfreisetzung wichtig ist, wie für Michaelis-Menten Model für Proteine mit nicht-ATP/GTP gesteuerter Produktfreisetzung vorhergesagt wurde. Dies konnte in dieser Arbeit experimentell nachgewiesen werden. Außerdem, wurde das kinetische Netzwerk für den funktionalen Zyklus von T4L gelöst. Es wurden zwei Relaxationszeiten $t_{R1} = 4$ µs und $t_{R2} = 230$ µs identifiziert, wobei $t_{R1}$ der „hinge bending" Bewegung entspricht, während $t_{R2}$ dem vorher unbekannten Schritt der Produktfreisetzung entspricht. Dieser Schritt kann für einen Protein, der schnelle Katalyse braucht, einen evolutionären Vorteil darstellen.

**Intrinsische Flexibilität von IF3**

Wie T4L, besteht *E.coli* Initiation Faktor 3 aus zwei Unterdomänen. Diese sind mit einer langen aber unstrukturierten Aminosäurenkette verbunden. IF3 spielt wichtige Rolle bei der

Initiation der Proteinbiosynthese am Ribosom, wo es die Zusammenstellung des Initiationskomplexes (IK), welcher aus der ribosomalen Untereinheit 30S, mRNA, initiator tRNA (fMet-tRNA$^{fMet}$ in bakterien) und drei Initiationsfaktoren IF1, IF2 und IF3 besteht, überwacht. Die Struktur der einzelnen Untereinheiten der IF3 ist bekannt und es wird angenommen, dass diese Untereinheiten sich wegen dem langen, unstrukturierten Linker frei und voneinander unabhängig bewegen können. Tatsachlich wurde es aber noch nicht gezeigt. Ausserdem, wurden die Konformation und die Dynamik des IF3 auf dem Ribosom mit den anderen Komponenten des IK bis jetzt nicht untersucht.

In dieser Arbeit wurde gezeigt, dass freies IF3 in der Lösung ein sehr flexibles Protein mit mindestens acht einzigartigen Zuständen und drei Relaxationszeiten von Nano- bis Millisekunden ist. Der Abstand zwischen den Fluorophoren, welche an N-terminaler und C-terminaler Untereinheit gebunden sind, reicht von 30 Å bis 69 Å. Die drei Relaxationszeiten von IF3 entsprechend (1) der umkehrbarer Untereinheitenassoziation, (2) dem strukturiert-zu-strukturlose Übergang des Linkers, (3) der Proteinkettendynamik des strukturlosen Linkers. Bei der Bindung des IF3 zu der ribosomalen Untereinheit 30S nehmen die meisten IF3 Moleküle eine von der zwei möglichen statischen Konformationen an. Diese Wahl ist bedingt durch die Natur der mRNA, die Startkodon und die Stärke der Shine-Dalgarno Sequenz. Die Flexibilität des IF3 ermöglicht es dem Protein die Annahme unterschiedlicher Konformationen auf der Zeitskala, die die Translation nicht behindert.

## Die Entfaltung und die Effekte von Tensiden auf die Denaturierung von T4L

Korrekte Faltung der Proteine ist eine Voraussetzung für deren korrekte Funktion. Inkorrekte Faltung kann zu einigen schweren oder sogar tödlichen Krankheiten wie Alzheimer, Huntington oder Parkinson führen. Deswegen sind die Untersuchungen zu Faltung und Entfaltung von Proteinen sehr wichtig.

In dieser Arbeit wurde T4L als Modellprotein verwendet, um dessen Entfaltung in saurem pH oder Harnstoff zu untersuchen. Strukturell passiert die Entfaltung von T4L in Harnstoff einen komplett gestreckten Zustand und endet in einer „molten globule" Konformation, welche möglicherweise durch Harnstoffmoleküle quervernetzt ist. Entfaltung durch die Protonierung (saures pH-Wert) führt zu einer polymerähnlichen, gestreckten, Random-Coil Struktur. Die Harnstoff Entfaltung zeigt mindestens einen Übergangszustand, während für die Entfaltung durch Protonierung die Übergangszustände vermutlich mit den entfalteten Zuständen zusammenfallen. Das kinetische

Verhalten für diese zwei Entfaltungsmethoden ist ebenfalls verschieden, wobei die Relaxationszeiten der Dynamik für die Protonierung von µs bis ms gehen, während für die Harnstoff Entfaltung diese nur von sub-ms bis ms reichen. Insgesamt konnte gezeigt werden, dass diese Entfaltungsmethoden unterschiedlichem Verlauf folgen.

Bei den Entfaltungsexperimenten werden oft nicht-ionische Tenside wie z.B. Tween 20 eingesetzt, um die Adsorption der Biomoleküle bei Einzelmoleküluntersuchungen zu vermeiden. Es wurde angenommen, dass diese Tenside das Verhalten der Proteine bei den Entfaltungsexperimenten nicht stören.

In dieser Arbeit wurde die Entfaltung von T4L in Harnstoff und saurem pH mit oder ohne den Tensid Tween20 getestet. Es wurde ein starker Einfluss von Tween 20 auf die Entfaltung von T4L festgestellt. Für die Entfaltung mit Harnstoff wurde bereits bei der niedrigsten gemessenen Konzentration von $10^{-5}$ % g/v Tween 20 ein Unterschied zu der Messung ohne Tween 20 festgestellt. Das Tensid und der Harnstoff zeigen kooperatives Verhalten, wobei der Harnstoff die sekundäre Struktur auflöst und zum Proteinrückgrat bindet, während Tween 20 ebenfalls das Protein bindet und es in dem entfalteten Zustand stabilisiert. Bei der Entfaltung durch Protonierung sind Unterschiede bei der Tween 20 Konzentrationen höher als $10^{-3}$ % g/v sichtbar. In diesem Fall bindet Tween 20 unspezifisch und nicht-kovalent zu den freigelegten hydrophoben Gebieten des Proteins, was zu der Stabilisierung in bestimmten Zuständen führt. Diese nicht-kovalente Bindung stellt sterische Hinderung dar, die die Formation von kompakten Zuständen verhindert, was ebenfalls für Harnstoff Entfaltung gilt. Die Zugabe von Tween 20 sollte also vermieden werden und andere Möglichkeiten, die Adsorption zu bekämpfen, müssen verwendet werden.

Content

# 1. Introduction

## 1.1 Protein structure and dynamics

Proteins are building blocks of life. For the last decades the view of the proteins came from wonderful frozen pictures provided to us by high-resolution X-Ray crystallography, cementing the assumption of the proteins as static structures. Only in the last couple of years the structure-function paradigm of proteins has started to include the intrinsic dynamic nature of biomolecules (Henzler-Wildman & Kern, 2007; Shaw et al, 2010; Weiss, 2000). As proposed by Peter Tompa (Tompa, 2012) to understand the supertertiary structure one needs not only to detect all conformational states, but also the relative probabilities of the states and the energy barriers between them. The experimental and computational work up to now has been concentrated on determining basins i.e. states within the conformational energy landscape (Abrahams et al, 1994; Gardino et al, 2009; Nojiri & Saito, 1997).

Structure and function are intimately connected to the question of folding and unfolding. All proteins are synthetized unfolded inside the cell and undergo folding either unassisted (Dill & MacCallum, 2012) or with the help of chaperon (Frydman, 2001). For the proper function of a protein it is essential to have the correct structure. The failure to achieve it, can lead not just to a non-functional protein, but in some severe cases it can be a cause for debilitating or even life threatening disease like Alzheimer, Huntington or Parkinson (Uversky, 2009). Understanding folding and unfolding pathways is extremely important and can be described as determining a conformational energy landscape of a protein under denaturing conditions.

There are several challenges on the path of determining the complete energy landscape of a protein. First step is to detect all conformers for a protein. While the major states are often relatively easy to determine with the help of X-Ray crystallography, nuclear magnetic resonance (NMR) or cryo-electron microscopy (cryo-EM), the minor states are much more complicated to detect and are not accessible with X-Ray, or require huge amount of statistics for Cryo-EM. In recent years, several publications using NMR techniques provided an insight into possible ways to determine these minor populations (Mittermaier & Kay, 2006; Mulder et al, 2001; Skrynnikov et al, 2002). Another important technique which allows for the structure determination was presented by Kalinin and co-workers (Kalinin et al, 2012; Sindbert et al, 2011b). In recent publications they presented the structure of the RNA four-way junction acquired using single molecule multiparameter fluorescence detection (smMFD) in combination with high precision structural modelling. This approach relies on

an extensive labelling network which provides information on distance distributions, allowing identification of major and minor states and is sufficient to capture these states.

**A**
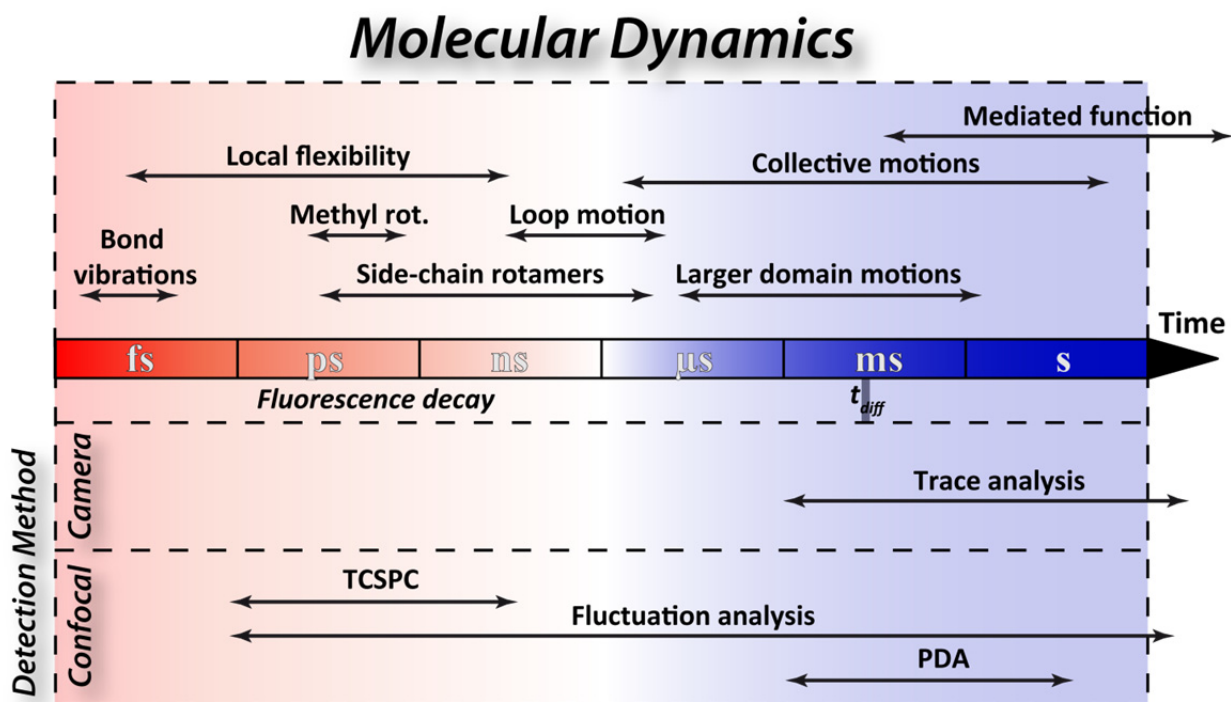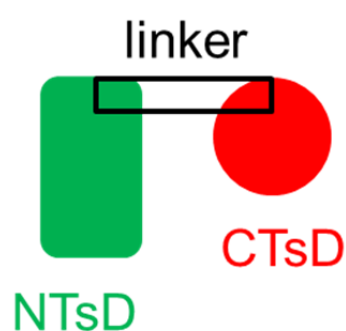


**B**      **C**
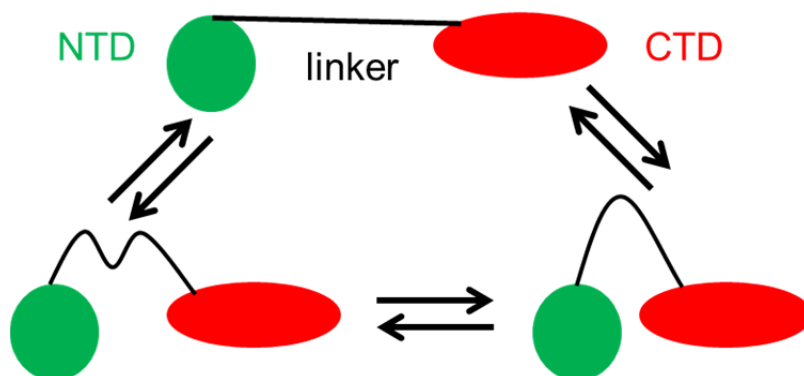


Figure 1.1. **Timescales accessible by a fluorescence experiment and schematic depiction of different models of two domain proteins**. (**A**) FRET can be used to study dynamic timerange covering over 10 decades in time. Figure extended from (Felekyan et al, 2013) (**B**) Schematic of a two domain protein with a structured linker corresponding to the T4 Lysozyme structure. (**C**) Schematics of possible conformations of a two domain protein with a flexible linker corresponding to the initiation factor 3.

The second challenge is the timescale of the dynamics, which covers a wide range, starting with the local motions of protein side chains on the fs to ns timescales and ending with the long range motions of slow domain movements on the timescale of seconds. Many techniques (e.g. NMR relaxation experiments, X-ray diffraction, fluorescence and molecular simulations) cover only some parts of the timescale. Copious analysis of the fluorescence data using all available tools allows for detection of the dynamics over 10 decades in time, making fluorescence a prime candidate to solve this challenge (Fig. 1.1a). In the context of supertertiary structure the object of the research is multidomain proteins with two-domain proteins representing the first step. Within this subgroup one prominent difference is the nature of the linker bridging the domains – it can be long or short, structured or disordered. To show the linker dependent relationship between structure, dynamics and function and its influence on the supertertiary energetic landscape two proteins were investigated: T4 lysozyme (T4L) with a structured α-helical linker (Fig. 1.1b) and initiation factor 3 (IF3) with a flexible linker (Fig. 1.1c).

## 1.2 Phage T4 Lysozyme

The bacteriophage T4 Lysozyme (T4L) contains 164 residues and consists of two subdomains: N-terminal (NTsD) (residues 13–75) and C-terminal (CTsD) (residues 76–164 and 1–12) connected by a long alpha helix C (Fig. 1.2a). The enzyme belongs to the group of glycoside hydrolases and is responsible for cleaving the glycosidic bond between N-acetylmuramic acid and N-acetylglucosamine of bacterial cell wall saccharides resulting in damages to the bacterial cell wall. There is a significant (>500 PDBs) database of structures of T4L, most of those coming from crystals (Baase et al, 2010; Dixon et al, 1992; Goto et al, 2001; Kuroki et al, 1993; McHaourab et al, 1997; Zhang et al, 1995). Remarkably, all these structures show nearly the same folding with a slight difference only in the orientation of the domains, where structure PDB 148L can be described as most closed conformation (Kuroki et al, 1993) while PDB 172L can be described as most open (Zhang et al, 1995). The agreed functional mechanism starts with lysozyme binding to the fourth sugar of the hexasaccharide (the D ring) and distorting it into a half-chair conformation, which allows the glycosidic bond to be easily broken. The catalysis of this reaction includes a hinge-bending motion between the open and closed conformation (Fig. 1.2b).

In addition to this "simple" bending motion various normal modes and fast motions in T4L were suggested by extensive computational simulations (Arnold & Ornstein, 1992; de Groot et al, 1998). Nevertheless, the temporal domain of many of these motions are still unknown mainly due to two reasons: i) States are low populated and ii) the exchange rate of these states lies in a temporal

domain not easily accessible to conventional tools (microsecond to millisecond). Therefore, T4L - a dynamic molecule - is well suited for studying conformational transitions with sub-microsecond resolution.

**A**                                                    **B**
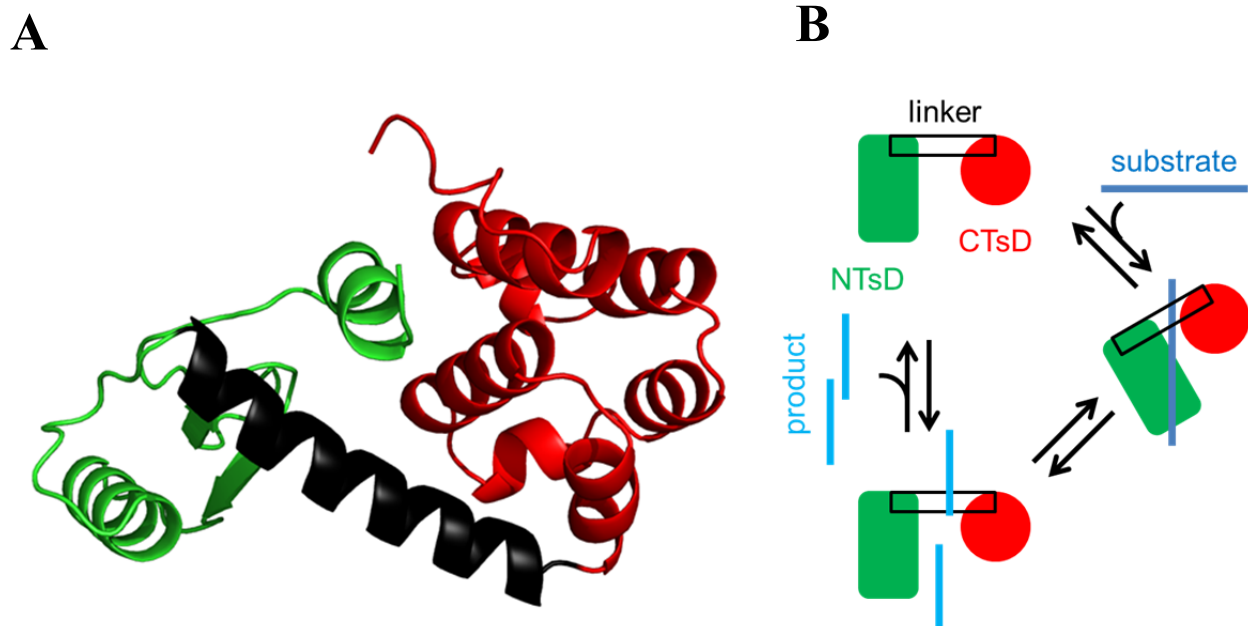


Figure 1.2. **Structure of T4 Lysozyme and the schematic of its enzymatic activity**. (**A**) Structure of T4L is derived from PDB 172L. The N-terminal subdomain is shown in green, C-terminal subdomain in red, connecting helix C in black. (**B**) Upon binding the substrate peptidoglycan, two subdomains move closer together representing a hinge bending motion. After the catalysis whereby the substrate is cleaved, the protein releases the product and relaxes into the open conformation.

## 1.3 Initiation Factor 3

IF3 (Fig. 1.3a) plays an important role during initiation of protein synthesis at the ribosome. 30S subunit, mRNA, initiator tRNA (fMet-tRNA$^{fMet}$ in bacteria), and three initiation factors (IFs), IF1, IF2 & IF3 constitute the initiation complex (IC)(Fig. 1.3b) and the interplay between them controls the efficiency of mRNA recruitment for translation, thereby determining the composition of the cellular proteome and controlling the response and adaptation to environmental stimuli (Allen & Frank, 2007; Gualerzi et al, 2001; Laursen et al, 2005; Milon & Rodnina, 2012; Myasnikov et al, 2009).

*E. coli* IF3 is contains 180 amino acids and consists of two domains: the N-terminal domain (NTD) (residues 1−77), of which the function is not known, and the C-terminal domain (CTD) (residues 91−180), which appears to carry out most of the IF3 activities in translation (Kycia et al, 1995; Moreau et al, 1997; Petrelli et al, 2001) and binds to the shoulder of the 30S subunit (Dallas &

Noller, 2001; Fabbretti et al, 2007; Julian et al, 2011; McCutcheon et al, 1999). The structure of the individual domains is known (Biou et al, 1995; Garcia et al, 1995a; Garcia et al, 1995b); however the structure of the full-length IF3 is not. The domains are connected by a linker (residues 79-90), which allows for different arrangements of the NTD and CTD with distance ranging from 45Å to 60Å between the centers of mass of the two domains (Moreau et al, 1997). The length of the linker may vary between just four residues (such as in IF3 from *Geobacillus stearothermophilus* (Kycia et al, 1995)) to 12 amino acids as found in *E. coli* and chloroplast IF3 (Hua & Raleigh, 1998; Moreau et al, 1997; Yu & Spremulli, 1997). The change in the linker length is caused by remodeling of the C-terminal α-helical segment of the NTD. NMR studies suggested that in *E. coli* IF3 the two domains move largely independently and the inter-domain linker is highly flexible (Moreau et al, 1997). Still the full impact of the linker on the supertertiary structure is not clear.
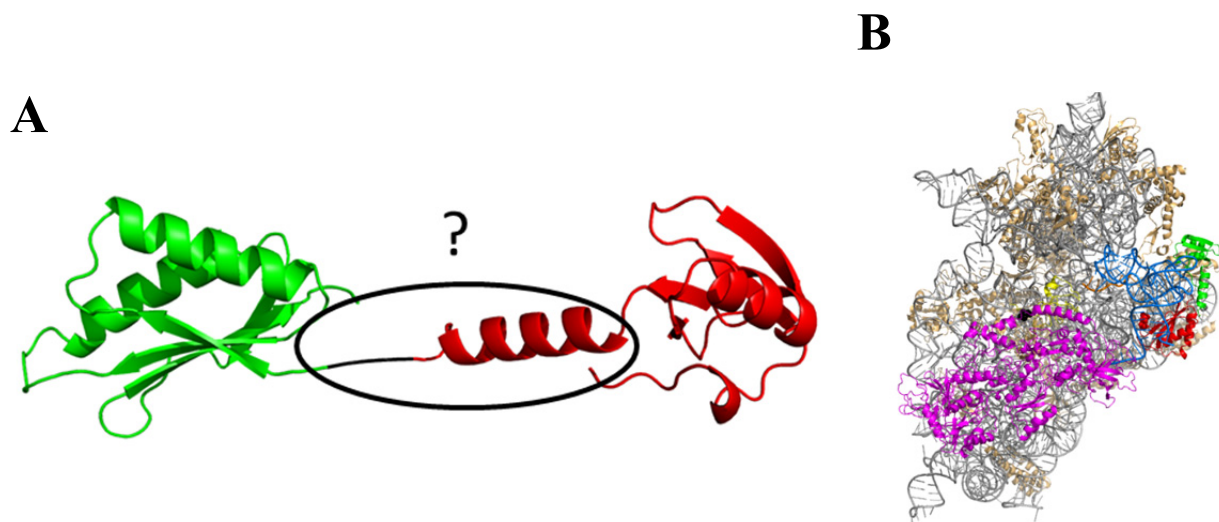


Figure 1.3. **Putative structure of IF3 and initiation complex (IC)**. (**A**) Model of IF3 based on the structure of individual domains (Homology model from 1TIF for N-terminal domain and 2IFE for C-terminal domain)(Biou et al, 1995; Moreau et al, 1997). N-terminal domain in green and C-terminal domain in red, the aminoacids missing in the linker region (black) for which no structural information was found in database were modeled using PyMol. The region indicated by black circle can adopt different conformations, e.g. α-helical or disordered. (**B**) Model of pre-initiation complex reconstructed by Cryo-EM (Julian et al, 2011; McCutcheon et al, 1999). N-terminal domain of IF3 is depicted in green, C-terminal domain of IF3 in red, IF1 in yellow, IF2 in magenta, mRNA in blue, fMet-tRNA^fMet in orange, 16S RNA in grey and 30S proteins in gold.

Furthermore, the position and conformation of the IF3 on the 30S initiation complex remains unclear, despite cryo-EM reconstruction of the 30S IC which suggested that on the ribosome IF3 is bound in an open, extended conformation (Julian et al, 2011; McCutcheon et al, 1999). Additionally, recent single-molecule FRET studyies in a TIRF setup suggested that 30S IC–bound IF3 can assume several conformations (Elvekrog & Gonzalez, 2013) depending on the presence of initiator tRNA and proper anticodon-codon interaction with the start codon within a completely assembled 30S IC. Nevertheless, the link between the dynamics of IF3 in solution and the conformations of the factor on the 30S subunit during initiation is unclear.

# 2. FRET Theory

## 2.1 Fluorescence

Förster Resonance Energy Transfer (FRET) is a strongly distance dependent non-radiative energy transfer. It occurs between two molecules called donor (D) and acceptor (A). Once the donor is in an excited electronic state it can follow several pathways to return to the ground state (Fig. 2.1a). One of those is the fluorescence, which involves the transfer of energy from donor to acceptor. The rate of this process is called $k_{FRET}$ and is defined as:

$$k_{FRET} = \frac{R_0^6}{\tau_{D(0)} R_{DA}^6}$$
(Eq. 2.1)

where $\tau_{D(0)}$ is the donor lifetime in absence of FRET, $R_{DA}$ is the distance between the donor and acceptor and $R_0$ is the Förster radius where the transfer efficiency is 50% and is calculated as

$$R_0 = \left[ \frac{9(\ln 10)}{128\pi^5 \cdot N_A} \cdot \frac{J(\lambda) \cdot \kappa^2 \cdot \Phi_{FD(0)}}{n^4} \right]^{\frac{1}{6}}$$
(Eq. 2.2)

where $J(\lambda)$ is spectral overlap integral between fluorescence of D and absorbance of A (Figure 2.1b), $\kappa^2$ is dye orientation factor (Figure 2.1d), $\Phi_{FD(0)}$ is fluorescence quantum yield of the donor in absence of acceptor, $N_A$ is the Avogadro constant and $n$ is the refractive index of the medium between the molecules.

FRET can be also quantified by FRET efficiency $E$. It is defined as the number of quanta transferred from D to A divided by total number of quanta absorbed by D.

$$E = \frac{n_A}{n_D + n_A}$$
(Eq. 2.3)

where $n_D$ and $n_A$ are the number of photons emitted by the donor and acceptor. After going through a series of transformations one finally arrives at the commonly used expression:

$$E = \frac{1}{1 + \left( \frac{R_{DA}}{R_0} \right)^6}$$
(Eq. 2.4)

With $R_0$ values of around 50 Å for the most common donor acceptor pairs, the FRET has the detection range between 20 Å and 100 Å. Fortunately, this is the scale fitting perfectly to the size of the proteins, with only the biggest ones exceeding (f.e. ribosomes) this limit. However, even for those due to variability of the possible dye attachment points, it is possible to design a donor-acceptor pair, which can be measured by FRET. In this work green fluorescent dye Alexa 488 was used as a donor and red fluorescent dye Alexa 647 was used as acceptor (see also section 3.5).
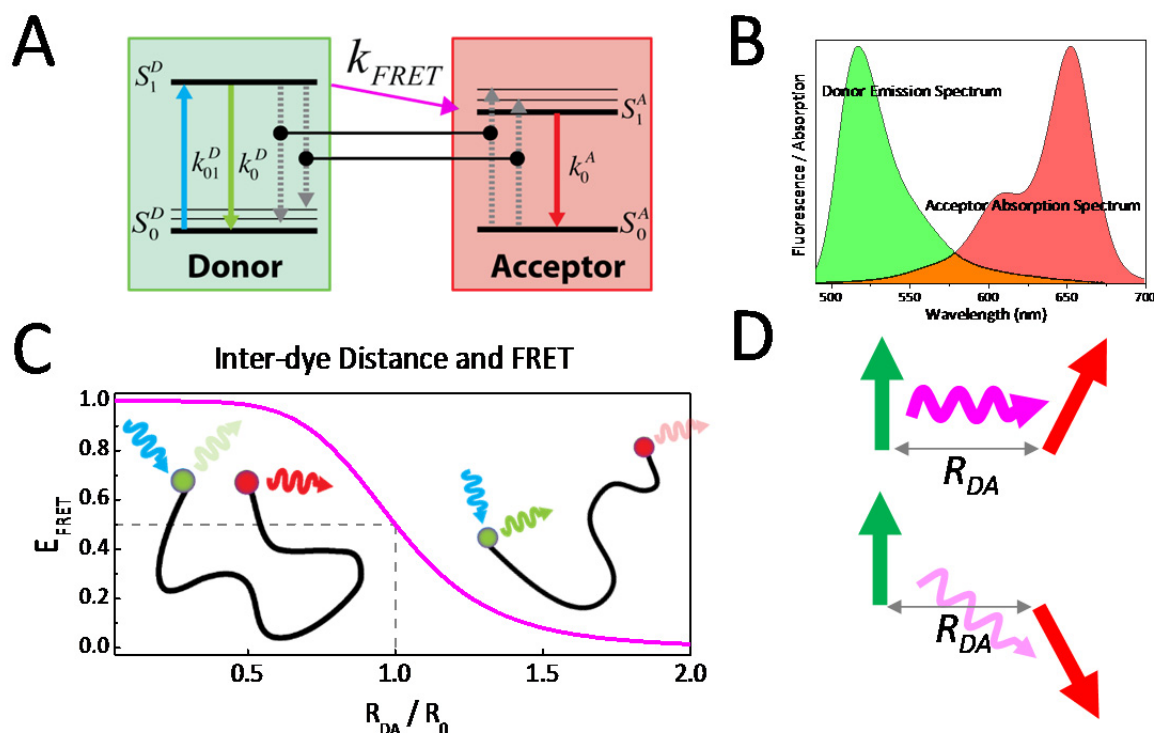


Figure 2.1. **Principles of FRET (adapted from Sanabria 2015, in preparation)**. Principles of FRET: (**A**) Simplified Perrin-Jablonski diagrams of D and A. D is excited at a rate $k_{01}{}^D$ to the first singlet state $S_1{}^D$ In the absence of A, it is depopulated with rate constant $k_{01}{}^D$. In presence of A the possible de-excitation of D and excitation of A can occur and energy transfer happens at a rate $k_{FRET}$ resulting in the excitation of A from $S_0{}^A$ to $S_1{}^A$ which is depopulated with a rate constant $k_0{}^A$. (**B**) The emission (Fl.) spectra of Alexa488 (A488) (Green) and the excitation (Abs.) of Alexa594 (A594) (Red). A488–A594 dyes constitute a commonly used D−A pair in FRET studies. The amount of the overlap between the emission of D and excitation of A (orange region) influences the value of the Förster radius $R_0$. (**C**) FRET efficiency ($E_{FRET}$) versus the normalized interdye distance ($R_{DA}/R_0$). The value of the Förster radius, $R_0$, defines the useful dynamic range of distances ($0.5 \leq R_{DA}/R_0 \leq 1.5$; $0.98 \geq E_{FRET} \geq 0.08$, normally 20Å < DA < 100Å) that can be measured with a specific dye pair. (**D**) $R_0$ strongly depends on the mutual orientation of the dipoles known as the $\kappa^2$. However, dynamic averaging ensures $\kappa^2 \approx 2/3$, which is the typically used.

## 2.2 Fluorescence anisotropy

Fluorophores absorb preferably the light which is polarised parallel to their absorption dipole moment and emit it with the polarisation parallel to their emission dipole moment. This leads to signal depolarization. The measure of this process is fluorescence anisotropy $r$ and is defined as:

$$r = \frac{r_0}{1 + \frac{\tau}{\rho}}$$

(Eq. 2.5)

where $r_0$ is the fundamental anisotropy, $\tau$ is the fluorescence lifetime and $\rho$ the rotational correlation time. This equation is called Perrin equation. $r_0$ in itself is defined only by the angle $\beta$ between the absorption dipole and emission dipole, called transition dipole moment.

$$r_0 = \frac{3\cos^2(\beta) - 1}{5}$$

(Eq. 2.6)

Rotational correlation time $\rho$ is defined by

$$\rho = \frac{\eta \cdot V}{R \cdot T}$$

(Eq. 2.7)

where $\eta$ is the viscosity of the solution, $V$ is the volume of the molecule, which is often approximated to be a sphere, $T$ is temperature in Kelvin and $R$ is the ideal gas constant.

For practical measurements, fluorescent signals $F_{\parallel}$ and $F_{\perp}$ are measured with $F_{\parallel}$ being the emission parallel to the excitation source and $F_{\perp}$ the emission perpendicular to the excitation source. Anisotropy is then defined as

$$r = \frac{G \cdot F_{\parallel} - F_{\perp}}{G \cdot F_{\parallel} + 2 \cdot F_{\perp}}$$

(Eq. 2.8)

where $G$ is a factor which corrects for different detection sensitivities for parallel and perpendicular polarized light.

# 3. Materials and Methods

## 3.1 Technical equipment

| Equipment | Modell | Company |
|-----------|--------|---------|
| Autoclave | Systec VE-150 | Systec |
| Cell density spectrometer | CO8000 Cell Density Meter | WPA Biowave |
| Centrifuge | Sorvall EvolutionRC | Thermo Scientific |
| Centrifuge | Megafuge 1.0 R | Heraeus |
| Chambered coverglass | Nunc Lab-Tek II | Thermo Scientific |
| Concentrators | Amicon® MW-15 10k/30k | EMD Millipore |
| Dialysis hose | Spectra/Por 4 | Spectrum Labs |
| Distillated water system | Arium 611 | Sartorius |
| Electrophoresis | Mini-PROTEAN II | Biorad |
| Fluorescence spectrometer | FluoroLog®-3 | Jobin Yvon Inc. |
| FPLC | ÄKTAprime™ plus | GE Healthcare |
| FPLC affinity column | HiTrap Chelating HP 1ml | GE Healthcare |
| FPLC cation exchanger monoS 5/5 | HiTrap SP FF 5ml | GE Healthcare |
| Highprecision scales | CP224S | Sartorius® |
| Incubators | Minitron | Infors HT |
| Membrane filters | 2μm | Nalgene |
| Peristaltic pump | P1 | GE Healthcare |
| pH-Meter | pH-Meter 766 Calimatic | Knick |
| Rotator | Stuart® SB3 | BioCote |
| Rotor | Sorvall SLA-3000 | Thermo Scientific |
| Rotor | Sorvall SL-34 | Thermo Scientific |
| Shaker | Orbitalschüttler 4010 | Köttermann |
| Sonificator and sonotrode | Sonopuls HD 2200, MS 72 | Bandelin |
| Syringe filters | 2μm | Whatman |
| Table centrifuge | Biofuge fresco | Heraeus |

| | | |
|---|---|---|
| TCSPC | 5000U | Jobin Yvon Inc. |
| UV-Vis Spectrometer | Cary 4000 UV-VIS | Agilent Technologies Inc. |
| UV-Vis Spectrometer | NanoDrop ND-1000 | peqlab |

## 3.2 Chemicals

All chemicals are of the *pro analysis* grade, except in cases where it is specifically specified.

| Chemical | Company | Ordering Number |
|---|---|---|
| 2-Mercaptoethanol 99 % p.a. | Carl Roth | 63689 |
| Acrylamid | Bio-Rad | 1610147 |
| Agar | Becton, Dickinson & Company | 221275 |
| Ampicillin | Carl Roth | 11257 |
| APS | Sigma-Aldrich | 248614 |
| Charcoal | Merck | |
| Chloramphenicol | Carl Roth | 3886.1-3 |
| Dimethyl sulfoxide (DMSO) | Merck | |
| EDTA Na$_2$ 2H$_2$O | AppliChem | 131669 |
| Ethanol p.a. | VWR | 83811 |
| Glycerol | Carl Roth | 6962.1-4 |
| Guanidinium Chloride | Carl Roth | G4505-25G |
| Hydrochloric acid 37% (HCl) | VWR | 1.13386.2500 |
| Imidazole | AppliChem | A3635 |
| Isopropyl β-D-1-thiogalactopyranoside (IPTG) | Carl Roth | 2316.1-5 |
| Liquid nitrogen | Linde | |
| Magnesium chloride | Sigma-Aldrich | M2670 |
| Peptidoglycan from *Micrococcus luteus* | Sigma Aldrich | 53243 |
| Potassium chloride | Sigma-Aldrich | 746436 |
| Sodium chloride p.a. | VWR | 85139 |
| Sodium hydroxide (NaOH, Pellets) | J.T.Baker | |

| | | |
|---|---|---|
| Sodium laurylsulfate pellets $\geq$ 99 % (SDS) | Carl Roth GmbH | 20765 |
| Tetramethylethylenediamine (TEMED) | Sigma-Aldrich | T7024 |
| Tris(hydroxymethyl)aminomethane (Tris) | VWR | 25,285-9 |
| Tris/Glycine/SDS premixed 10x | BIO-RAD Lab. Inc. | 161-0732 |
| Trolox | Sigma-Aldrich | 56510 |
| Tryptone | Becton, Dickinson & Company | 211638 |
| Tween 20 (10 ml) | Sigma-Aldrich | P7949 |
| Unnatural amino acid para-Acetylphenylalanine (pAcF) | SynChem OHG | |
| Urea | Appli Chem | 131754 |
| Yeast extract | Becton, Dickinson & Company | |

## 3.3 Bacterial strains

| Strain | Genotype | Company |
|---|---|---|
| BL21 | F - ompT hsdSB(rB − mB − ) dcm + dam + Tet R galλ (DE3) endA Hte [argU, ileY, leuW, Cm R ] | Stratagene |

## 3.4 Dyes

Following dyes were used in this work:

| **Donordye: Alexa Fluor® 488 C$_5$ Maleimid (Invitrogen, Inc)** | |
|---|---:|
| Excitation maximum [nm] | 494 |
| Emission maximum [nm] | 519 |
| Extinction coefficient in Absorption maximum ε [cm$^{-1}$M$^{-1}$] | 73000 |
| Fluorescent quantum yield ɸ | 0.92 |
| Fluorescent lifetime [ns] | 4.1 |
| Correction factor for λ = 280 nm | 0.11 |
| Net charge | -1 |
| Linker length (Count of C-Atoms/Overall atom count) | C5/A10 |
| Invitrogen ordering number | A-10254 |
| AV model length | 20 Å |
| AV model width | 4.5 Å |
| AV model dye radius | 3.5 Å |
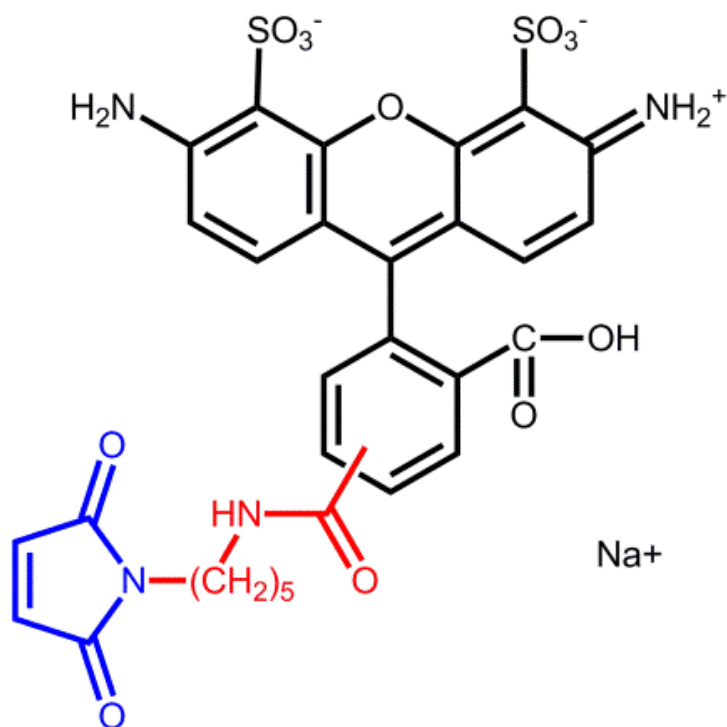
**Akzeptordye: Alexa Fluor® 647 C$_2$ Maleimid**

**(Invitrogen, Inc)**

| | |
|---|---|
| Excitation maximum [nm] | 651 |
| Emission maximum [nm] | 672 |
| Extinction coefficient in Absorption maximum $\varepsilon$ [cm$^{-1}$M$^{-1}$] | 250000 |
| Fluorescent quantum yield $\phi$ | 0.33 |
| Fluorescent lifetime [ns] | 1 |
| Correction factor for $\lambda = 280$ nm | 0.03 |
| Net charge | -3 |
| Linker length (Count of C-Atoms/Overall atom count) | C2/A12 |
| Invitrogen ordering number | A-20347 |
| AV model length | 22 Å |
| AV model width | 4.5 Å |
| AV model dye radius | 3.5 Å |
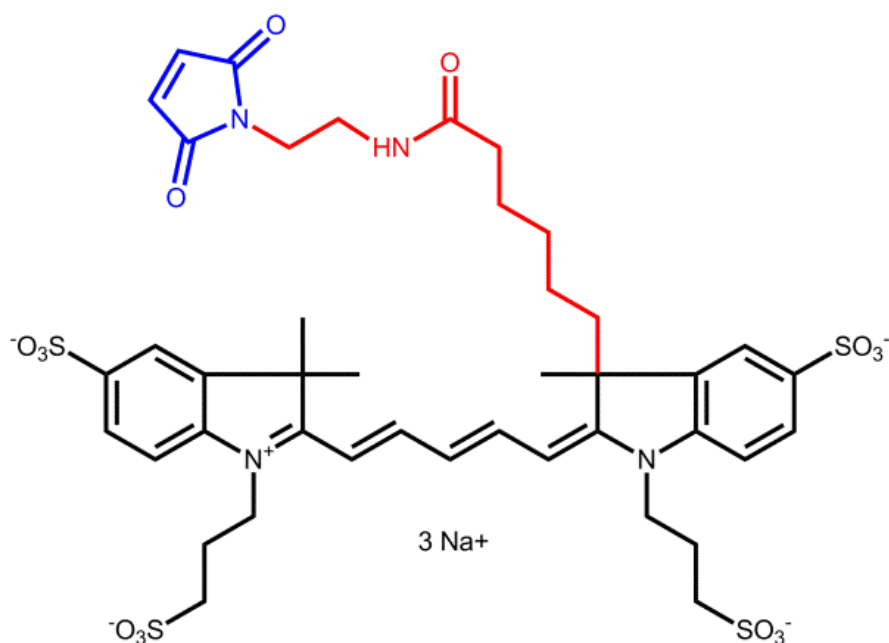
## 3.5 Buffers

### 3.5.1 Buffers for T4 Lysozyme project

LB (Luria Bertani) Medium:

10 g/l tryptone

10 g/l NaCl

5 g/l yeast extract

1 pellet NaOH

LB Agar:

1% Agarose solved in LB medium

Lysis buffer:

50 mM HEPES

1 mM EDTA

5 mM DTT

pH 7.5 at RT

Thiol Labeling buffer:

50 mM Sodium Phosphate Buffer

150 mM NaCl

pH 7.5 at RT

Keto Labeling Buffer

50 mM Sodium Acetate

150 mM NaCl

pH 4 at RT

Measuring buffer:

50 mM Phosphate Buffer

150 mM NaCl

pH 7.5 at RT

## 3.5.2 Buffers for IF3 Project

Measuring Buffer:

50 mM Tris-HCl

70 mM $NH_4Cl$

30 mM KCl

7 mM $MgCl_2$

pH 7.6 at RT


30S activation buffer:

50 mM Tris-HCl

70 mM $NH_4Cl$

30 mM KCl

100 mM $MgCl_2$

pH 7.6 at RT

## 3.6 Biochemical Methods

### 3.6.1 Methods for the T4 Lysozyme Project

#### 3.6.1.1 T4 Lysozyme purification

T4L cysteine and amber (TAG) mutants were generated by Katharina Hemmen or Hugo Sanabria via site directed mutagenesis as previously described in the pseudo-wild-type containing the mutations C54T and C97A (WT*) and subsequently cloned into the pET11a vector(Brustad et al, 2008; Fleissner et al, 2009; Lemke, 2011).

The plasmid containing the gene with the desired mutant was co-transformed with pEVOL(Lemke, 2011) into BL21(DE3) *E. coli* and plated onto LB- agar plates supplemented with the respective antibiotics, ampicillin and chloramphenicol. A single colony was inoculated into 100 mL of LB medium containing the above mentioned antibiotics and grown overnight at 37 °C in a shaking incubator. 50 mL of the overnight culture were used to inoculate 1 L of LB medium supplemented with the respective antibiotics and 0.4 g/L of pAcF (SynChem) and grown at 37°C until an $OD_{600}$ of 0.5 was reached. The protein production was induced for 6 hours by addition of 1 mM IPTG and 4 g/L of arabinose.

The cells were harvested, lysed in lysis buffer and purified using a monoS 5/5 column (GE Healthcare) with an eluting gradient from 0 to 1 M NaCl according to standard procedures. High-molecular weight impurities were removed by passing the eluted protein through a 30 kDa Amicon concentrator (Millipore), followed by subsequent concentration and buffer exchange to the measuring buffer with a 10 kDa Amicon concentrator.

#### 3.6.1.2 High Performance Liquid Chromatography

Binding of labeled T4L mutants to peptidoglycan from *Micrococcus luteus* (Sigma-Aldrich) was monitored by reverse phase chromatography using a C-18 column out of ODS-A material (4 X 150 mm, 300 Å) (YMC Europe, GmbH). The protein was eluted with a gradient from 0 to 80% acetonitrile containing 0.01% trifluoroacetic acid for 25 min at a flow rate of 0.5 ml/min. The labelled complex elution was monitored by absorbance at 495 nm.

## 3.6.1.3 Fluorescence labelling

Site specific labelling of T4L uses orthogonal chemistry. For labelling the Keto handle at the N-terminus the hydroxylamine linker chemistry was used for Alexa 488 and Alexa 647 (Fig. 3.1). Cysteine mutants were labeled via a thiol reaction with maleimide linkers of the same dyes. Double mutants were labeled sequentially - first thiol and second the keto handle, as suggested by Brustad et al. (Brustad et al, 2008). Single mutants were labeled in one step reaction. The thiol reaction was carried out overnight at room temperature in labelling buffer in presence of 5 molar excess of dye. The keto reaction was done in keto labelling buffer with 5 molar excess of dye for over 12 hours. After each reaction, excess of unreacted dye was removed via a desalting column PD-10 and further concentrated using Amicon 10kDa concentrators. For labeling the Keto function of the *p*-acetyl-L-phenylalanine (pAcF) amino acid at the N-terminus, hydroxylamine linker chemistry was used for Alexa 488 and Alexa 647. To become aware of specific fluorophore effects in FRET measurements, labeled samples were prepared in both possible configurations (named "(DA)" when the donor (acceptor) is attached to the NTsD (CTsD), and "(AD)" for the reverse order).



Figure 3.1 **Site specific labelling with unnatural amino acid (adapted from T4L manuscript). (A, B)** Orthogonal chemistry for thiol and keto labelling, respectively
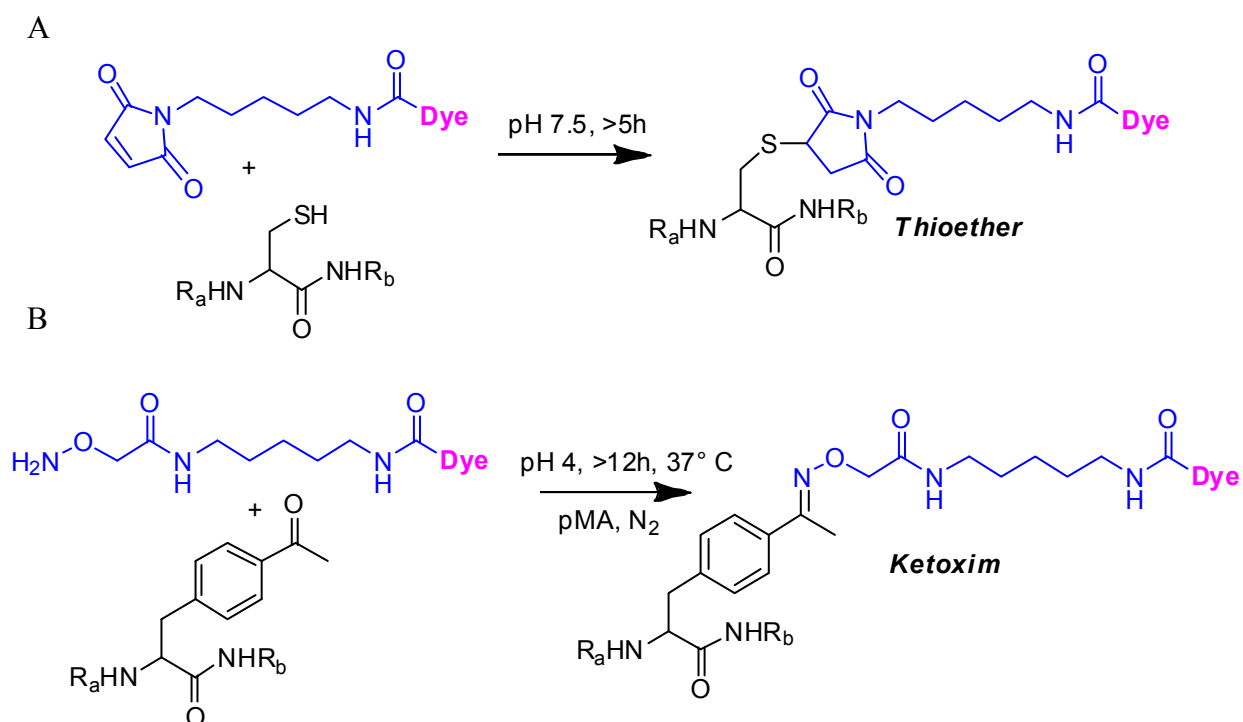
## 3.6.2 Methods for the IF3 Project

### 3.6.2.1 Complex preparation for single molecule measurements

The complexes were prepared as described in (Milon et al, 2007). 30S subunits were reactivated in 30s activation buffer at final 20 mM $MgCl_2$ concentration for 1 hr at 37°C. The final complex was assembled without any IF3, at 10 μM 30S concentrations. All other components were added in 1.5x concentration over 30S subunits. Final complex was incubated for 15' at 4°C and centrifuged afterwards shortly at 13000 rpm in table centrifuge. The complex was diluted to the appropriate concentration in the well of the measuring chamber, double labeled IF3 was added at a single molecule concentration and measured for at least an hour at single molecule setup. For FCS experiment the concentration of double labeled IF3 was increased to the FCS concentration (aprox. 3 molecule in focus, adjusted upon every measurement).

## 3.7 Computational methods (Part 1)

### 3.7.1 Accessible volume modeling (AV modeling)

For proteins with known structures it is possible to access best labelling positions using either AV-modelling or molecular dynamics simulations. As shown by Wozniak et al (Sindbert et al, 2011a; Woźniak et al, 2008), the quality of AV modelling is comparable with MD simulations, while providing superior speed for the simulation. The AV considers the dyes as hard sphere models connected to the protein via flexible linkers (modeled as a flexible cylindrical pipe) (Sindbert et al, 2011a; Woźniak et al, 2008) with all positions having equal probability. The overall dimension (width and length) of the linker is based on their chemical structures. Parameters used in this work are shown in the section 3.4.

For IF3 the positions were pre-ordained by our collaborators at MPI. Nevertheless we performed AV modeling on a putative structure of IF3 to access the expected FRET distance and guard against possible artefacts of the labelling positions (Fig. 3.2).

For T4L we generated a series of AV's for donor and acceptor dyes attached to T4L placing the dyes at multiple separation distances. For each pair of AV's, we calculated the distance between dye mean positions ($R_{mp}$)

$$R_{mp} = \left| \left\langle \vec{R}_{D(i)} \right\rangle - \left\langle \vec{R}_{A(j)} \right\rangle \right| = \left| \frac{1}{n} \sum_{i=1}^{n} \vec{R}_{D(i)} - \frac{1}{m} \sum_{j=1}^{m} \vec{R}_{A(j)} \right|,$$  (Eq. 3.1)

where $\vec{R}_{D(i)}$ and $\vec{R}_{A(i)}$ are all the possible positions that the donor fluorophore and the acceptor fluorophore can take. However, in ensemble TCSPC (eTCSPC) the mean donor-acceptor distance is observed:

$$\left\langle R_{DA} \right\rangle = \left| \left\langle \vec{R}_{D(i)} - \vec{R}_{A(j)} \right\rangle \right| = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \left| \vec{R}_{D(i)} - \vec{R}_{A(j)} \right|,$$  (Eq. 3.2)

which can be modeled with the accessible volume calculation.

The relationship between $R_{\mathrm{mp}}$ and $\langle R_{\mathrm{DA}} \rangle$ can be derived empirically following a third order polynomial from many different simulations.
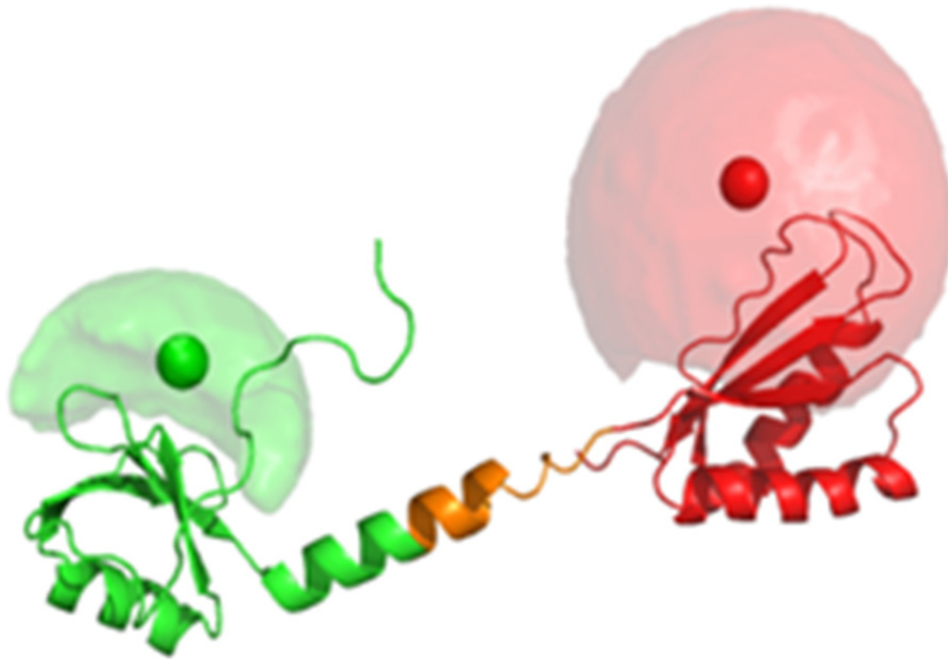


Figure 3.2. **Putative IF3 structures with AV-clouds for Alexa 488 and Alexa 647 (taken from IF3 manuscript)**. N-terminal domain is shown in green and C-terminal domain in red. The amino acids in the linker region (orange) for which no reliable structure is available were modeled for visualization using PyMol. Accessible volumes (AV) for the dye Alexa 488 (pale green surface) and Alexa 647 (pale red surface) were simulated (Sindbert et al, 2011a), with mean dye positions depicted as green and red spheres, respectively.

## 3.8 Spectroscopic Methods

### 3.8.1 Absorption spectra

Absorption spectra were recorded using UV-Vis spectrophotometer Cary 300-Bio from Varian. The absorption of light is defined by the Lambert-Beer equation

$$Abs = -\log \frac{I_{Trans}}{I_0} = \varepsilon \, c \, d \qquad \text{(Eq. 3.3)}$$

where $I_0$ is the intensity of the incident light, $I_{trans}$ is the intensity of the transmitted light, $\varepsilon$ is the extinction coefficient, $c$ is the concentration of the sample and d is the optical length of the cell. The data collection was performed in double beam mode, using a reference cell. The recording was simultaneous, allowing for correction of the measurement results in real time.

### 3.8.2 Fluorescent spectra

Fluorescent spectra were recorded with Fluorolog-3 and Fluoromax-3 fluorometers from Horiba Jobin Yvon, SPEX. The fluorescene intensity F is expressed as

$$F = \Phi_F \, I_{Abs} \qquad \text{(Eq. 3.4)}$$

where $I_{Abs}$ is the intensity of the absorbed light and $\Phi_F$ is the quantum yield of the dye, defined as

$$\Phi_F = \frac{Emitted \; photons}{Absorbed \; photons} \qquad \text{(Eq. 3.5)}$$

Assuming $I_{Abs} = I_0 - I_{trans}$, we can express Eq. 3.4 as following:

$$F = \Phi_F \, I_0 \left(1 - 10^{-\varepsilon c d}\right) \qquad \text{(Eq. 3.6)}$$

## 3.8.3 Single molecule multiparameter fluorescence detection (sm MFD)

Various dimensions (fluorescence lifetime, anisotropy, spectral properties) can be measured for a single-molecule in a confocal experiment using MFD (Kühnemuth & Seidel, 2001; Widengren et al, 2006). Combining these parameters, it is possible to extract structural information and exchange rate constants in dynamic systems.

In our smFRET experiments using an MFD setup, the fluorescence sample is diluted into low picomolar concentration ($10^{-12}$ M = 1 pM) and placed in a confocal microscope, where a sub-nanosecond laser pulse excites labeled molecules freely diffusing through a detection volume. A typical confocal volume is <4 femtoliters or fL. At such low concentrations only single molecules are detected at a time. The emitted fluorescence from the labeled molecules is collected through the objective and spatially filtered using a pinhole with typical diameter of 100 μm. This step defines an effective confocal detection volume. Then, the signal is split into parallel and perpendicular components at two (or more) different spectral windows (e.g. "green" and "red"). Each photon detector channel is then coupled to time correlated single photon counting (TCSPC) electronics for data registration (Fig. 3.3).
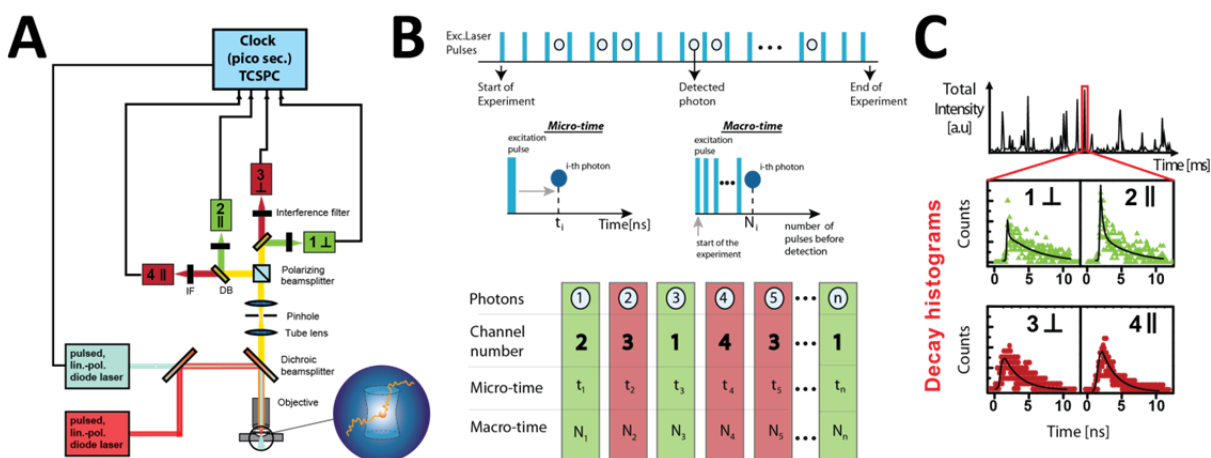


Figure 3.3. **Experimental setup and data registration (Taken from Sanabria 2015, in preparation)**. (**A**) A typical Multiparameter Fluorescence Detection setup is shown and consists of four detectors covering two different spectral windows. Detectors are connected to the time-correlated single photon counting (TCSPC) electronics. (**B**) In TCSPC, each photon is identified by three parameters: (i) micro-time or time after the excitation pulse, (ii) macro-time or number of excitation pulses from the start of the experiment, and (iii) channel number. These three parameters are required for off-line analysis. (**C**) Single molecules diffuse freely through the confocal volume and photons are emitted leaving a burst of photons as a function of time. Each selected burst is fitted accordingly and used for displaying multi-dimensional histograms.

Three parameters are recorded for each photon and used for further off-line data analysis: micro-time, macro-time and channel-number. The micro-time is the elapsed time from the previous excitation pulse. The macro-time is the elapsed time from the start of the experiments. The channel number indicates the spectral window and polarization information of the photon. For each burst fluorescence parameters (e.g. $E, r$) are calculated and decay histograms are formed and fitted. The picosecond timing accuracy allows measurements covering a time-span of >10 orders of magnitude without gaps. Therefore, any mechanism causing temporal fluctuations slower than the fluorescence lifetime can be studied.

MFD for confocal single molecule Förster Resonance Energy Transfer (smFRET) measurements was done using a 485 nm diode laser (LDH-D-C 485 PicoQuant, Germany, operating at 64 MHz, power at objective 110 µW) exciting freely diffusing labeled T4L molecule that passed through a detection volume of the 60X, 1.2 NA collar (0.17) corrected Olympus objective. The emitted fluorescence signal was collected through the same objective and spatially filtered using a 100 µm pinhole, to define an effective confocal detection volume. Then, the signal was divided into parallel and perpendicular components at two different colors ("green" and "red") through band pass filters, HQ 520/35 and HQ 720/150, for green and red respectively, and split further with 50/50 beam splitters. In total eight photon-detectors are used- four for green (τ-SPAD, PicoQuant, Germany) and four for red channels (APD SPCM-AQR-14, Perkin Elmer, Germany). A time correlated single photon counting (TCSPC) module (HydraHarp 400, PicoQuant, Germany) was used for data registration.

For smFRET measurements samples were diluted in measuring buffer with 1 µM unlabeled T4L to pM concentration assuring ~ 1 burst per second. Collection time varied from several minutes up to 10 hours. To avoid drying out of the immersion water during the long measurements an oil immersion liquid with refraction index of water was used (Immersol, Carl Zeiss Inc., Germany). NUNC chambers (Lab-Tek, Thermo Scientific, Germany) were used with 500 µL sample volume. Standard controls consisted of measuring water to determine the instrument response function (IRF), buffer for background subtraction and the nM concentration green and red standard dyes (Rh110 and Rh101) in water solutions for calibration of green and red channels, respectively. To calibrate the detection efficiencies we used a mixture solution of double labeled DNA oligonucleotides with known distance separation between donor and acceptor dyes.

## 3.8.4 Static and dynamic FRET lines in a single molecule experiment

The relationship between the ratio of the donor fluorescence over the acceptor fluorescence $F_D/F_A$ and the fluorescence weighted donor lifetime obtained in burst analysis $\langle\tau_{D(A)}\rangle_f$ depends on specific experimental parameters such as fluorescence quantum yields of the dyes ($\Phi_{FD(0)}$ and $\Phi_{FA}$ for donor and acceptor respectively), background ($\langle B_G\rangle$ and $\langle B_R\rangle$ for green and red channels), detection efficiencies ($g_G$ and $g_R$ for green and red respectively) and crosstalk ($\alpha$). In the $F_D/F_A$ vs. $\langle\tau_{D(A)}\rangle_f$ 2D representations it is useful to represent a static FRET line such as:

$$\left(\frac{F_D}{F_A}\right)_{static} = \frac{\Phi_{FD(0)}}{\Phi_{FA}}\cdot\left(\frac{\tau_{D(0)}}{\tau_{D(A)}}-1\right)^{-1}. \tag{Eq. 3.7}$$

In Eq. 3.7, $\tau_{D(A)} = \langle\tau_{D(A)}\rangle_f$ is the fluorescence averaged lifetime obtained via the maximum likelihood estimator when fitting ~100 green photons per burst. $\tau_{D(0)}$ is the donor fluorescence lifetime in the absence of acceptor.

The corrected fluorescence ($F_D$ and $F_A$) depends on the detection efficiencies of green ($g_G$) and red ($g_R$) channels as follows:

$$F_D = \frac{S_G - \langle B_G\rangle}{g_G}, \tag{Eq. 3.8}$$

$$F_A = \frac{S_R - \alpha F_G - \langle B_R\rangle}{g_R}, \tag{Eq. 3.9}$$

where the total signal in green and red channels are $S_G$ and $S_R$, respectively. The ratio ($F_D/F_A$) is weighted by the species fractions.

To properly describe the FRET line, one needs to consider that fluorophores are moving entities coupled to specific places via flexible linkers. This in turns generates a distance distribution between two fluorophores governed by the linker dynamics. Additionally, one needs to consider that the measured lifetime per burst is the fluorescence weighted average lifetime $\langle\tau_{D(A)}\rangle_f$. Therefore Eq. 3.7 is only valid in the ideal scenario.

The goal is to transform Eq. 3.7 to include the linker dynamics which is slower than the fluorescence decay time.

$$\left(\frac{F_D}{F_A}\right)_{static,L} = \frac{\Phi_{FD(0)}}{\Phi_{FA}} \cdot \left(\frac{\tau_{D(0)}}{\langle \tau_{D(A)} \rangle_{x,L}} - 1\right)^{-1}.$$ (Eq. 3.10)

To include this correction, the first thing to consider is a distance distribution between two fluorophores. We assume a Gaussian probability distance distribution with standard deviation $\sigma_{DA}$ and mean value $\langle R_{DA} \rangle$ such as

$$p(R_{DA}) = \frac{1}{\sqrt{2\pi} \cdot \sigma_{DA}} \exp\left(\frac{-(R_{DA} - \langle R_{DA} \rangle)^2}{2\sigma_{DA}^2}\right).$$ (Eq. 3.11)

For each $R_{DA}$ one can calculate the corresponding species lifetime following to

$$\tau_{D(A)}(R_{DA}) = \tau_{D(0)} \cdot \left(1 + \left(R_0 / R_{DA}\right)^6\right)^{-1},$$ (Eq. 3.12)

where each species corresponds to a different distance between the two fluorophores. For simplicity the donor lifetime is treated as mono exponential decay and $\tau_{D(0)} = \langle \tau_{D(0)} \rangle_x = \langle \tau_{D(0)} \rangle_f$. Each $\tau_{D(A)}(R_{DA})$ has a probability defined by the corresponding distribution $p(\tau_{DA}(R_{DA}))=p(R_{DA})$. The average species lifetime, due to linker dynamics, can be defined in the continuous approximation as

$$\langle \tau_{D(A)} \rangle_{x,L} = \int \tau_{D(A)}(R_{DA}) p(R_{DA}) dR_{DA},$$ (Eq. 3.13)

and the fluorescence average lifetime as

$$\langle \tau_{D(A)} \rangle_{f,L} = \langle \tau_{D(A)} \rangle_f = \frac{\int (\tau_{D(A)}(R_{DA}))^2 p(R_{DA}) dR_{DA}}{\langle \tau_{D(A)} \rangle_{x,L}}.$$ (Eq. 3.14)

Thus, we can set a pair of parametric relations with respect to $\langle R_{DA} \rangle$ corresponding species to the species and fluorescence average lifetime such as

$$\langle \tau_{D(A)} \rangle_{x,L}(R_{DA}) \text{ and } \langle \tau_{D(A)} \rangle_{f,L}(R_{DA}).$$ (Eq. 3.15)

Generally it is analytically impossible to solve Eq. 3.15 using Gaussian distributions. Thus, Eq. 3.15 is solved numerically covering a range for $\langle R_{DA} \rangle = 1$ Å to [5 $R_0$] Å. From the numerical solution we can create an empirical relation between the species and fluorescence average lifetimes for the selected range of $\langle R_{DA} \rangle$'s using an $i^{th}$ order polynomial function of with coefficients $A_{i,L}$ like

$$\left\langle \tau_{D(A)} \right\rangle_{x,L} = \sum_{i=0}^{n} A_{i,L} \left( \left\langle \tau_{D(A)} \right\rangle_{f,L} \right). \tag{Eq. 3.16}$$

Finally, we introduce Eq. 3.16 into Eq. 3.7 and obtain the static FRET line corrected for dye linker movements as

$$\left( \frac{F_D}{F_A} \right)_{\text{static,L}} = \frac{\Phi_{FD(0)}}{\Phi_{FA}} \cdot \left( \frac{\tau_{D(0)}}{\sum_{i=0}^{3} A_{i,L} \left( \left\langle \tau_{D(A)} \right\rangle_{f,L} \right)^{i}} - 1 \right)^{-1}. \tag{Eq. 3.17}$$

Coefficients ($A$'s) vary by variant. Experimentally, there is not difference on the observable thus, unless otherwise specified, we use in all figures and captions the assumption that all measured average lifetimes include the linker effect or $\langle \tau_{D(A)} \rangle_f = \langle \tau_{D(A)} \rangle_{f,L}$.

In the case of transition between two different states, one can also get an equation for a dynamic FRET line. In this case, a mixed fluorescence species arises from the interconversion between two conformational states. For the simplest case the dynamic FRET line can be analytically presented as (Sisamakis et al, 2010)

$$\left( \frac{F_D}{F_A} \right)_{\text{dyn,L}} = \frac{\Phi_{FD(0)}}{\Phi_{FA} \tau_{D(0)}} \cdot \frac{\langle \tau_1 \rangle_f \cdot \langle \tau_2 \rangle_f}{\left( \langle \tau_1 \rangle_f + \langle \tau_2 \rangle_f - \sum_{i=0}^{3} C_{i,L} \left( \left\langle \tau_{D(A)} \right\rangle_f \right)^{i} \right) - \frac{\langle \tau_1 \rangle_f \cdot \langle \tau_2 \rangle_f}{\tau_{D(0)}}} \tag{Eq. 3.18}$$

where $\langle \tau_{D(A)} \rangle_{f,L}$ is the mixed fluorescence lifetime and $\Phi_{FD(0)}$, $\Phi_{FA}$ are the quantum yields of the donor and acceptor dyes, respectively. $\langle \tau_1 \rangle_f$ and $\langle \tau_2 \rangle_f$ are two donor fluorescence lifetimes in presence of acceptor at the beginning and end points of the interconverting states. The $C_{i,L}$ coefficients are determined for each FRET pair and differ from the $A$ coefficients. The $L$ sub index notation is to identify and specify the linker effects.

## 3.8.5 Ensemble time-correlated single photon counting (eTCSPC)

Ensemble Time Correlated Single Photon Counting (eTCSPC) measurements were performed using an IBH-5000U (IBH, Scotland) system. The excitation sources were a 470 nm diode laser (LDH-P-C 470, PicoQuant, Germany) operating at 10 MHz for donor excitation and a 635 nm diode laser (LDH-8-1-126, PicoQuant, Germany) for acceptor excitation. The corresponding slits were set to 2 nm (excitation path) and 16 nm (emission path). Cut-off filters were used to reduce the

contribution of the scattered light (>500 nm for donor and >640 nm for acceptor emission, respectively) and the monochromator was set to 520 nm for green detection and 665 nm for detecting the emission of the acceptor fluorophore. For the measurement of acceptor sensitized emission, the donor was excited at 470 nm and the emission of acceptor fluorophore was detected at 665 nm.

Fluorescence intensity decays were fitted using the iterative re-convolution approach with various models using single, double or triple exponential decays. In general these multi-exponential relaxation models can be described by Eq. 3.19:

$$F_{D(A)}(t) = \left(1 - X_{D(0)}\right) \cdot \sum_i x_{D(A)}^{(i)} \exp(-t / \tau_{D(A)}^{(i)}) + X_{D(0)} \cdot \sum_j x_{D(0)}^{(j)} \exp(-t / \tau_{D(0)}^{(j)}) \qquad \text{(Eq 3.19)}$$

where $\tau_{D(0)}^{(i)}$ is the donor fluoresce lifetime, $x_{D(0)}^{(i)}$ are the pre-exponential factors. $X_{D(0)}$ is the donor only labeled fraction and $\tau_{D(A)}^{(i)}$ are the donor lifetime excepting FRET and the corresponding pre-exponentials $x_{D(A)}^{(i)}$. Excluding dynamic donor quenching the average inter dye distance can be

obtained from $\langle R_{DA} \rangle = R_0 \cdot \left( \dfrac{\tau_{D(0)}}{\tau_{D(A)}} - 1 \right)^{-\frac{1}{6}}$, valid only when $\kappa^2$ is 2/3. $R_0$ is the Förster distance and for the Alexa 488 and Alexa647 FRET pair is 52 Å. The fluorescence decay of the donor in the absence of FRET was multi-exponential, most likely, due to local quenching. To account for this multi-exponential behavior of donor lifetime, it is best to consider that each donor lifetime excerpts the same FRET rate. This is true if quenching does not change the donor radiative lifetime, the spectral overlap and relative orientation factor, known as $\kappa^2$. Then, instead of an empirical multi exponential decay function the donor fluorescence decay in the presence of FRET is given by

$$F_{D(A)}(t) = \sum_i x_{D(0)}^{(i)} \int_{R_{DA}} p(R_{DA}) \exp\left( -\frac{t}{\tau_{D(0)}^{(i)}} \left[1 + (R_0 / R_{DA})^6\right] \right) dR_{DA} \qquad \text{(Eq. 3.20)}$$

In Eq. 3.20, $\tau_{D(0)}^{(i)}$ are the lifetimes of the donor in the absence of FRET, and $x_{D(0)}^{(i)}$ are the corresponding amplitudes (pre-exponential factors). We assume that the distance distribution ($p(R_{DA})$) can be modeled as a Gaussian with mean value $\langle R_{DA} \rangle$ and a half-width of $\sigma_{DA}$. Then Eq. 3.20 can be rewritten as

$$F_{D(A)}(t) = \sum_i x_{D(0)}^{(i)} \int_{R_{DA}} \frac{1}{\sqrt{2\pi} \cdot \sigma_{DA}} \exp\left( -\frac{\left(R_{DA} - \langle R_{DA} \rangle\right)^2}{2\sigma_{DA}^2} \right) \exp\left( -\frac{t}{\tau_{D(0)}^{(i)}} \left[1 + (R_0 / R_{DA})^6\right] \right) dR_{DA} \qquad \text{(Eq. 3.21)}$$

Exemplary fit for the IF3 data using three exponentials is presented in the Figure 3.4 below.
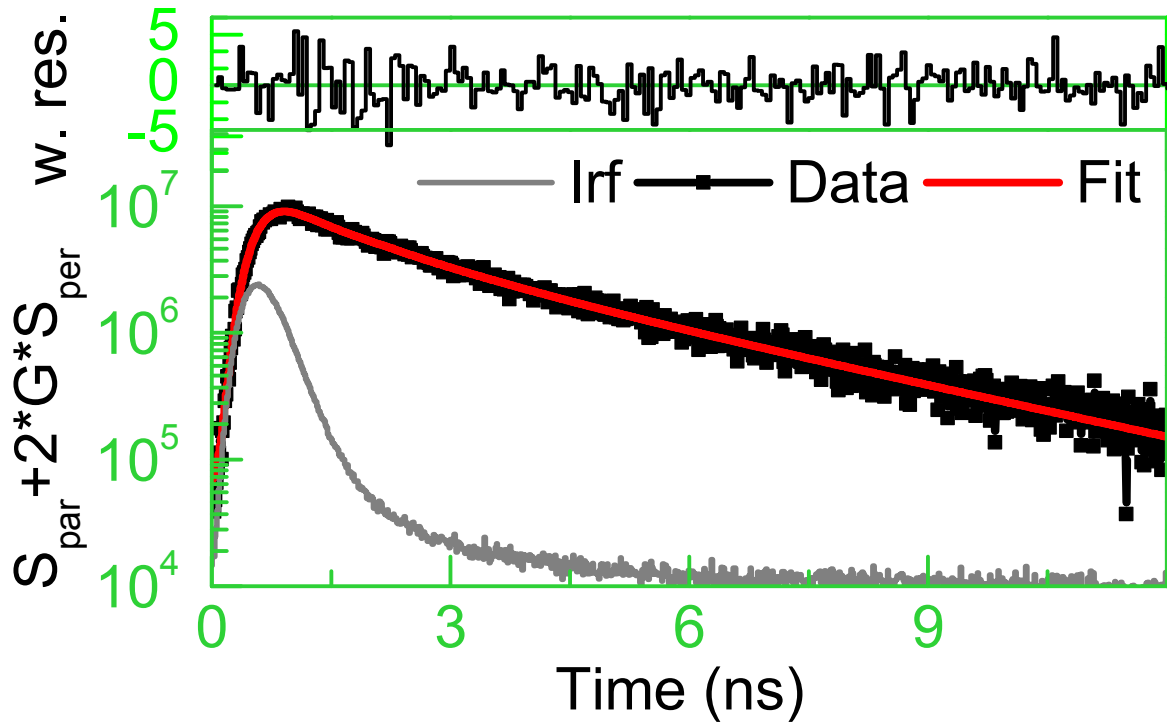


Figure 3.4. **Exemplary lifetime analysis of the IF3 free in solution (taken from IF3 manuscript).** All donor photons are accumulated in a decay histogram for seTCSPC analysis. Three lifetimes fit was used. Experimental data of total signal S is shown in black, the instrument response function in grey and the fit in red. The fit yielded $\tau_1$ = 0.18 ns ($x_1$ = 0.44), $\tau_2$ = 1.02 ns ($x_2$ = 0.19), $\tau_3$ = 3.27 ns ($x_3$ = 0.37), G-Factor = 1.046.

The statistical uncertainties of the fits were estimated by exploring the $\chi^2$-surface of the model function using the Metropolis-Hastings algorithm with at least 20 independent Markov-chains with 50000 steps each starting from the minimum fit-result $p_{min}$ with a step-width of $w = p_{min} \cdot r \cdot w_0$ (where r is a random sample from a uniform distribution over [-0.5, 0.5] and $w_0$=0.005) up to the maximum confidence level $conf_{max}$ = 1-10$^{-5}$. The error-margins of the individual fitting parameters are the projection to the individual parameter-dimension. The maximum allowed $\chi^2_{r,max}$ for a given confidence-level ($P$; e.g. for 2σ $P$ = 0.95) was calculated by:

$$\chi^2_{r,max}(P) = \chi^2_{r,min} \cdot \left[1 + n/\upsilon \cdot \mathrm{cdf}^{-1}(F(n,\upsilon,P))\right] \qquad \text{(Eq. 3.22)}$$

where cdf$^{-1}(F(n,v,P))$ is the inverse of the cumulative distribution function of the $F$-distribution for $n$ number of free parameters, and with $v$ degrees of freedom. $\chi^2_{r,min}$ is the minimum determined $\chi^2_r$ (Lakowicz, 2006).

### 3.8.6 Fluorescence correlation spectroscopy (FCS)

To study fluctuations of any signal one can compute the correlation function of such. In this respect, fluorescence correlation spectroscopy (FCS) (Elson, 2013; Elson & Magde, 1974; Magde et al, 1972) in combination with FRET (FRET-FCS) was developed as a powerful tool (Slaughter et al, 2004; Torres & Levitus, 2007). FRET-FCS allows for the analysis of FRET fluctuations covering a time range of nanoseconds to seconds. Hence, it is a perfect method to study conformational dynamics of biomolecules, complex formation, folding and catalysis (Al-Soufi et al, 2005; Gurunathan & Levitus, 2010; Johnson, 2006; Levitus, 2010; Price et al, 2011; Price et al, 2010; Slaughter et al, 2004; Slaughter et al, 2005a; Slaughter et al, 2005b; Slaughter et al, 2007; Torres & Levitus, 2007).

Structural fluctuations are reflected by the correlation function, which in turn provides restraints on the number of conformational states. In this section we briefly describe the simplest case of FRET-FCS and discuss various experimental scenarios. A detailed review was recently published covering all advantages and challenges of these methods (Felekyan et al, 2013).

The auto/cross-correlation of two correlation channels $S_A$ and $S_B$ is given by:

$$G_{A,B}(t_c) = 1 + \frac{\langle \delta S_A(t) \cdot \delta S_B(t+t_c) \rangle}{\langle S_A(t) \rangle \cdot \langle S_B(t) \rangle} \tag{Eq. 3.23}$$

If $S_A$ equals $S_B$ the correlation function is called an autocorrelation function, otherwise it is a cross-correlation function. If all species are of equal brightness, the amplitude at time zero of the autocorrelation function, $G(t_c = 0)$, allows one to determine the mean number of molecules $N$ in the detection volume, $V_{det}$, or the concentration, $c$, if the parameters of detection volume are known (Eq. 3.24).

$$G(t_c = 0) = 1 + \frac{1}{N} \cdot G_{diff}(t_c) \qquad or \qquad c = \frac{N}{V_{det}} \tag{Eq. 3.24}$$

When assuming a 3-dimensional (3D) Gaussian shaped detection/illumination volume, the normalized diffusion term is given by:

34

$$G_{diff}(t_c) = \left(1 + \frac{t_c}{t_{diff}}\right)^{-1} \cdot \left(1 + \left(\frac{\omega_0}{z_0}\right)^2 \cdot \frac{t_c}{t_{diff}}\right)^{-\frac{1}{2}}$$ (Eq. 3.25)

whereas $\omega_0$ and $z_0$ are shape parameters of the detection volume, which is defined by $w(x,y,z) = \exp(-2(x^2 + y^2)/\omega_0^2)\exp(-2z^2/z_0^2)$. For 1-photon excitation the characteristic diffusion time related to the diffusion coefficient $D_{diff}^{(i)}$ is expressed by $t_{diff}^{(i)} = \omega_0^2/4D_{diff}^{(i)}$. The autocorrelation function allows direct assessment of the diffusion constant $D_{diff}$.

To study the conformational dynamics of free IF3 with high temporal resolution, FCS curves (green-to-green (GG) and red-to-red (RR) autocorrelations, green-to-red and red-to-green (GR and RG) color cross-correlations) were computed from smMFD data by custom-made software correlator (LabVIEW, National Instruments Co.) (Fig. 3.5). The data were analyzed (Felekyan et al, 2013) using the following equations:

$$G_{G,G}(t_c) = 1 + \frac{1}{N_{G,G}} \cdot G_{diff}(t_c) \cdot \left[1 - \sum_{i=1}^{4}\left(AC_{G,G}^{(i)} \cdot \left(1 - e^{-\frac{t_c}{t_R(i)}}\right)\right)\right]$$

$$G_{R,R}(t_c) = 1 + \frac{1}{N_{R,R}} \cdot G_{diff}(t_c) \cdot \left[1 - \sum_{i=1}^{4}\left(AC_{R,R}^{(i)} \cdot \left(1 - e^{-\frac{t_c}{t_R(i)}}\right)\right)\right]$$ (Eq. 3.26)

$$G_{G,R}(t_c) = 1 + \frac{1}{N_{G,R}} \cdot G_{diff}(t_c) \cdot \left[1 - \sum_{i=1}^{4}\left(CC_{G,R}^{(i)} \cdot e^{-\frac{t_c}{t_R(i)}}\right)\right]$$

where $N_x$ is an average number of molecules for corresponding correlation, $G_{diff}(t_c)$ – diffusion term, $AC_{G,G}^{(i)}$ and $AC_{R,R}^{(i)}$ – fractions of respective bunching terms, respectively, $CC_{G,R}^{(i)}$ – fractions of respective anti-correlation terms in cross-correlation curves, and $t_R^{(i)}$ – is the $i$-th relaxation time shared in $i$-th bunching or anti-correlation terms in global fit of FCCS curves.

To determine the fractions of IF3 free and bound to the 30S subunit via their characteristic diffusion times $t_{d\text{-}free}$ and $t_{d\text{-}complex}$, respectively, the following equation was used to fit auto-correlation curves $G_{G,G}(t_c)$ of the green signal $S_G$:

$$G_{G,G}(t_c) = 1 + \frac{1}{N} \cdot \left(x_{free} \cdot G_{diff-free}(t_c) + (1 - x_{free}) \cdot G_{diff-complex}(t_c)\right) \cdot \left(1 - T + T \cdot e^{-\frac{t_c}{t_T}}\right)$$ (Eq. 3.27)
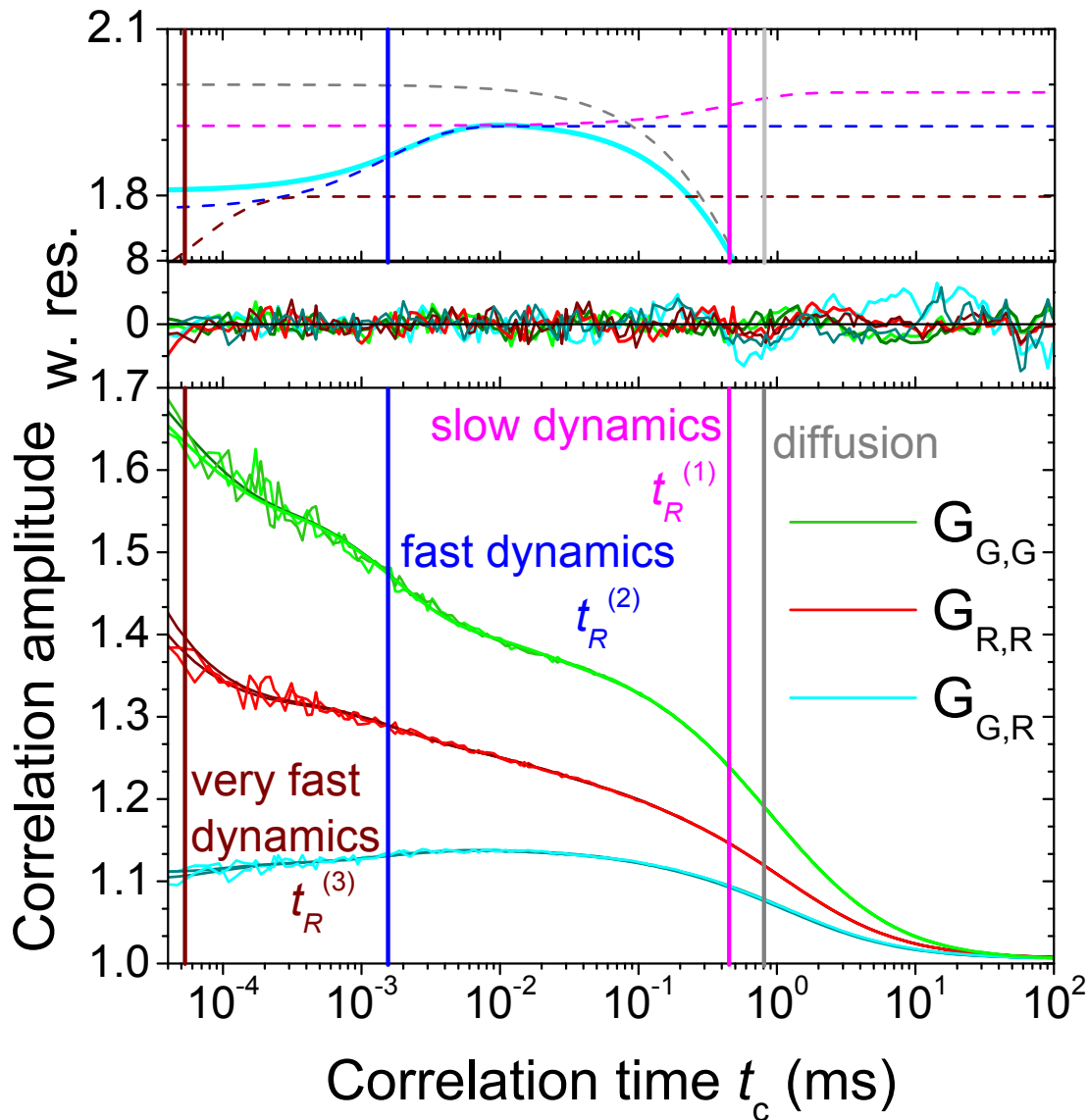
Figure 3.5. **Exemplary FCS analysis taken from IF3 manuscript**. Six correlations were calculated, two green auto-correlations $G_{G,G}$ (green lines), two red auto-correlations $G_{R,R}$ (red lines) and two cross-correlations $G_{G,R}$ green to red and red to green (cyan lines). The curves were globally fitted by a model with six bunching terms or the auto-correlation function and two bunching and four anti-bunching term for the cross-correlation function. The fit yield $t^{(1)}_R = 470$ µs (magenta vertical line); $t^{(2)}_R = 1.6$ µs (blue vertical line); and $t^{(3)}_R = 60$ ns (dark vertical line); diffusion time is 800 µs as indicated by gray vertical line. Respective residuals are shown on top. The top panel indicates the contribution of different dynamic terms to the overall shape of the cross-correlation curve – $t^{(3)}_R$ in dark red, $t^{(2)}_R$ in blue, $t^{(1)}_R$ in magenta, and the diffusion term in gray. The resulting cross-correlation term is shown in cyan.

where $t_c$ is the correlation time, $N$ is the total number of molecules, $x_{free}$ is the fraction of the unbound IF3 with the diffusion term $G_{diff\text{-}free}$, $G_{diff\text{-}complex}$ is the diffusion term of IF3 bound to the 30S subunit, and $T$ is the fraction of molecules in triplet (dark) state with the characteristic triplet time $t_T$.

### 3.8.7 Filtered fluorescence correlation spectroscopy (fFCS)

To separate species, we use filtered FCS (fFCS) (Böhmer et al, 2002; Felekyan et al, 2012). fFCS differs from standard FCS and FRET-FCS by interrogating the "species" (conformational states) fluctuations instead of photon count rates. We define the species cross- correlation function (sCCF) as

$$G^{(i,m)}(t_c) = \frac{\left\langle F^{(i)}(t) \cdot F^{(m)}(t+t_c) \right\rangle}{\left\langle F^{(i)}(t) \right\rangle \cdot \left\langle F^{(m)}(t+t_c) \right\rangle} = \frac{\left\langle \left( \sum_{j=1}^{d \cdot L} f_j^{(i)} \cdot S_j(t) \right) \cdot \left( \sum_{j=1}^{d \cdot L} f_j^{(m)} \cdot S_j(t+t_c) \right) \right\rangle}{\left\langle \sum_{j=1}^{d \cdot L} f_j^{(i)} \cdot S_j(t) \right\rangle \cdot \left\langle \sum_{j=1}^{d \cdot L} f_j^{(m)} \cdot S_j(t+t_c) \right\rangle} \quad \text{(Eq. 3.28)}$$

where (*i*) and (*m*) are two selected "species" in a mixture. A set of filters $f_j^{(i)}$ that depend on the arrival time of each photon after each excitation pulse is used (Fig. 3.6). The signal $S_j(t)$, obtained via pulsed excitation is recorded at each j = 1 ... $L$ TCSPC-channel. The signal and filters per detector, *d*, are stacked in a single array with dimensions $d \cdot L$ for global minimization according to (Felekyan et al, 2012). Filters are defined in such a way that the relative "error" difference between the photon count per species ($w^{(i)}$) and the weighted histogram $f_j^{(i)} \cdot H_j$ is minimized as defined in Eq. 3.29.

$$\left\langle \left( \sum_{j=1}^{d \cdot L} f_j^{(i)} \cdot H_j - w^{(i)} \right)^2 \right\rangle \to \min. \quad \text{(Eq. 3.29)}$$

where brackets represent time averaging.

The requirement is that the decay histogram $H_j$ can be expressed as a linear combination of the conditional probability distributions $p_j^{(i)}$, such as $H_j = \sum_{i=1}^{n} w^{(i)} p_j^{(i)}$, with $\sum_{j=1}^{d \cdot L} p_j^{(i)} = 1$. Hence, the species cross- correlation $G^{(i,m)}(t_c)$ provides maximal contrast for intercrossing dynamics (Felekyan et al, 2012). One major advantage of sCCF is that if photophysical properties are decoupled from species selection the intercrossing dynamics is recovered with great fidelity. Exemplary fFCS curves with the respective fits are shown in Figure 3.7.
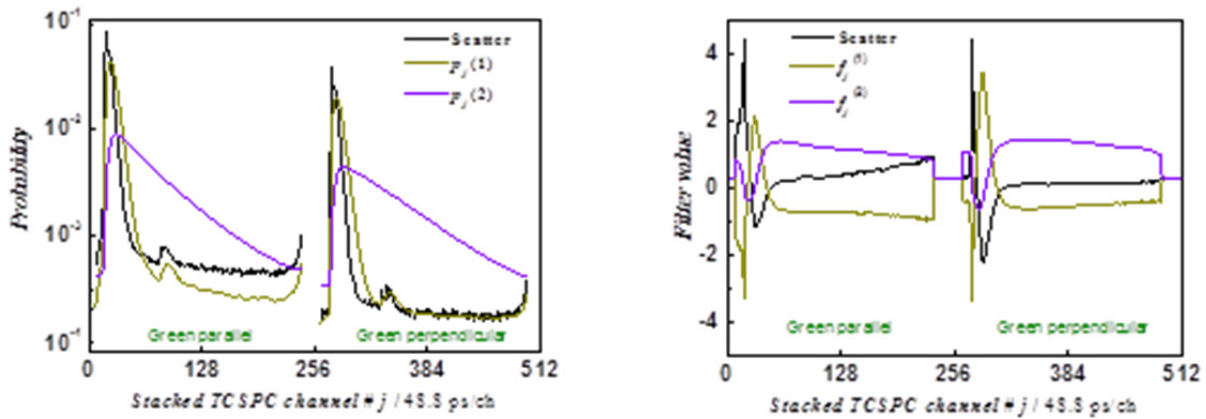
Figure 3.6. **Exemplary fFCS analysis of the S44pAcF/I150C-DA mutant from the T4L manuscript**. (**A**) Generated decays for two pseudo-species. The parameter of the decays are fluorescence lifetime $\tau_1 = 0.25$ ns and rotational correlation $\rho_1 = 3.3$ ns for decay 1 (yellow) and fluorescence lifetime $\tau_2 = 2.97$ ns and the rotational correlation $\rho_1 = 3.3$ ns for decay 2 (magenta). Scatter profile in black for used for both decays. (**B**) Filters $f_j^{(i)}$ were calculated according to Eq. 35 using the decays from (A).
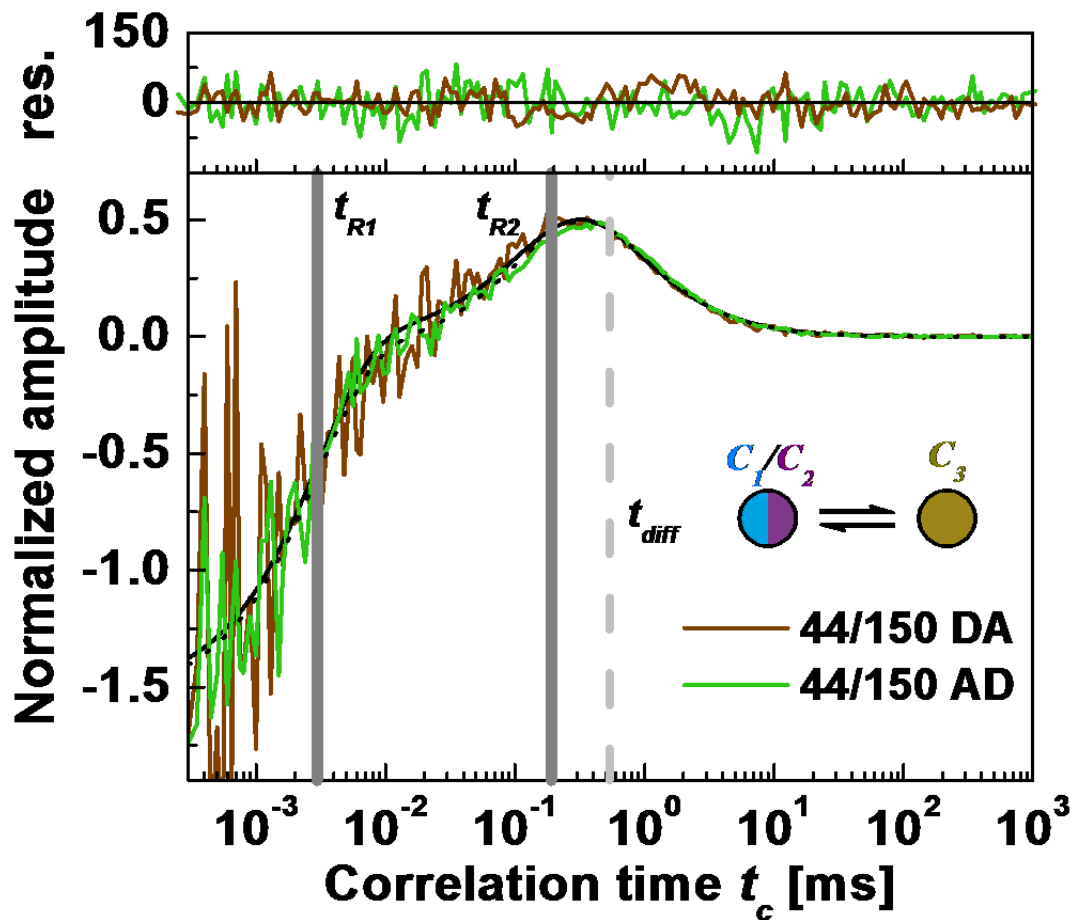


Figure 3.7. **Exemplary fFCS curves (taken from T4L manuscript)**. Overlay of the normalized sCC functions of S44pAcF/I150C-DA and S44pAcF/I150C-AD. The fit shows two relaxation times $t_{R1} = 4.0 \pm 2.4$ µs and $t_{R2} = 230 \pm 28$ µs and diffusion time $t_{diff} = 0.54$ ms as indicated by grey lines.

38

To properly fit the species cross-correlation function we used

$$G(t_c) = 1 + \frac{1}{N} \cdot G_{diff}(t_c) \cdot [1 - G_K(t_c)]$$ (Eq. 3.30)

where $G_K(t_c)$ is defined as

$$G_K(t_c) = \sum_{t_{Ri}}^{t_{Rn}} A_{Ki} \exp(-t_c / t_{Ri})$$ (Eq. 3.31)

In Eq. 3.31 the summation is over $n$ reaction times $t_{Rn}$. The same 3-dimensional Gaussian shaped volume element is assumed. We assume that $G_{diff}(t_c) = G_{diff}^{(i)}(t_c) = G_{diff}^{(m)}(t_c)$ take the form of Eq. 31. The normalized correlation function is presented as:

$$g(t_c) = N \cdot (G(t_c) - 1).$$ (Eq. 3.32)

Filtered FCS requires prior knowledge of the time-resolved fluorescence and polarization decays for each species. For a mixture of more than two species, we generated two decays corresponding to two "pseudo-species". Using the scatter profile as the excitation pulse the parallel and perpendicular decay components $F_{\parallel}(t)$ and $F_{\perp}(t)$ for each "pseudo-species" were modeled as

$$F_{\parallel}(t) = F(t) \cdot (1 + (2 - 3l_1) \cdot r(t)) / 3$$
$$F_{\perp}(t) = F(t) \cdot (1 - (1 - 3 \cdot l_2) \cdot r(t)) / 3$$ (Eq. 3.33)

where $F(t)$ is the time-resolved fluorescence decay at magic angle (typically mono-exponential), and $l_1 = 0.01758$ and $l_2 = 0.0526$ are correction factors (Koshioka et al, 1995; Schaffer et al, 1999). The anisotropy decay $r(t)$ is given by

$$r(t) = r_{0,ov} \exp(-t / \rho_{overall}) + r_{0,ba} \exp(-t / \rho_{backbone}) + r_{0,li} \exp(-t / \rho_{linker})$$ (Eq. 3.34)

Background signal consists of dark counts (uniformly distributed over TCSPC channels) and scatter contribution.

### 3.8.8 Photon distribution analysis (PDA)

System properties can be qualitatively described by visual inspection of FRET histograms guided by FRET lines. To extract parameters like FRET efficiencies, rate-constants and population fractions quantitatively fitting of FRET efficiency histograms is a common method (Gopich & Szabo, 2005; Gopich & Szabo, 2003; Gopich & Szabo, 2007; Sisamakis et al, 2010; Zhang & Yang, 2005). In PDA this is accomplished by calculating the probability $P(S_G,S_R)$ to observe a certain combination of photons collected in the "green" ($G$) and "red" ($R$) detection channel given a certain time-window:

$$P(S_G,S_R) = \sum_{F_G+B_G=S_G;F_R+B_R=S_R} P(F)P(F_G,F_R|F)P(B_G)P(B_R) \qquad \text{(Eq. 3.35)}$$

where the fluorescence intensity distribution, $P(F)$, is obtained from the total signal intensity distribution $P(S)$ assuming that the background signals $B_G$ and $B_R$ are distributed according to Poisson distributions, $P(B_G)$ and $P(B_R)$, with known mean intensities $\langle B_G \rangle$ and $\langle B_R \rangle$. The conditional probability $P(F_G, F_R|F)$ is the probability of observing a particular combination of green and red fluorescence photons, $F_G$ and $F_R$, provided that the total number of registered fluorescence photons is $F$, and can be expressed as a binomial distribution (Antonik et al, 2006):

$$P(F_G,F_R\,|\,F) = \frac{F!}{F_G!(F-F_G)!}\, p_G^{\,F_G}\left(1-p_G\right)^{F-F_G} \qquad \text{(Eq. 3.36)}$$

here $p_G$ is the probability a detected photon is registered in the "Green" detection channel. The value of $p_G$ is unambiguously related to the FRET efficiency $E$ according to:

$$p_G = \left(1 + \alpha + \gamma \cdot \frac{E}{(1-E)}\right)^{-1} \qquad \text{(Eq. 3.37)}$$

where $\gamma = \dfrac{\Phi_{F,A}^{(A,0)}}{\Phi_{F,D}^{(D,0)}} \cdot \dfrac{g_R}{g_G}$ and $\alpha$ is the crosstalk from G to R channel.

Knowledge of $P(S_G, S_R)$ is sufficient to generate theoretical 1D histograms of any FRET-related parameter, which can be expressed as a function of $S_G$ and $S_R$ (e.g. signal ratio $S_G/S_R$ or FRET efficiency $E$ (Kalinin et al, 2007)) (Fig. 3.8). Fitting such histograms obtained for a single species requires only one floating parameter, $p_G$.

To analyze dynamic processes quantitatively Sisamakis et al. (Sisamakis et al, 2010) developed a toolkit, referred to as dynamic PDA, that combines the mathematical formalism of Palo et al. (Palo et al, 2006) and Gopich et al. (Gopich & Szabo, 2007) with all essential steps towards the complete description of experimental data (including brightness variations, shortening of the observation time due to diffusion and contribution of multimolecular events). Aforementioned authors laid the groundwork for dynamic PDA. It was applied to characterize exchange between two or more FRET states using global analysis of multiple time windows. Dynamic PDA can quantify the processes with relaxation times of 0.1-10 fold of the selected time window, ranging from 0.1 ms (limited by photon statistics) up to 10 ms (limited by diffusion time) . In order to compliment PDA and its variations FCS and lifetime filtered FCS, is an additional, powerful tool that extends the possible range of kinetics studies by six decades in time.
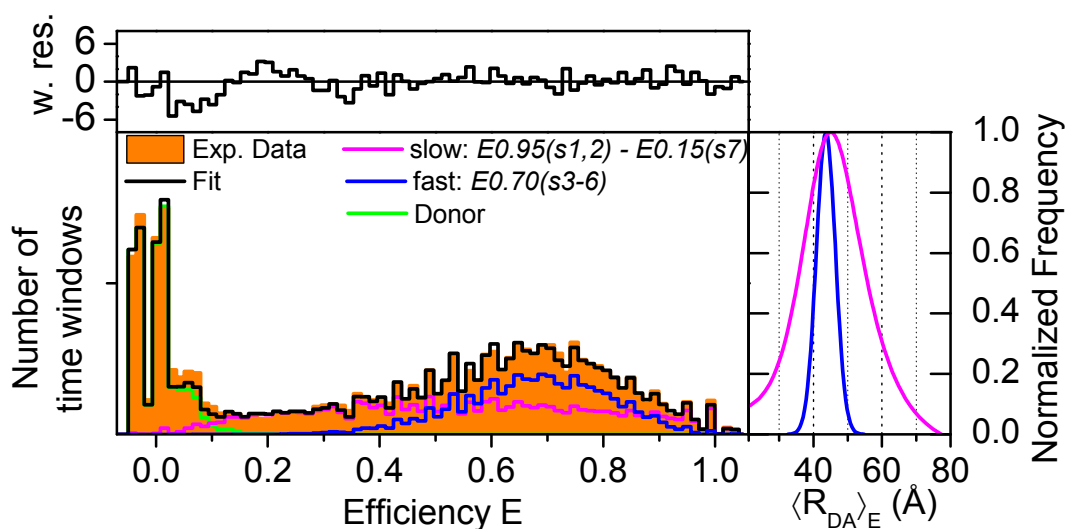


Figure 3.8. **Exemplary PDA analysis of the E values of free IF3 in a time window of 1 ms (taken from IF3 manuscript)**. The black line represents the fit containing two dynamic and one donor-only population. Magenta line represents the dynamic population with the end distance of 30 Å and 69 Å and $k_{12}$ = 2.08 ms$^{-1}$ and $k_{21}$ = 1.76 ms$^{-1}$; blue line represents dynamic population with very fast interconversion rates resulting in a quasi-static population with Gaussian distance distribution with $\langle R_{D(A)} \rangle_E$ = 45 Å and width w = 8 Å. The panel on the right side indicates the distributions of the corresponding FRET populations.

## 3.9. Computational methods (Part 2)

### 3.9.1. Simulation of the FRET data

To verify or decipher complex kinetic schemes a multistep simulation procedure was employed. In the first step a multistate matrix is conceived based on the data previously obtained from seTCSPC, 2D histograms, PDA and (f)FCS analysis. This matrix specifies number of states, FRET efficiency $E$ or lifetime $\tau$ of these states (from seTCSPC or PDA) and the rate transition matrix using the relaxation times acquired by (f)FCS. The assignment of the transition to the states can be free or bound by some pre-knowledge. In case of IF3, three relaxation times were assigned based on the expected mechanism of action, f.e. very fast relaxation time of 65 ns was assigned as free movement of the different domains (IF3 manuscript).

In the next step the expected 1D and 2D histograms for certain amount of bursts were simulated from the state matrix and rate transition matrix. For detailed description of the simulation process refer to IF3 manuscript supplementary information section 12. Furthermore, seTCSPC parameters of lifetime and fractions can be easily obtained from equilibrium fractions and efficiencies. The simulation also allows to calculate resulting FCS curves for all correlation combinations, e.g. $G_{G,G}$, $G_{R,R}$, $G_{G,R}$ and $G_{R,G}$. These three different outputs can be compared to the measured experimental data and if needed re-assignment of the states/rates or adjustment of the parameters is performed to achieve better results as can be judged by a $\chi^2$ parameter.

### 3.9.2 Discrimination for the minimal simulation model

To compare different models and validate the choice between them we employed a probability density function method to combine all $\chi2$ values from all types of analysis into a single parameter, thus enabling the comparison of the previously not comparable results. This allows comparing different results from simulations as in the case of IF3, as well as different mutants as in the case of T4L distance network.

As described in IF3 manuscript and T4L manuscript, supplementary section 14 we calculate two values, $F$ and $p$:

$$F = \chi^2(dataset_1) / \chi^2(dataset_2) \tag{Eq. 3.38}$$

where $dataset_1$ and $dataset_2$ are two set which are compared with their corresponding $\chi^2$ values and

$$p = F.DIST(F, dof(dataset_1), dof(dataset_2), TRUE) \hspace{2cm} \text{(Eq. 3.39)}$$

where *F.DIST* is the probability density function as calculated in Excel (Microsoft), *dof* is the number of degrees of freedom for the given data set. *p* is the confidence level with which the *dataset_1* describes the data better than *dataset_2*. Values for *p* vary from 0.5 to 1, with 0.5 indicating no difference between the sets and 1 meaning 100% confidence for *datatset_1* to describe the experiment better than *dataset_2*.

The corresponding values are listed in Table 5.S4 for IF3 and in Table X.X for T4L. For both proteins, more complicated eight states (IF3) and three states (T4L) models describe the data significantly better than the simpler models. In the case of IF3 this is the first time a comparison of three techniques (five different analyses) were compared within a framework of a single test.

# 4. T4 lysozyme manuscript

This chapter corresponds to an almost finished manuscript "Dynamics and function of transient states in single enzyme molecules", which is very close to submission.

The manuscript is based on collaborative work in a joint project of following persons: Hugo Sanabria, Dmitro Rodnin, Katherina Hemmen, Thomas Peulen, Suren Felekyan, Mark R Fleissner, Mykola Dimura, Felix Koberling, Ralf Kühnemuth, Wayne Hubbell, Holger Gohlke and Claus A.M. Seidel.

The contributions to this manuscript are as follows: H.S. and K.H. purified and labeled the protein. H.S., T.P., K.H., and D.R. measured and analyzed the FRET experiments. T.P., M.D. and H.G. performed structural screening. M.R.F. performed and analyzed EPR experiments. W.H. performed study design and EPR analysis. S.F. developed fluorescence analysis tools. F.K. and R.K. developed fluorescence instrumentation and gave technical support. All authors discussed the results and commented on the manuscript. H.S., W.H. and C.A.M.S. wrote the paper. C.A.M.S. supervised the project

HS: 20%

DR: 20%

KH: 20%

TP: 8%

MRF:2%

SF: 4%

MD: 3%

FK: 2%

RK: 3%

WH: 3%

HG: 4%

CAMS: 11%

4.1 T4 lysozyme manuscript main text

# Dynamics and function of transient states in single enzyme molecules

**Authors:** Hugo Sanabria[1,2*], Dmitro Rodnin[1], Katherina Hemmen[1], Thomas Peulen[1], Suren Felekyan[1], Mark R Fleissner[3]†, Mykola Dimura[1,5], Felix Koberling[4], Ralf Kühnemuth[1], Wayne Hubbell[3], Holger Gohlke[5], Claus A.M. Seidel[1*]


#: contributed equally
*: corresponding authors

## Affiliations

[1]Institut für Physikalische Chemie, Lehrstuhl für Molekulare Physikalische Chemie, Heinrich-Heine-Universität, Düsseldorf, Germany.

[2]Department of Physics and Astronomy, Clemson University, Clemson, South Carolina, U.S.A.

[3]Jules Stein Eye Institute and Department of Chemistry and Biochemistry, University of California, Los Angeles, U.S.A.

[4]PicoQuant GmbH, Berlin, Germany.

[5]Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität, Düsseldorf, Germany.

*Correspondence to: cseidel@hhu.de

Keywords:

## Abstract

To understand enzyme function it is key to visualize transiently populated conformational states and their corresponding exchange pathways. Here, we map over seven decades in time (ns-ms) the conformational landscape of T4 Lysozyme (T4L) in three different enzyme states defined by substrate interaction. To accomplish this, we introduce a hybrid approach involving FRET-based multiparameter fluorescence and EPR spectroscopy, mutagenesis, and FRET-positioning and screening. To describe enzyme function, we use a state matrix of three enzyme states and three short lived conformational states (open, closed, and compact) sampled at 4 μs and 230 μs. The 230 μs bending motion has not been identified before and leads to the compact state, which is more closed than any known structural model of T4L. After screening over 400 structural models in the protein data bank (PDB), using 24 FRET-derived distances of each conformational state, we conclude that the best representative model for the open state is PDB ID 172L, for the closed state PDB ID 1C67, and no structure could be correlated to the new compact state. Furthermore, the closed state 1C67

differs by only 0.7 Å RMSD from the adduct form of T4L (PDB ID 148L). Moreover, the observed changes in equilibrium between transiently populated conformational states in mutants that mimic the catalytic states suggest that the new compact state is involved in the product release state as predicted by the Michaelis-Menten model for non-ATP/GTP hydrolyzing enzymes. We believe that this step can be an evolutionary advantage when fast catalysis is required. Furthermore, the results highlight the potential of our hybrid approach to observe transiently populated conformational states in enzyme catalysis.

## Significance Statement

Thermal fluctuations between distinct protein conformations are critical for function and disease. However, transient states are often too short lived to be detectable by classical biophysical methods. We introduce a quantitative FRET-based approach to probe low-populated states of enzymes from ns to ms timescale, and describe their kinetic network connections. The power of this approach is demonstrated for the hitherto two-state enzyme T4 Lysozyme, where a third short-lived state is unexpectedly detected. Thus, we suggest that T4L samples three functional conformational states, with the third state being involved in product release. In the context of Michaelis-Menten kinetics this will have impact on many non-ATP/GTP-hydrolyzing enzymes where stochastic product release is expected.

## 4.1.1. Introduction

Enzymes are biological catalysts that accelerate chemical reactions. During catalysis enzymes sample defined structural conformations (Hammes, 1964), including low populated states over a wide temporal domain (Henzler-Wildman & Kern, 2007). Their structure and function are linked by conformational transitions (Smock & Gierasch, 2009). Transient states are also related with disease (Selkoe, 2003). During the last decade, various experimental and theoretical tools were introduced to study the enzymes' dynamic behavior (Chung et al, 2009; Henzler-Wildman & Kern, 2007) and our view of rigid enzymes has changed. However, gathering a full overview of enzyme dynamics and characterizing excited conformational states is one of the most difficult experimental problems in molecular biophysics. Transient states are frequently termed "hidden states" because they are often short lived and hidden in classical methods used in structural biology. All spectroscopic techniques such as nuclear magnetic resonance (NMR) (Kleckner & Foster, 2011), electron paramagnetic resonance (EPR) (McHaourab et al, 1996), FRET (Woźniak et al, 2008), and lower resolution methods (Henzler-Wildman & Kern, 2007), face different method-specific restrictions in sensitivity, size, molecular flexibility as well as spatial and temporal resolution. X-ray crystallography provides structure information at very high resolution but it is constrained by the crystal; thus, biomolecular dynamics is usually lost. NMR spectroscopy does provide detailed equilibrium dynamic information but it is limited by sample size and requires complex labeling for large systems. NMR relaxation methods are used to probe protein dynamics at the fast end (picoseconds to nanoseconds) as well as in a much slower range (sub-milliseconds and slower) of this broad spectrum of time scales (Mulder et al, 2001). Finally, other spectroscopic approaches can provide extremely high temporal and spatial resolution information about dynamic behavior, but this information is very detailed with respect to the local environment around the spectroscopic probe (e.g., UV-vis, IR, fluorescence, EPR), or comprises sparse long-range distance information (pulsed EPR, NMR, FRET) reporting global structural changes, where the FRET advantage over EPR is the precise determination of reaction rate constants.

Here we introduce a unique hybrid approach combining single-molecule and ensemble techniques to probe low-populated states of enzymes over a range of seven decades of time (ns-ms). Our toolkit as outlined in Figure 4.1A uses fluorescence lifetime measurements (ensemble Time Correlated Single Photon Counting; eTCSPC), confocal Multiparameter Fluorescence Detection (MFD) of single diffusing proteins (Sisamakis et al, 2010), and filtered fluorescence correlation spectroscopy (fFCS; see SI section 4.2.1.12 and (Böhmer et al, 2002; Felekyan et al, 2012)).

Moreover, we corroborated our results by the use of other biophysical tools, such as site-directed mutagenesis, chromatography, and EPR spectroscopy (McHaourab et al, 1996).
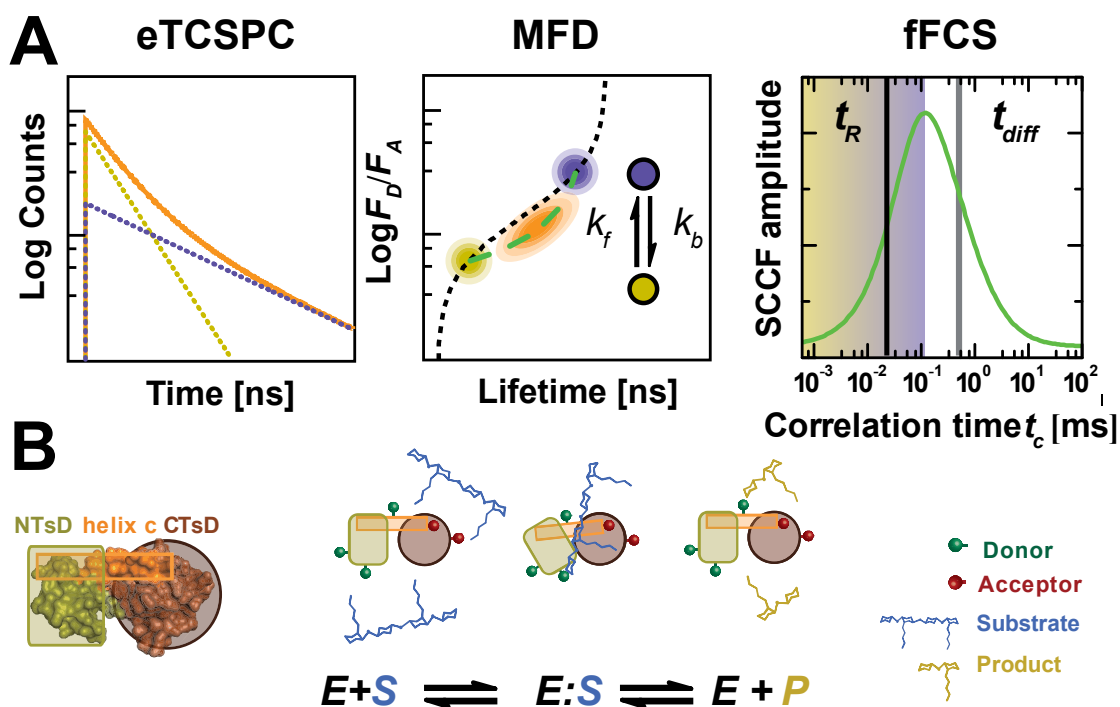


Figure 4.1 (**A**) **Schematic of the high precision FRET and fluctuation analysis toolkit**. eTCSPC resolves stable populations on the ns timescale. MFD experiments of freely diffusing single molecules are analyzed by the FRET indicator $F_D/F_A$ (ratio of donor over acceptor fluorescence) accumulated during single burst duration on the ms timescale. Unimodal distributions occur when dynamic conformational mixing is faster than the burst duration and follow the dashed green line in the MFD plot. fFCS computes the species-specific cross-correlation function (sCCF, Eq. (4.2.S20)). The observed anti-correlation has a characteristic relaxation time $t_R$ related to the exchange rate constant. (**B**) **Schematic of enzymatic activity of T4L**. Upon cleavage of its substrate peptidoglycan, the N-terminal lobe (green) and the C-terminal lobe (brown) of T4L are closing (Goto et al, 2001; Kuroki et al, 1993). This change in closure between the two lopes can be followed by the induced change in FRET.

We study the hinge-bending dynamics of the bacteriophage T4 lysozyme (T4L) and how enzyme dynamics relates to its function. T4L cleaves the glycosidic bond between *N*-acetylmuramic acid and *N*-acetylglucosamine of saccharides of the bacterial cell wall via an oxocarbenium ion-like intermediate (Kuroki et al, 1993). Structurally, T4L (Matthews & Remington, 1974) consists of two subdomains (N-terminal subdomain, NTsD, and C-terminal subdomain, CTsD) connected via a long alpha helix (Helix **c**) (Fig. 4.1B) (Llinas et al, 1999a). There are now more than 500 entries in the Protein Data Bank (PDB) after the original structure (PDB ID: 1LZM; (Matthews & Remington,

1974)). From these studies, two major conformations of T4L are known. In the absence of substrate T4L adopts a conformation $C_1$ which is open at various degrees (Goto et al, 2001; Zhang et al, 1995). However T4L adopts a closed conformation $C_2$, when substrate is bound (Kuroki et al, 1993). This movement corresponds to a classic hinge bending motion of the NTsD. In view of the Michaelis-Menten mechanism for enzyme catalysis, T4L has been considered a typical two-state system (Fig. 4.1B) (Yirdaw & McHaourab, 2012). Here, typically the product is expected to dissociate stochastically from enzymes. Recently, Lu and coworkers (Zheng & Lu, 2014) proposed for Horse reddish peroxidase that the product may be actively spilled out in the presence of strong enzyme-product interactions (Zheng & Lu, 2014). Therefore, a modified Michaelis-Menten model for enzyme kinetics with the inclusion of a product release state is debated (Johnson & Goody, 2011; Kou et al, 2005). However, also several studies of T4L provided some hints that indeed more than two states could be involved in catalysis. Several models of hinge-bending motions were suggested from early observations of crystal structures. Conversely, often-contentious results are found (Yirdaw & McHaourab, 2012). Various computational simulations estimated a nanosecond hinge bending motion (de Groot et al, 1998; Lange & Grubmüller, 2008; Zacharias, 2008). NMR studies (Goto et al, 2001) suggested that inter-domain closure occurs faster than 170 µs. EPR (McHaourab et al, 1996) and fluorescence correlation spectroscopy (Yirdaw & McHaourab, 2012) found evidence for dynamics in the micro- to nanoseconds timescales. Moreover Bouvignies *et al.* observed internal dynamics (about 1 ms) in the CTsD for the L99A mutant of T4L (Bouvignies et al, 2011). From experiments on enzyme catalysis under various conditions (turbidity, FRET, and carbon nanotube currents (Choi et al, 2012; Lu, 2011; Tsugita & Inouye, 1968)) the substrate dependent turnover times were estimated to be 10 - 50 ms with additional faster electrical current fluctuations of a few ms observed in functionalized carbon nanotubes. Recently, single molecule FRET experiments suggested that T4L exhibits complex conformational motions beyond the hinge-bending dynamics (Lu & Lu, 2014).

To address this goal, we map the conformational landscape of T4L by using the above hybrid approach (Fig 4.1A) and our findings show a hidden third conformational state whose function in the enzymatic cycle is discussed. To warrant our results we did a systematic study of 24 doubly labeled variants capturing all possible motions of the T4L backbone (Fig 4.S1A) and applied our toolkit with three methods to resolve all states and their equilibrium fractions and probe their time evolution. To derive the energy landscape of the T4L's cleavage cycle, mutations affecting two active site residues were introduced to capture the enzyme at various steps during substrate hydrolysis.

## 4.1.2. Results

To study the conformational space and dynamics of T4L we performed ensemble and single molecule FRET studies on 24 FRET variants (the workflow is outlined in Figure 4.S1). Each of the 24 variants is site-specifically labeled at the two subdomains by covalently attaching one fluorophore to a cysteine and the second fluorophore to the unnatural amino acid *p*-acetyl-L-phenylalanine (pAcF) via orthogonal chemistry (Brustad et al, 2008; Fleissner et al, 2009) (Fig. 4.2A). This strategy allowed us to heed specific fluorophore effects in FRET measurements by preparing labeled samples in both possible configurations, named "(DA)" when the donor is attached to the NTsD, and "(AD)" for the reverse order.

We selected the dye positions based on the wealth of structural information already available in the PDB. We choose representative structures for the known open and closed conformations $C_1$ and $C_2$ of T4L (PDB ID: 172L and 148L, respectively), and use FPS (FRET positioning and screening) to FRET label *in silico* and compute the expected experimental interdye distances (see section 4.1.4.2 and SI section 4.2.1.15). The mean interdye distance $\langle R_{DA} \rangle$ is directly measured from time resolved FRET experiments (eTCSPC). Treating the dimensionality of the fluorophores is required for proper structural interpretation but care must be taken of the fact that $\langle R_{DA} \rangle$ should not be confused with the backbone (Cα-Cα) distance (Table 4.S1B,C). To obtain a suitable conversion between $\langle R_{DA} \rangle$ and the backbone distance an appropriate structural model is required (Sindbert et al, 2011a) as shown in SI section 4.2.1.3 and 4.2.1.15. Based on our initial network design, the hinge-bending motion should lead to a two state model with FRET distance fluctuations as a function of time. Instead, as we will show below, the first measurements showed three conformational states. Thus we designed additional variants (Table 4.S1) to determine the total conformational space of the protein and confirm the existence of the third state. We note that some mutants were designed in such a way, that $C_1$ and $C_2$ conformations were expected to yield the same distance. In this case the data for a two state system would yield a single FRET quenching term of the donor. Thus, we have two classes of variants: (1) $C_1$ and $C_2$ cannot be distinguished; (2) $C_1$ and $C_2$ can be distinguished. Furthermore, we oversampled the conformational space with FRET pairs to reduce the uncertainty associated to single pair FRET analysis.

To track the protein dynamics over seven orders of magnitude in time we applied our fluorescence spectroscopic toolkit with three complementary methods in the following order. At first we used high-precision ensemble TCSPC measurements to resolve the ensemble of conformational states which live longer than the fluorescence lifetime of the donor dye $\tau_D \approx 4$ ns. Next, we applied

confocal MFD to characterize the fluorescence bursts of single diffusing proteins and to check whether during the burst duration of a few milliseconds the ensembles are averaged out, which was the case for T4L. Finally, we probed the transitions in time range between nano- and milliseconds by fFCS which can analyze subpopulations in a mixture by using filters based on species specific fluorescence decays (Figure 4.1A and SI section 4.2.1.12). This is essential for unraveling the molecular kinetics and extracting rate constants of complex networks under equilibrium conditions. Thus, no sophisticated strategies to synchronize kinetics are needed.

**Detection of a third conformational state by eTCSPC**.

To analyze the conformational ensembles of free T4L with high time resolution and precision (see SI section 4.2.1.6), we performed eTCSPC measurements with small time bin width (14.1 ps/ch), a very high number of collected photons ($30 \times 10^6$ photons) and an additional instrument linearization procedure. We systematically tested our conclusions by statistical and global analysis using three differentfit models. To minimize the number of fit parameters, we analyzed the data set of each variant (i.e. the decay of the Donor-only- and DA-labeled sample) by a global fit (eq. 4.S13, 4.S14 and SI section 4.2.1.11), which allows us to determine the donor fluorescence lifetimes and to extract the additional donor quenching by FRET described by either one, two or three mean donor-acceptor distances $\langle R_{DA} \rangle$ in the DA sample. To include a small fraction ($\approx 0.1$) of locally quenched donor, we had to fit the fluorescence decays of all donors in the absence of FRET by two exponentials (Table 4.S1). Typical fluorescence decay curves of a DA- and the corresponding D-only-sample of the variant K19pAcF/R119C are shown in Fig. 2B. First, we tried to describe the additional FRET related quenching in the DA sample with a single $\langle R_{DA} \rangle$, which didn't even work for mutants (Table 4.S1B,C), where the AV simulations predict only a single distance for the two known conformers $C_1$ and $C_2$. Next we used a two-state model (Table 4.S1B). However, by looking to all data and checking its consistency as a single global set, clear features emerged indicating the need for a third state. All variants, including those that were designed as controls with a single $\langle R_{DA} \rangle$ for $C_1$ and $C_2$, needed at least two $\langle R_{DA} \rangle$. Moreover, the species fractions and the corresponding interdye distances of the 2-state model yielded an absolutely inconsistent assignment to the two known conformers (Fig. 4.S3).

Given the discrepancy on the two states model, we decided to fit all decays with the three states model (Table 4.S1C) and assigned the obtained mean distances $\langle R_{DA} \rangle$ to the known conformers $C_1$ and $C_2$ as predicted by AV simulations. The remaining distance was assigned to the new state $C_3$.

To test whether the improvement in $\chi^2_r$ of the three states model is statistically significant, we postulated the null hypothesis that a two states model would be sufficient to describe our data and calculated the F-value given by the ratio $\chi^2_{r-rel} = \chi^2_r(2\,states)/\chi^2_r(3\,states)$ for computing the corresponding p-values *p(2vs3)*. Plotting *p(2vs3)* against the interdye distance difference $\Delta R = \left| \left\langle R_{DA}^{(C1)} \right\rangle - \left\langle R_{DA}^{(C2)} \right\rangle \right|$ obtained by the three states model we recognize the existence of three classes of variants of distinct statistical significance values: (1) $C_1$ and $C_2$ aren't distinguishable (Fig 4.2C, black points); (2) $C_1$ and $C_2$ are distinguishable but the distance difference of $C_1$ or $C_2$ to $C_3$ is small ($\Delta R < 5$ Å, blue points); (3) all three conformers are well resolved by FRET ($\Delta R > 5$ Å, red points). Next we treated the whole data as a single global set and calculated the threshold of the F-value for our null hypothesis with a confidence level of 99 %. The obtained F-value 1.022 (SI section 4.2.2.1) is clearly smaller than the F-value of our data (1.046), thus we need at least a three states model to describe our data.

The newly found state $C_3$ can be considered a "hidden" state, because it has eluded identification based on other structural methods. In fact, $C_3$ has a population of 27 % ± 11 % in all variants (Table 4.S1C).

**Submillisecond conformational averaging detected by single-molecule MFD**

Single molecule experiments of freely diffusing molecules with burst analysis (see sections 4.2.1.8 and 4.2.1.10) provide direct evidence on the temporal evolution of the three found conformational states (Sisamakis et al, 2010). The two-dimensional frequency histogram of the FRET indicator fluorescence intensity ratio of the donor over the acceptor, $F_D/F_A$, against the fluorescence averaged donor lifetime in the presence of acceptor, $\langle \tau_{D(A)} \rangle_f$, shows the surprising result (Fig 4.2E), that the variant K19pAcF/R119C–(DA) exhibits only a single, slightly tailing population – apart from a small DOnly-fraction in the grey shaded area. With the pre-knowledge that three different, well- separated states are identified in ensemble conditions (Figure 4.2D), we would expect to resolve them in a similar form in our MFD plot. This apparent discrepancy can be rationalized by comparing the FRET indicator levels (colored lines) derived from eTCSPC with the 1D projections (Fig 4.2E) of the single-molecule burst analysis for both parameters $F_D/F_A$ (right) and $\langle \tau_{D(A)} \rangle_f$ (top). The values derived from eTCSPC measurements do not coincide with the peaks of the 1D projection but lie flanking to either side of the peaks. This is a hallmark for dynamic averaging at sub-millisecond timescales to merge into a tailing population during burst duration (~ 1 ms).
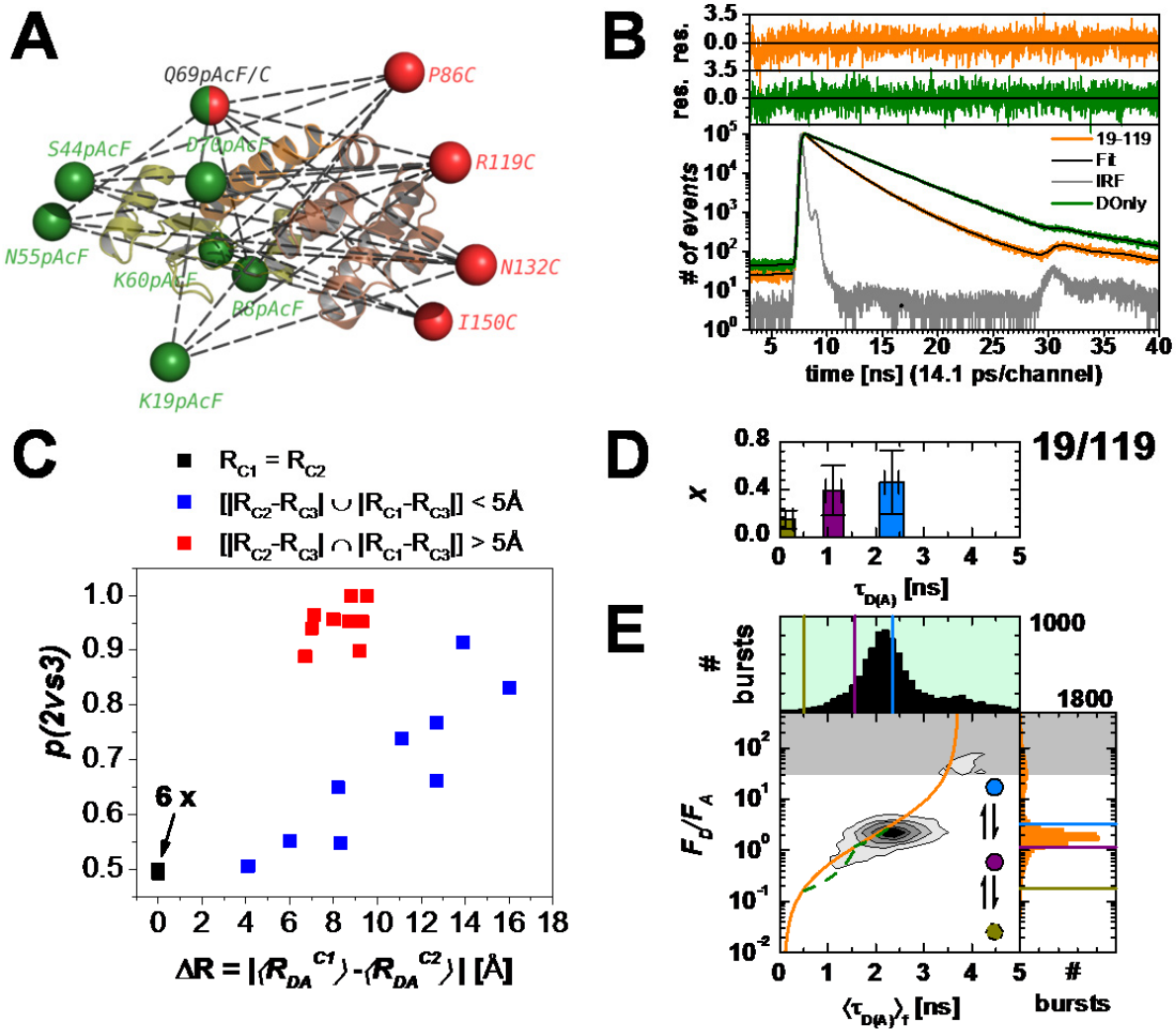
Figure 4.2 **Uncovering of a third conformational state** (A) Distance network of 24 DA FRET pairs. At donor positions (green), the inserted unnatural amino acid pAcF was labeled with Alexa488-hydroxylamine, the cysteine at the acceptor positions (red) was labeled with Alexa647-maleimide. Position 69 was used as both donor and acceptor position. (B) Time-resolved fluorescence decay for K19pAcF/R119C-(DA) (orange) respective DOnly-sample (green) fitted with three states model (black line), respectively, with the following values for $\langle R_{DA} \rangle_i$ and (species fraction $x_i$): 55.0 Å (0.46), 46.3 Å (0.39), 35.2 Å (0.15) (for more details see Table 4.S1C). Residuals are shown on top. (C) All variants were fitted with the two and three states model. Here, the p-value p(2vs3) calculated from the F-value $\chi^2_{r-rel} = \chi^2_r(2\,states)/\chi^2_r(3\,states)$ is plotted against the absolute difference in mean distances C1 and C2 as obtained from the three states model. For six variants, the mean distance of the states C1 and C2 cannot be distinguished (black, arrow). (D) Recalculated lifetimes of the variant K19pAcF/R119C-(DA) from eTCSPC data, error bars indicate uncertainty in fraction and lifetime. (E) MFD histogram of K19pAcF/R119C-(DA) labeled T4L. Two dimensional histogram $F_D/F_A$ vs. lifetime of donor in the presence of acceptor $\langle \tau_{D(A)} \rangle_f$. One dimensional projections for $F_D/F_A$ and $\langle \tau_{D(A)} \rangle_f$ are also shown. FRET line (Eq. (4.S11)) is shown in orange. Pure donor and acceptor fluorescence ($F_D$ and $F_A$) are corrected for background ($\langle B_G \rangle = 0.5$ kHz, $\langle B_R \rangle = 0.3$ kHz), spectral cross-talk ($\alpha = 1.7$ %) and detection efficiency ratio ($g_G/g_R = 0.75$). Shaded area in gray is the region of donor only. Horizontal guidelines for states $C_1$, $C_2$, and $C_3$ according to the eTCSPC results of the same sample are added as references. Ignoring the donor only population a single distribution is observed in all $F_D/F_A$ vs. $\langle \tau_{D(A)} \rangle_f$.

**Kinetic network of T4 Lysozyme conformational states resolved by filtered FCS**

To identify the timescales for the transitions between all observed conformational states and the minimal number of involved states, we use fFCS of the mutant S44pAcF/I150D-(DA). We chose this variant because all functional mutations were later added to this backbone. As a control, this particular variant was also reversely labeled in the -(AD) scheme where the acceptor is located on the N-terminal domain and the donor is located at the C-terminal domain. fFCS maximizes the contrast in fluctuation analysis by using filters (Fig 4.S4A-D) for species-specific fluorescence properties (fluorescence lifetime, spectral window, and fluorescence anisotropy), to provide unique crosscorrelation functions, sCCF (Fig 4.3A and Fig 4.S4E-I) (Felekyan et al, 2012; Felekyan et al, 2013). To calculate the sCCFs by fFCS it is required to select two species. However, we have three different populations. To solve this, we combined $C_1$ and $C_2$ into a single pseudo-species and cross-correlated with $C_3$.

Remarkably, when normalized, the sCCF of the –(DA) and –(AD) labeled T4L variants overlap extremely well showing two relaxation times at $t_{R1}$ = 4 μs, and $t_{R2}$ = 230 μs when globally fitting various mutants, indicative of the three exchanging states (Fig. 4.3A). Mathematically, the observed relaxation times in sCCF correspond to the eigenvalues of the kinetic matrix (SI Section 4.2.1.12 and 4.2.2.5, Eq. 4.S32). From the overlap of the two sCCF's we conclude that the location of the dyes does not alter the dynamics observed on the biomolecule. Besides, we see a similar behavior on the MFD diagrams of other variants (Fig. 4.S4E-I), which supports that three states interconvert a sub-millisecond timescale. Moreover we performed an analysis with time windows of 0.5 ms of the selected bursts to increase the time resolutions of the MFD diagrams. Particularly for the mutant S44pAcF/I150D-(DA) we observe a small but significant contribution of nearly static bursts (~5%) at the location of the $C_3$ state (Fig. 4.S7A). The fact that this population is visible shows two points about this state. First, observing bursts with integration time in the ms timescale means that the population is "fixed" in that particular state for at least the burst duration. Additionally, this shows that in fact the total contribution population of the $C_3$ states consists of a dynamic part $C_{3d}$ (overall 22% ± 10%) that exchanges with $C_2$ and $C_1$, as in the other mutants, but that on top we have a static fraction $C_{3t}$. Because we only observe two relaxation times from μs to ms, we ignore for the time being the existence of the 5% of $C_{3t}$, as it is not needed to discuss the connectivity between $C_1$, $C_2$ and $C_3$ at faster timescales.

Assuming a three-state system, there are only few possibilities ($C_1 \rightleftharpoons C_2 \rightleftharpoons C_3$; $C_2 \rightleftharpoons C_1 \rightleftharpoons C_3$, $C_1 \rightleftharpoons C_3 \rightleftharpoons C_2$; $C_1 \rightleftharpoons C_2 \rightleftharpoons C_3 \rightleftharpoons C_1$) to describe the connectivity of states. Considering the microscopic

reversibility of transitions we discriminate the full cycle, which could start in $C_1$, $C_2$ or $C_3$. Then, we are only left with three possibilities. Using dynamic FRET lines (SI section 4.2.1.18), which trace the connectivity between states, we clearly see that the transition $C_1$-$C_3$ is forbidden (Fig. 4.S5, 4.S7 and Fig. 4.2E), otherwise some bursts would follow the FRET line; particularly because the mean average FRET state ($F_D/F_A$) is somewhere between $C_1$ and $C_2$. Then, we are left with two options, for which we identify that either $C_1$-$C_2$ transition is fast or the transition between $C_2$-$C_3$ is fast. Considering the dynamic FRET lines the solution is straightforward. Mixing occurs between $C_1$ and $C_2$ with almost no elongation of the FRET population. Therefore, the fast transition occurs between $C_1$-$C_2$, while the exchange between $C_2$ and $C_3$ is slow. However, to be quantitative at this respect, we decided to analytically solve this kinetic problem by combining our results from eTCSPC and fFCS in a joint state matrix and conduct Brownian dynamics simulations (section 4.2.2.4). The state matrix, shown in SI section 4.2.2.4, has two possible analytical solutions (Fig 4.3C-D). We analyzed both solutions burst wise in various time windows (Fig 4.3C-D, Fig 4.S7) (Kalinin et al, 2010b) to take advantage of the various temporal domains of the transitions. If the transitions are in the regime of the time window, then a significant change on the distribution of bursts will be observed. By comparing both simulations ($C_1$-$C_2$ fast vs. $C_1$-$C_2$ slow) in a TW analysis of 0.5 ms, it is clear that only the first option is able to reproduce our data (Fig 4.3C-D, Fig 4.S7). The rate constants for the transition between $C_{3d}$ and $C_{3t}$ were semi-empirically selected such that their transition is slower than the diffusion of the molecules and also it clear and that this state is reached by transitioning from $C_{3d}$; otherwise we would observe a smear of bursts along the FRET lines from $C_{3t}$ to $C_1$-$C_2$.

Next we proceeded to do fFCS as done for the experimental data by selecting the mixture of $C_1$-$C_2$ and cross- correlated it with the species $C_3$. The sCCF shows almost identical behavior as the experimental data (Fig 4.3G). Also note that the slow transition between $C_{3d}$ and $C_{3t}$ is so slow that it is not observed on the sCCF, similar to what happened on the experimental data. Therefore, we conclude that we are able to reproduce our experimental data by Brownian dynamic simulations (SI section 4.2.1.14, 4.2.2.4).

**Detection of $C_3$ by EPR**

To look for an independent proof of the existence of $C_3$, we used EPR spectroscopy. We took the mutant S44pAcF/I150C and site specifically labeled it with the appropriate spin labels for Double Electron Electron Resonance (DEER) experiments (SI section 4.2.1.5). In those experiments the distribution of expected distances is very broad which means that there is a distribution of states between the open ($C_1$) and closed state ($C_2$). This observation corresponds to the expected hinge-

bending motion. In addition, the distance distribution clearly shows a side population with inter EPR label distance of ~ 35 Å. To make sure the side state is not an artifact of the Tikhonov regularization algorithm (Chiang et al, 2005a; Chiang et al, 2005b) or by rotamer populations, we decided to probe the distance distribution in other conditions. We observed that the population of the $C_3$ state increases at low pH (Fig. 4.3B), a resemblance of the acidic state found on the Hen Egg White Lysozyme (Klein-Seetharaman et al, 2002). Consequently, we did the same DEER experiment and showed that there is a redistribution of states towards a compact distance. Our FRET experiments show a similar redistribution of states at low pH. Thus, we concluded that in fact EPR and FRET show the compact state $C_3$. The full pH dependence of T4L is well beyond the scope of this paper, and it is left for a further study.

With independent verification of the new state, we decided to look deeper into the functional implication of the existence of $C_3$. To accomplish this goal, we mutate residues at the active site based on the backbone S44pAcF/I150C. We use this mutant because of the advantage to trap T4L in the state $C_{3t}$ as mentioned in the previous section. This trapped state is only observed in this backbone and on the Q69AcF/P86C-(DA). However, only this variant S44pAcF/I150C-(DA) shows three well-separated states. Based on this assumption, we expect to use the population of $C_{3t}$ as an indicator of the role of this new state during enzyme catalysis.

**Trapped chemical states of T4L**

The E11A mutation is known to inactivate T4L (Shoichet et al, 1995). Moreover, recently it was shown that the variant E11A could bind the substrate but it is unable to catalyze it (Hu & Lu, 2004). To make sure this statement is correct, we did intensity based FCS where the correlation time or diffusion component is slowed significantly in presence of purified peptidoglycan from *Micrococcus luteus* (Fig. 4.4C). Without substrate the correlation shows a typical diffusion time of 0.54 ms that agrees very well with the hydrodynamic radius of ~2 nm for the T4L (Durant et al, 2002). Single molecule FRET experiments of E11A/S44C/I150C-(DA) in presence of its substrate showed a relative shift on kinetics toward the $C_2$ state compared to the free enzyme (Fig. 4.4D-E). Note that the conformational state $C_{3t}$ was not observed in this case. Consequently, most likely the conformation $C_3$ is not relevant for substrate binding and dissociation of substrate happens stochastically by the opening due to the hinge bending dynamics without populating $C_{3t}$. Moreover, additional evidence of binding is observed by the shift towards larger anisotropy values upon incubation with substrate (Fig 4.S4).
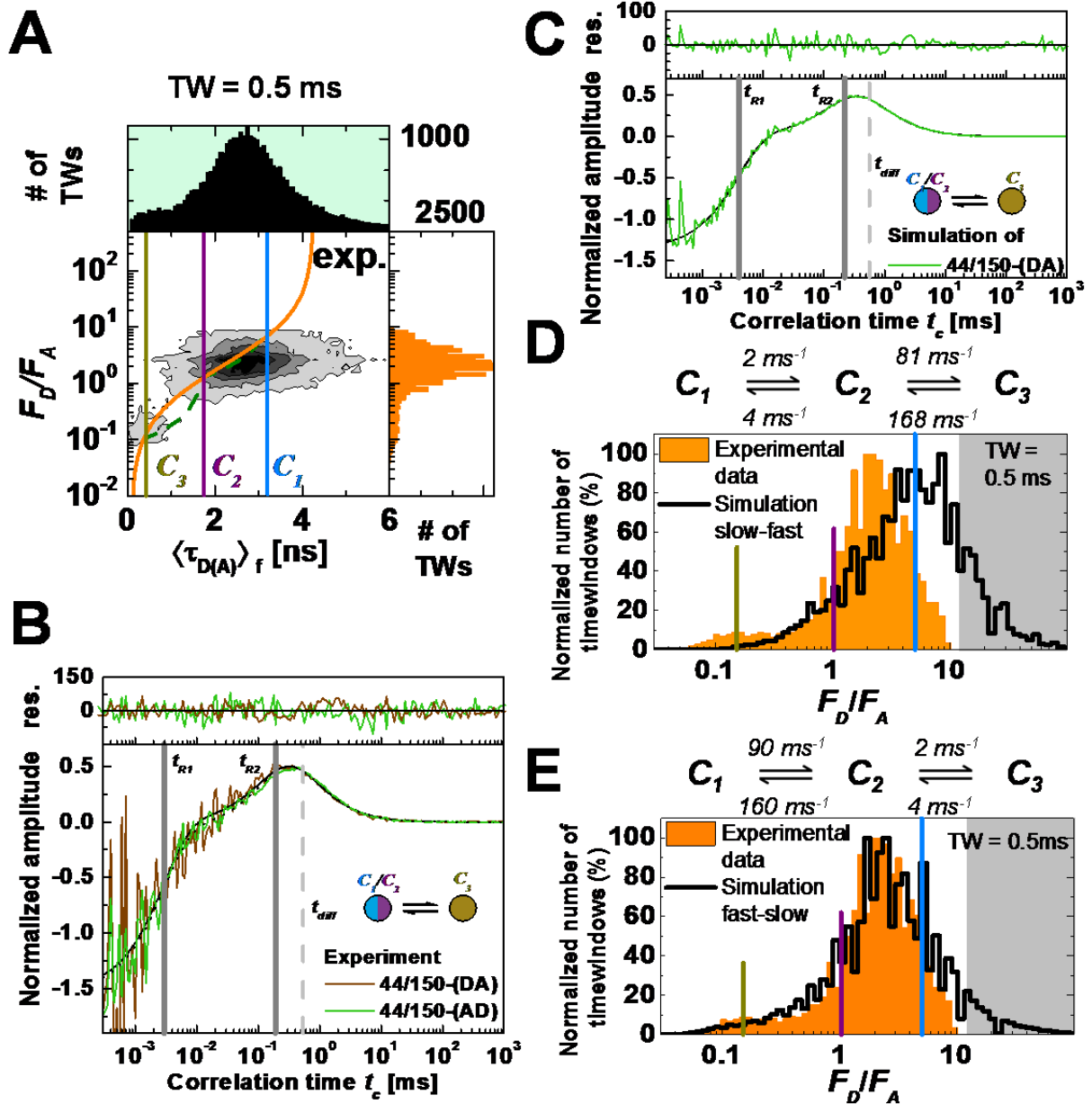
Figure 4.3 **Kinetic connectivity** (**A**) MFD histogram of S44pAcF/I150C-(DA) labeled T4L analyzed in 0.5 ms time-windows. Two dimensional histogram $F_D/F_A$ vs. lifetime of donor in the presence of acceptor $\langle\tau_{D(A)}\rangle_f$. One dimensional projections for $F_D/F_A$ and $\langle\tau_{D(A)}\rangle_f$ are also shown. FRET line (Eq. (4.S11)) is shown in orange. Pure donor and acceptor fluorescence ($F_D$ and $F_A$) are corrected for background ($\langle B_G\rangle$ = 1.6 kHz, $\langle B_R\rangle$ = 0.8 kHz), spectral cross-talk ($\alpha$ = 1.2 %) and detection efficiency ratio ($g_G/g_R$ = 0.77). Shaded area in gray is the region of donor only. Horizontal guidelines for states $C_1$, $C_2$, and $C_3$ according to the eTCSPC results of the same sample are added as references (**B**) Overlay of the normalized sCC functions (Eq. 4.S23) of S44pAcF/I150C-(DA) and S44pAcF/I150C-(AD). The Global fit with other mutants shows two relaxation times ($t_{R1}$ = 4.0 ± 2.4 µs, $t_{R2}$ = 230 ± 28 µs) and diffusion time $t_{diff}$ = 0.54 ms. (C) sCCF between the pseudo-states consisting of the $C_1/C_2$ mix and the $C_3$ for simulated data. Fit of this sCCF curve returns two relaxation times of 4 µs and 220 µs, consistent with our input parameters.(**D,E**) Overlay of time window analysis, fast-slow in A and slow-fast in B. Details concerning the model and the simulation can be found in SI Section 4.2.2.4

57
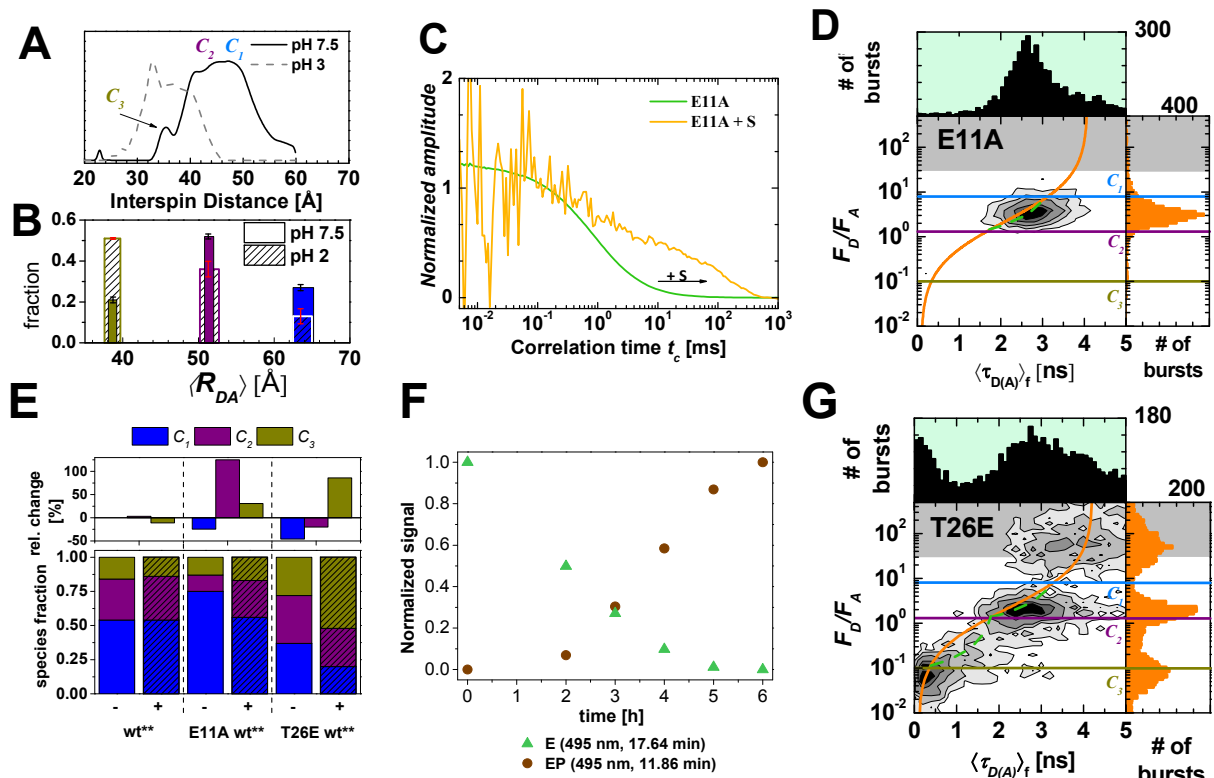
Figure 4.4 **T4 Lysozyme functional states** (**A, B**) Normalized interspin distance distributions for S44(K1)/I150(R1), T26E/S44(K1)/I150(R1) and FRET distance distributions for S44pAcF/I150C-(DA) from eTCSPC. Due to linker effects the interlabel distances are not directly comparable. The shoulder, indicated by the arrow, of the distribution at short distance agrees with the short interdye distance of state $C_3$. For eTCSPC data, the width of the bar represents the uncertainty in distance (empty bars: pH 7.5, hatched bars: pH 2). The statistical uncertainty in the amplitudes is indicated by error bars (black: pH 7.5, red: pH 2). The state $C_3$ increases in population at low pH. (**C, D**) Effects of the substrate on E11A/S44C/I150C-(DA). (**C**) Overlay of normalized sCCF of E11A/S44C/I150C-(DA) with and without substrate. Consistent with the larger rotational correlation, we observe a shift of $t_{diff}$ toward longer times for the variant E11A/S44C/I150C-(DA) when incubated with the substrate. (**D**) MFD histograms for variant E11A/S44C/I150C-(DA) with substrate. Upon addition of the substrate we observe a shift toward lower $F_D/F_A$ values. (E) Species-fractions of the variants S44pAcF/I150C-(DA), E11A/S44C/I150C-(DA) and T26E/S44pAcF/I150C-(DA) used to mimic free enzyme (E), enzyme-substrate complex (ES) and enzyme-product bound state (EP). Upon addition of peptidoglycan (+), the fractions of states of the functional variants are shifted towards $C_2$ and $C_3$, respectively (shaded bars). On top, the relative change in fractions upon addition of peptidoglycan is shown. (**F, G**) Effects of the substrate on T26E/S44pAcF/I150C-(DA). (**F**) T4L lysozyme binding to peptidoglycan as observed by reverse phase HPLC. Samples were taken at different incubation times of the T26E variant with peptidoglycan and separated using C-18 reverse phase HPLC. Elution for free enzyme (E) and product-bound (EP) was monitored at 496 nm which corresponds to the absorbance of Alexa488. Plotted is the normalized signal at 17.64 min (E) and at 11.86 min (EP) after background subtraction. (**G**) Two dimensional $F_D/F_A$ vs. $\langle \tau_{D(A)} \rangle_f$ histograms of variant T26E/S44pAcF/I150C-(DA) incubated with peptidoglycan. An accumulation of the high FRET states ($C_{3d}/C_{3t}$) is observed.

The next designed variant was the T26E which is the mutant used in the crystallization of the adduct form of T4L (Kuroki et al, 1993). This variant is able to cleave its substrate but is unable to release the product. To confirm the catalysis we used HPLC and monitored the elution of the free and bound form of labeled T4L (T26E/S44pAcF/I150C-(DA) variant) at various incubation times (Fig. 4.4F). We noticed that the mutated enzyme (T26E) elutes earlier from the column after incubation with substrate indicative of the adduct form of T4L (Fig. 4.4F). Mass spectroscopy confirmed the increased molecular weight change of 940 Da (Data not shown). The T26E/S44pAcF/I150C-(DA) mutant was then incubated in single molecule conditions in similar conditions as the HPLC experiments. We observe a significant increase on the population of $C_{3t}$. eTCSPC also showed a significant increase on the population of $C_3$. However, in EPR (Fig. 4.4A) the T26E mutant mostly populates $C_2$. Note that the EPR labels are much smaller than the FRET labels and most likely the trapped mechanism is an effect of the long dyes that act as a latch. Nevertheless, in EPR $C_3$ is still observed to a larger extent in the T26E/S44pAcF/I150C than in the S44pAcF/I150C alone. This is consistent with the FRET measurement. The difference is that $C_{3t}$ overwhelms the response in the FRET data due to the $C_{3t}$ mechanism. What is clear is that the state $C_3$ is populated in greater extent in the T26E mutant for both the EPR and FRET data. This is the first hint that the new state was in fact involved in the catalytic cycle of T4L. Contrary to what has been indicated in literature about the T4L, where a two-state model with a single hinge-bending motion would suffice to describe its function, it is clear that there is a need of a 3$^{rd}$ state.

**Structural features of conformational states**

To date, more than 580 different structural models of T4L have been deposited in the PDB; many of those contain single or several mutations, helix insertions or deletions or are fused to G-coupled receptors. We used FPS to screen 409 of those PDB-models with the obtained FRET-restraints in order to identify the structures that best describe our experimentally determined conformational states. The PDB-models with severe modifications were not included in the screening. FPS is used to *in silico* label a known structure in order to compute an error function

$$\chi^2_{r,FPS}{}^{(j)} = \frac{1}{N}\sum_{i=1}^{N} \frac{(\langle R_{DA}\rangle^{(i)}_{model,j} - \langle R_{DA}\rangle^{(i)}_{experiment})^2}{\Delta R^2_{DA}(\kappa^2, k_{FRET})^{(i)}} ,$$ where each measured distance $\langle R_{DA}\rangle^{(i)}_{experiment}$ is

compared to the modeled distance $\langle R_{DA}\rangle^{(i)}_{model}$ and weighted by the experimental uncertainties $\Delta R^2_{DA}(\kappa^2, k_{FRET})^{(i)}$ over the number of measurements ($i$=1...N) for each conformational state (j) (SI section 4.2.1.15).

We used this error function to screen a 409x409 matrix of the structures in order to simultaneously satisfy states $C_1$ and $C_2$. In total 167,281 structure combinations were computed. Finally, we ordered the result according to lowest $\chi^2_{r,FPS}$. $\chi^2_{r,FPS}$ shows a funnel towards the structure that best represent the experimental states (Fig. 4.5A). The lowest $\chi^2_{r,FPS}$ is ~ 0.85 and corresponds to PDB ID structure 172L for the state $C_1$ while $C_2$ can be described best by PDB ID 1C67 (Quillin et al, 2000) – a T4L variant containing two alanine mutations inside the (hydrophobic) core of the CTD and that was crystallized with Krypton bound. An alignment of both crystal structures shows the expected result; in 1C67 the N- and C-terminal lobes are closer together as compared to 172L. Also, the structure 1C67 is only 0.7 Å RMSD (Root Mean Squared Deviation) of the known adduct state PDB ID 148L.

To judge the goodness of the screening, we plot the modeled distances against the experimentally determined distances (Fig. 4.5B). In the ideal case, the black curve has a slope of 1. The linear regression, shown in red, is very close to this idealized case of slope 1.

In the case of state $C_3$, we performed a separate single screening, not to bias the result for the $C_1$ and $C_2$ states due to the much worse $\chi^2_{r,FPS}$ values for $C_3$. The best screened structure found for $C_3$ is PDB ID 3C7Y (Mooers et al, 2009), which contains only a single mutation at position 96 (R96A) and displays a $\chi^2_{r,FPS}$ of 4.3. This is significantly worse than the minimum $\chi^2_{r,FPS}$ found for $C_1$ and $C_2$. This is clearly visible by the anti-correlated behavior of simulated and measured FRET distances (Fig 4.5B). This result implies that there is no structure currently deposited on the PDB that describes the newly found $C_3$.
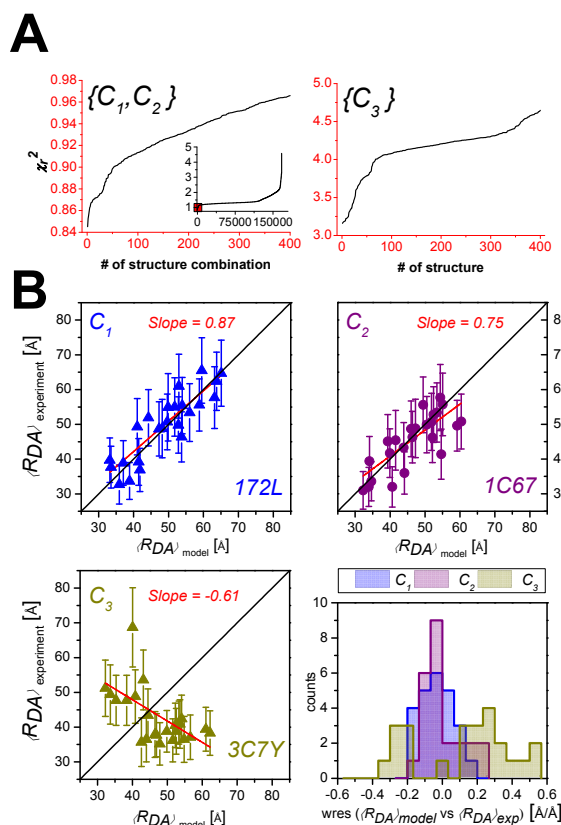
Figure 4.5 **PDB screening** (**A**) A global screening containing the determined distances for $C_1$ and $C_2$ was performed against 409 out of 583 total structures deposited in the PDB for T4L. The resulting 409x409 structural combinations were sorted according to an increase in reduced $\chi^2_{r,FPS}$. The inset shows the development of $\chi^2_{r,FPS}$ of all 167'281 possible structure combinations. (**B**) The distances from the PBD structure describing our experimentally determined distances set best are plotted against the experimentally determined distances and fitted linearly (red line). Best structures are: $C_1$ 172L, $C_2$ 1C67, $C_3$ 3C7Y. $C_3$ was screened single, not to bias the result of the combined $C_1/C_2$ screening. Black lines have a slope of 1. Weighted residuals for the difference between model and experiment as histogram are shown in the right lower panel.

## 4.1.3. Discussion

**Determination of $C_3$ and its structural properties**

Additional supporting data that adds to the already overwhelming conclusion of the existence of $C_3$ comes from few specific variants. For example, the FRET variant Q69pAcPhe/P86C-(DA) is especially informative to understand how T4L can generate such a compact state ($C_3$). Here the donor is placed in the middle of helix c (Orange Fig. 4.1B) connecting both domains. The acceptor is located in the middle of helix d, which is part of the CTD (Brown Fig. 1B). According to the AV calculations in this configuration $C_1$ (PDB ID 172L) and $C_2$ (PDB ID 148L) are hardly distinguishable by FRET. If a new state would be observed, it would require for the interdye distance to be larger than the distances for $C_1$ and $C_2$. FRET exactly confirms the hypothesis (Table 4.S3). This is due to the extension of the linkers and the angles expected to occur as the two subdomains approach each other. Assuming that both domains preserve their secondary structure, the observation that $C_3$ has the longest distance can only be explained by a kink in helix c between amino acids 69 and 86. The helix breaker Gly 77 is the perfect candidate for introducing such a kink. Previous mutagenesis studies revealed enhanced protein thermostability for the G77A variant (Matthews et al, 1987). Introducing the G77A mutation into the S44pAcPhe/I150C-(DA) variant showed that the $C_3$ state is partially reached with a separation distance of 42.6±4.4 Å compared to 38.8±1.5 Å in the double mutant (Fig. 4.S3/S4).

It is known that salt bridges could also contribute to the hinge bending and to its function. To test this we used the inactive mutant R137E. When this particular mutation is added to the backbone S44pAcPhe/I150C-(DA) we do not see the trapped state ($C_{3t}$), a phenotype also observed on the inactive mutant E11A/S44C/I150C-(DA).

Key to the analysis and determination of $C_3$ is the use of global analysis of time-resolved fluorescence decays. It was important to recognize that oversampling compensates the variability and uncertainty associated with a single pair FRET experiments. Proper statistics and model comparison, allowed us to build the confidence and assure the existence of $C_3$. Ignoring the existence of $C_3$ and relying only on two states ($C_1$ and $C_2$) would results in nonsense screening results (Fig. 4.S2), because the obtained distances do not provide a good enough value for $\chi^2_{r,FPS}$ and there is not a linear correlation between modeled and experimental interdye distances.

**Relationship between conformational state and its function**

We find that $C_3$ is preferentially populated after substrate hydrolysis. However, its exact function is yet to be determined. Using the relative changes observed by the population of $C_3$ across the functional mutants, ignoring the trapped state $C_{3t}$, and only considering its dynamic part, we build a state matrix that relates the conformational state with the designed enzymatic state (Fig. 4.6). For the free enzyme $E$ we use the variant S44pAcPhe/I150C-(DA), for the substrate bound enzyme $ES$ we use the results from the E11A/S44C/I150C-(DA) in the presences of excess substrate. And for the product state $EP$, we use the adduct variant T26E/S44pAcPhe/I150C-(DA) after substrate hydrolysis. Nonetheless, one cannot compare them directly because each of these variants in the absence of substrate shows slightly different equilibrium conditions (Fig. 4.4E). Thus, we apply the relative change in the presence of substrate to look at changes on the energy landscape in the enzymatic states. The areas of the numbers on the matrix (Fig. 4.6A) are coded to represent the relative population fractions. Reaction rate constants are obtained similarly as done before in Fig. 4.3. Using Arrhenius formalism we computed the relative energy landscape (Fig. 4.6C). We use a Gaussian distribution for each state; the width of the distribution (2-3 Å) is adjusted to satisfy the activation energies. We can observe some very well-known characteristics of T4L and some new insights on what could be a possible mechanistic use of $C_3$. It is obvious that in the presence of substrate, $ES$, the most favored state is $C_2$; something that was identified three decades ago. After catalysis, in the $EP$ state, also the most probable state is $C_2$. However, if we take into consideration the contribution of the latch mechanism ($C_{3t}$) we can observe that we could lock T4L in this compact state. This is only possible by placing the fluorescent dyes in that particular location. The state $C_{3t}$ only shows that there is a need of the enzyme to pass through a more compact state after catalysis, and not so upon substrate binding. All our evidence indicates that the conformational state $C_3$, which is much more compact than any other structure known of T4L, is needed after catalysis and not before. Note that even in saturated conditions of $ES$ and $EP$ states, the enzyme is still in dynamic equilibrium and it exchanges between conformational states. This view departs for our rigid view of enzymes that are locked in particular states; however, this effect is difficult to observe.

To date the mechanistic view of T4L presumably starts by a proton that is donated by the Glutamic acid in position 11 (E11) to the oxygen of the glycosidic bond of the substrate. In a single displacement reaction, water molecules that are hydrogen bonded to Aspartic Acid at position 20 (D20) and Threonine in position 26 (T26) can attack the C-1 carbon of the substrate, inverting the substrate configuration to give the product. Following catalysis, the proton is presumably restored from D20 to E11 through solvent transfer (Kuroki et al, 1995). For this to occur no significant

structural changes can occur on T4L; otherwise the proton transfer would be lost over larger distances. However, what happens after hydrolysis is what makes $C_3$ important. Considering the extended Michaelis-Menten like formalism (Kou et al, 2005) (Fig. 4.6E) where the substrate $S$ binds reversibly with an enzyme $E$ to form an enzyme-substrate complex $ES$. $ES$ undergoes a transition to the excited state $E^*$ in order to actively spill out the product $P$ through an intermediate step $EP$, where the enzyme is still bound to the product, similarly to the case of T4L. Finally, the enzyme is regenerated to the original enzyme $E$ from $E^*$. In this case, the rate constants for the reaction are $k_1$, $k_{-1}$, $k_2$, $k_{-2}$, $k_3$, $k_{-3}$, $k_4$, $k_{-4}$, and $\delta$. From this model we identify that each observed conformational state ($C_1$, $C_2$ and $C_3$) is linked to each fundamental enzyme state $E$, $ES$, and $EP$ during catalysis. Substrate binds to the open conformation $C_1$ and links states $E$ to $ES$, followed by the cleavage of the glycosidic bond in the closed conformation $C_2$ linking state $ES$ to $EP$, and the product release from the compact conformation $C_3$ that links state $EP$ to $E$. The last step closes the enzymatic cycle and leads to a regeneration of the original enzyme $E$ from $EP$ via an excited conformational state $E^*$, where the compact state $C_3$ would correspond to the conformation of $E^*$ in order to spill out the product. This must not to be confused with the $EP$ state, although we use the T26E adduct form to trap the $E^*$ based on the longer linkers and their location. This mechanism can be an evolutionary advantage to release the product when fast catalysis is required for function even at room temperature. On the contrary, considering a system with only two conformational states and without an active cleaning mechanism, stochastic dissociation of the product can become rate limiting if the affinity of the product to the enzyme in the $EP$ state is high or comparable to substrate affinity. Indeed, the *in vivo* conditions for T4L are such that there is always a large excess of substrate. Thus, decoupling substrate access and product release via a three-state system can overcome substrate inhibition that occurs in a two-state system when the route to the active site is clogged by excess substrate concentrations (Gohlke et al, 2013). In view of the modified Michaelis-Menten mechanism of enzyme kinetics, our results stress the importance of considering also short-lived functional states in kinetic descriptions. It waits to be seen whether such compaction and motions are common for other proteins with comparable subdomain architecture.
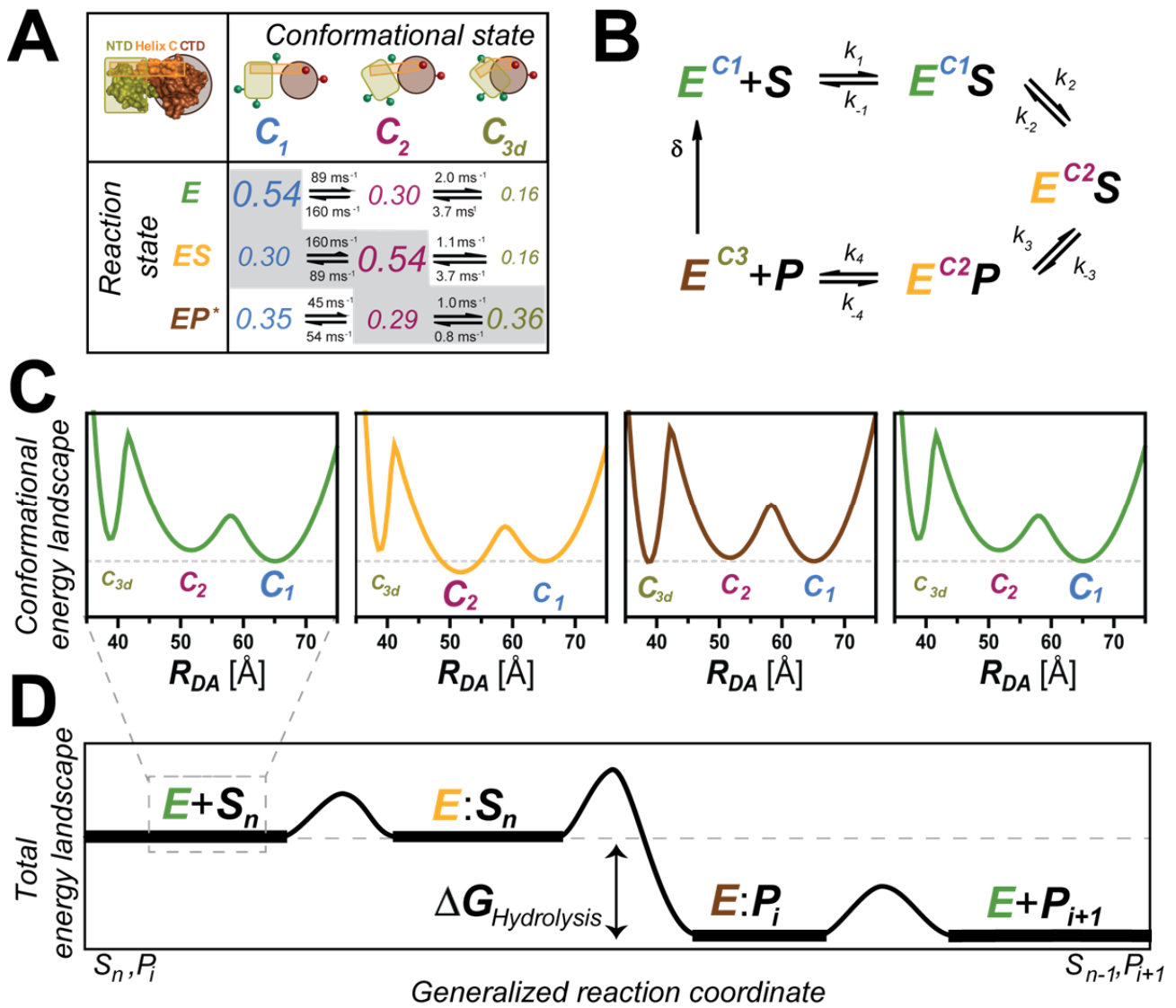
Figure 4.6 **Energy landscape of T4L.** (**A**) T4L interconverts between three major conformational states, $C_1$, $C_2$, and $C_3$. Population fractions of $C_1$, $C_2$, and $C_3$ are normalized to the variant S44AcF/I150C-(DA) in absence of substrate using the relative changes in Fig. 4.4E to correct for influence on specific mutations in the absence of substrate. The different font sizes represent the species fractions $x_i$ for each conformer according to Table 4.S4 and satisfy $x_1+x_2+x_{3d}=1$. The three enzyme states were mimicked by three enzyme variants: (i) the free enzyme state $E$ by S44AcF/I150C-(DA); (ii) the enzyme-substrate state $ES$ by E11A/S44C/I150C-(DA); (iii) the enzyme product state $EP$ by T26E/S44AcF/I150C-(DA) mutant after substrate cleavage. * For the sake of clarity the population of the state $C_{3t}$ is not displayed, only for the $EP$ state. Reaction rates constants were calculated according to Fig. 4.3B. Confidence intervals on the reaction rate constants are shown in Table 4.S4. (**B**) The catalytic activity of enzymes can be understood using a Michaelis-Menten like mechanism (Kou et al, 2005): a substrate $S$ binds reversibly with an enzyme $E^{C1}$ to form an enzyme-substrate complex $E^{C1}S$. The rearrangement of the protein toward the closed conformation $E^{C2}S$ is closely followed by the hydrolysis and the transition to $E^{C2}P$ state. Release of the cleaved half of the product allows the protein to evolve into the excited state $E^{C3}$ in order to release the product $P$. Finally, the

enzyme is regenerated to the original enzyme $E^{C1}$ from $E^{C3}$. The rate constants for the reaction are $k_1$, $k_{-1}$, $k_2$, $k_{-2}$, $k_3$, $k_{-3}$, $k_4$, $k_{-4}$, and $\delta$. (**C,D**) Generalized reactions coordinate of T4L. Relative Gibbs free energy landscapes were calculated using $\Delta G^{\circ} = -k_B T \ln\left(\dfrac{k_{ji}}{k_{ij}}\right)$, where $k_B$ is the Boltzmann constant and $T$ is the temperature; $k_{ij}$ are the reaction rate constants between states $C_i$ and $C_j$ for the data presented in panel A. The activation energies were calculated according to $\Delta G^{\circ+} = -k_B T \ln\left(\dfrac{k_{ji}}{k_0}\right)$ assuming $k_0 = 10^3$ ms$^{-1}$ as an arbitrary constant. Solid line distributions consider $C_1$, $C_2$, and $C_{3d}$ to follow a Gaussian distribution as a function of the interdye distance $R_{DA}$. Gaussian widths ($\sigma_t$) were adjusted to satisfy the energy differences and calculated activation energies. The contribution of $C_{3t}$ is only shown for the condition $EP$ with a dashed line where $C_3 = C_{3t} + C_{3d}$.

Moreover, some of the structural features yet to be determined might show that indeed the product could be released via this excited state $C_3$ or $E^*$. It is our next challenge, to determine the structure of this hidden state to validate the possibility of this product release step based on its three-dimensional structure. This could be possible using the FRET restraints already acquired, but significant developments on the field of structure determination based on FRET measurements are needed. What is clear is that there is a hidden structure of T4L yet to be determined.

## 4.1.4 Materials and Methods

**Samples**

Mutagenesis of the pseudo-wild-type (WT*) with the mutations C54T and C97A resulting in 24 T4L mutants was carried out as previously described (Fleissner et al, 2009). Site specific labeling of T4L was accomplished using orthogonal chemistry for FRET and EPR with the keto function of the p-acetyl-L-phenylalanine (pAcF) and cysteine (C). For FRET Alexa 488 was used as donor and Alexa647 as acceptor dye with a Förster Radius $R_0 = 52$ Å. All FRET measurements were performed at room temperature using an aqueous buffer (50 mM sodium phosphate, pH 7.5, 150 mM NaCl). For single-molecule measurements we added 40 μM TROLOX to minimize acceptor blinking and 1 μM unlabeled T4L to prevent adsorption to the cover glass. fFCS were performed in a custom-built confocal microscope with a dead time-free detection scheme using 8 detectors (four green (τ-SPAD, PicoQuant, Germany) and four red channels (APD SPCM-AQR-14, Perkin Elmer, Germany)). The detected photon counts were registered by a customized time-correlated single photon counting (TCSPC) module with 8 synchronized input channels (Felekyan et al, 2005) (HydraHarp 400, PicoQuant, Germany) saving the data in the Time-Tagged Time-Resolved (TTTR) mode. Double electron electron resonances were measured on a Bruker Elexys 580 spectrometer as shown

previously (Fleissner et al, 2009). FRET positioning and screening was done according to Kalinin *et al*. (Kalinin et al, 2012). More detailed description of all materials and methods is given in SI (section 4.2.1).

**Simulation of interdye distances**

In FPS, each fluorophore is coarse-grained where the chemical linker is modeled as a cylinder with radial width (*w:* 4.5 Å and 3.5 Å for Alexa 488 and Alexa 647 respectively) and length (*l:* 20 Å and 22 Å for Alexa 488 and Alexa 647 respectively), and chromophores are modeled as three spheres with determined radii (donor and acceptor radii $R_1$: 5 Å and 11 Å $R_2$: 4.5 Å and 3 Å $R_3$: 1.5 Å for both) (Sindbert et al, 2011a). All possible positions of the center of mass of each fluorophore are simulated using a spring like model allowing the fluorophore to sample all conformational space restricted only by steric clashes with the surface of the biomolecule. The result is that each dye forms a "cloud" with all possible spatial coordinates where the center of mass of the dye could be located; we call this the accessible volume or AV (SI section 4.2.1.13). From all donor and acceptor positions ($\overrightarrow{R_D^{(i)}}$ and $\overrightarrow{R_A^{(j)}}$) the interdye distance $\langle R_{DA} \rangle$, measured by TCSPC, is the average

$$\langle R_{DA} \rangle = \left| \left\langle \overrightarrow{R_D^{(i)}} - \overrightarrow{R_A^{(j)}} \right\rangle \right| = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \left| \overrightarrow{R_D^{(i)}} - \overrightarrow{R_A^{(j)}} \right|.$$

**Controls for FRET**

Most of the potential problems with smFRET come from the complexities associated with the labels. We did ten controls to check for potential label artifacts. Additional data is found in Supporting Information section 4.2 (SI).

1. The labeling does not alter enzyme function. Labelled T26E mutant shows expected adduct formation (Fig. 4.3/4.S5).
2. Local quenching of donor dye is considered when calculating distances and cross correlations (Table 4.S1-S3).
3. Triplet state quenchers do not affect the observed relaxation times and amplitudes on the species cross-correlation (Fig. 4.S7).
4. Acceptor cis-trans isomerization does not contribute to the signal on the species correlation analysis (Fig. 4.S7).

5. $\kappa^2$ distributions show that our assumption of $\kappa^2 = 2/3$ is valid (Table 4.S5). Table 4.S5 summarize the residual anisotropies ($r_\infty$) of D - donor, A - acceptor and A(D) by the FRET sensitized emission of acceptor used for calculating $\kappa^2$ distributions (Sisamakis et al, 2010).

6. The existence and the population fraction of the new conformational state $C_3$ is consistent across our library of mutants (Table 4.S2/S3), $x_3 = 28 \pm 10$ %. The statistical uncertainty of this species fraction of all FRET wt* variants determined from fluorescence decay analysis is about 2.2 %. The larger variation of the experimental uncertainty is consistent with common knowledge that mutations slightly affect the conformational stability of T4L, which was measured for example in chemical denaturation experiments. Therefore, we attribute this slightly increased variability of population fractions to mutation effects.

7. All 24 variants provide a consistent view of T4L conformational states and after screening we find consistency of two known structured determined by X-ray crystallography of T4L without outliers (Fig. 4.5, Table 4.S3).

8. We oversample FRET restraints to reduce uncertainty introduced by each point mutation (Fig. 4.5, Table 4.S3).

9. Thermodynamic stability and proper folding of our variants were verified by chemical denaturation with urea and by measuring CD spectra for unlabeled and labeled T4L (Data not shown).

10. We fit time resolved fluorescence decays with various models providing a consistent view of the conformational space (SI section 4.2.1.11).

**Author contributions:** H.S. and K.H. purified and labeled the protein. H.S., T.P., K.H., and D.R. measured and analyzed the FRET experiments. T.P., M.D. and H.G. performed structural screening. M.R.F. performed and analyzed EPR experiments. W.H. performed study design and EPR analysis. S.F. developed fluorescence analysis tools. F.K. and R.K. developed fluorescence instrumentation and gave technical support. All authors discussed the results and commented on the manuscript. H.S., W.H. and C.A.M.S. wrote the paper. C.A.M.S. supervised the project.

**Author Information:** H.S. current address: Clemson University, SC, U.S.A.; M.R.F. current address: Boston University, MA, U.S.A. Correspondence and requests should be addressed to C.A.M.S. (cseidel@hhu.de).

4.2 T4 lysozyme manuscript supporting information

# Dynamics and function of transient states in single enzyme molecules

**Authors:** Hugo Sanabria[1,2*], Dmitro Rodnin[1], Katherina Hemmen[1], Thomas Peulen[1], Suren Felekyan[1], Mark R Fleissner[3]†, Mykola Dimura[1,5], Felix Koberling[4], Ralf Kühnemuth[1], Wayne Hubbell[3], Holger Gohlke[5], Claus A.M. Seidel[1]*

# Affiliations

[1]Institut für Physikalische Chemie, Lehrstuhl für Molekulare Physikalische Chemie, Heinrich-Heine-Universität, Düsseldorf, Germany.

[2]Department of Physics and Astronomy, Clemson University, Clemson, South Carolina, U.S.A.

[3]Jules Stein Eye Institute and Department of Chemistry and Biochemistry, University of California, Los Angeles, U.S.A.

[4]PicoQuant GmbH, Berlin, Germany.

[5]Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität, Düsseldorf, Germany.

*Correspondence to: cseidel@hhu.de

## 4.2.1. Materials and Methods

## 4.2.1.1 Experimental design

We use a hybrid approach that combine single-molecule FRET (Förster Resonance Energy Transfer) using a MFD (Multiparameter Fluorescence Detection) system, ensemble TCSPC (Time Correlated Single Photon Counting), mutagenesis to build a network of FRET labeled samples and correlation analysis using filtered FCS (Fluorescence Correlation Spectroscopy). To resolve the kinetic network we used dynamic photon distribution analysis (PDA) of bursts with variable time windows and the analytical solution of a three-state kinetic network. All these methods and how they were used are explained. Their interaction is better illustrated in Fig. 4.S1.
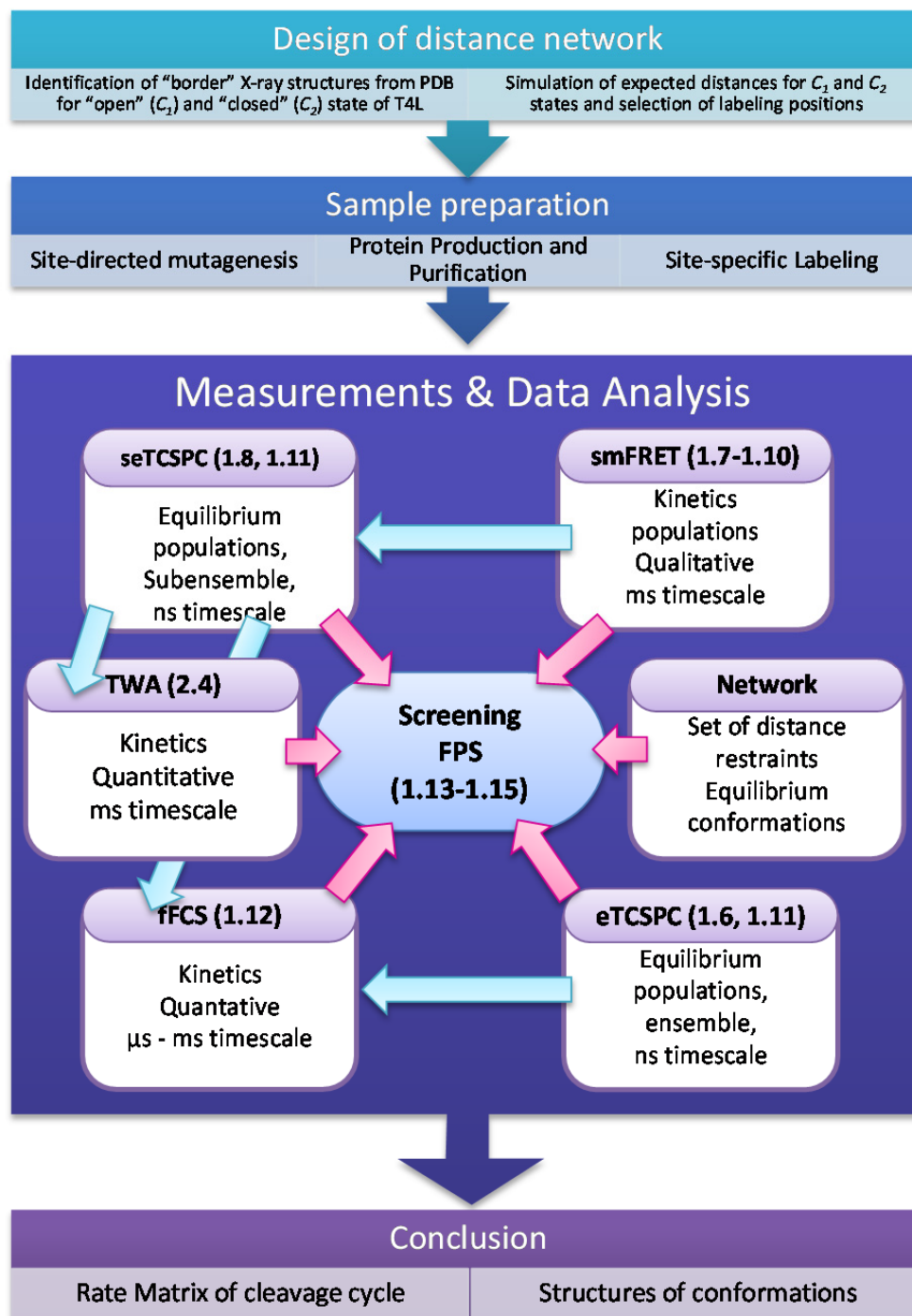
Figure 4.S1 **Flowchart of data acquisition and data analysis.** After identification of border X-ray structures, a distance network spanning the whole T4 lysozyme is built. T4L variants are generated via site-directed mutagenesis, produced in *E. coli*, purified and site-specifically labeled. Of all variants the following experiments are conducted: ensemble time correlated single photon counting (eTCPSC), multiparameter fluorescence detection (MFD) and filtered fluorescence correlation spectroscopy (fFCS). Of single molecule data, population-specific data can be cut and sub-ensemble TCSPC (seTCPSC) can be done. Time-window analysis (TWA) is used to discriminate between different kinetic models.

## 4.2.1.2 T4 Lysozyme purification and site specific mutation

T4L cysteine and amber (TAG) mutants were generated via site directed mutagenesis as previously described in the pseudo-wild-type containing the mutations C54T and C97A (WT*) and subsequently cloned into the pET11a vector (Life Technologies Corp.)(Brustad et al, 2008; Fleissner et al, 2009; Lemke, 2011). The plasmid containing the gene with the desired mutant was co-transformed with pEVOL(Lemke, 2011) into BL21(DE3) *E. coli* (Life Technologies Corp.) and plated onto LB- agar plates supplemented with the respective antibiotics, ampicillin and chloramphenicol. A single colony was inoculated into 100 mL of LB medium containing the above-mentioned antibiotics and grown overnight at 37 °C in a shaking incubator. 50 mL of the overnight culture were used to inoculate 1 L of LB medium supplemented with the respective antibiotics and 0.4 g/L of pAcF (SynChem) and grown at 37°C until an $OD_{600}$ of 0.5 was reached. The protein production was induced for 6 hours by addition of 1 mM IPTG and 4 g/L of arabinose.

The cells were harvested, lysed in 50 mM HEPES, 1 mM EDTA, and 5 mM DTT pH 7.5 and purified using a monoS 5/5 column (GE Healthcare) with an eluting gradient from 0 to 1 M NaCl according to standard procedures. High-molecular weight impurities were removed by passing the eluted protein through a 30 kDa Amicon concentrator (Millipore), followed by subsequent concentration and buffer exchange to 50 mM PB, 150 mM NaCl pH 7.5 of the protein flow through with a 10 kDa Amicon concentrator. For the double cysteine mutant containing E11A, the temperature was reduced to 20 °C after induction and the cells were grown additional 20 hours to increase the fraction of soluble protein. This mutant was produced and purified as described above, except that only ampicillin for selection and IPTG for induction were needed.

## 4.2.1.3 High Performance Liquid Chromatography

Binding of labeled T4L mutants to peptidoglycan from *Micrococcus luteus* (Sigma-Aldrich) was monitored by reverse phase chromatography using a C-18 column out of ODS-A material (4 X 150 mm, 300 Å) (YMC Europe, GmbH). The protein was eluted with a gradient from 0 to 80% acetonitrile containing 0.01% trifluoroacetic acid for 25 min at a flow rate of 0.5 ml/min. The labeled complex elution was monitored by absorbance at 495 nm.

## 4.2.1.4 Fluorescence and spin Labeling

Site specific labeling of T4L was accomplished using orthogonal chemistry. For labeling the Keto function of the *p*-acetyl-L-phenylalanine (pAcF) amino acid at the N-terminus, hydroxylamine linker chemistry was used for Alexa 488 and Alexa 647 (Life Technologies Corp.). Cysteine mutants were labeled via a thiol reaction with maleimide linkers of the same fluorescent dyes. For spin labeling, the S44pAcF amino acid was labeled with a hydroxylamine nitroxide (HO-4120) and the cysteine at I150C was labeled with a methanthiosulfonate nitroxide (HO-225), both as previously reported(Fleissner et al, 2009). Double mutants were labeled sequentially - first thiol and second the keto handle, as suggested by Brustad *et al.* (Brustad et al, 2008). Single mutants were labeled in a one-step reaction.

The thiol reaction was carried out overnight at room temperature in 50 mM PB, 150 mM NaCl at pH 7.5 in presence of 5 times molar excess of dye. The keto reaction was done at pH 4.0 in 50 mM sodium acetate and 150 mM NaCl with 5 times molar excess of the dye for over 12 hours at 37°C. After each reaction, excess of unreacted dye was removed via a desalting column PD-10 (GE Healthcare) and the labeled sample was further concentrated using Amicon 10kDa concentrators (Millipore Corp.).

## 4.2.1.5 EPR Spectroscopy

For double electron electron resonance (DEER) measurements of doubly spin labeled proteins, ~ 200 μM spin-labeled T4L containing 20% glycerol (v/v) was placed in a quartz capillary (1.5 mm i.d. X 1.8 mm o.d.; VitroCom) and then flash-frozen in liquid nitrogen. Sample temperature was maintained at 80 K. The four-pulse DEER experiment was conducted on a Bruker Elexys 580 spectrometer fitted with an MS-2 split ring resonator. Pulses of 8 ns ($\pi/2$) and 16 ns ($\pi$) were amplified with a TWT amplifier (Applied Engineering Systems). Pump frequency was set at the maximum of the central resonance, and the observe frequency was 70 MHz less than the pump frequency. Dipolar data were analyzed by using a custom program written in LabVIEW (National Instruments Co.). Distance distributions were acquired using Tikhonov regularization (Chiang et al, 2005a).

## 4.2.1.6 Ensemble Time Correlated Single Photon Counting with high precision

Ensemble Time Correlated Single Photon Counting (eTCSPC) measurements were performed using an IBH-5000U (IBH, Scotland) system. The excitation sources were a 470 nm diode laser (LDH-P-C 470, PicoQuant, Germany) operating at 10 MHz for donor excitation and a 635 nm diode laser (LDH-8-1-126, PicoQuant, Germany) for acceptor excitation. The corresponding slits were set to 2 nm (excitation path) and 16 nm (emission path). Cut-off filters were used to reduce the contribution of the scattered light (>500 nm for donor and >640 nm for acceptor emission, respectively) and the monochromator was set to 520 nm for green detection and 665 nm for detecting the emission of the acceptor fluorophore. For the measurement of acceptor sensitized emission, the donor was excited at 470 nm and the emission of acceptor fluorophore was detected at 665 nm.

The TAC-histograms were recorded with a bin-width of 14.1 ps within a time-window of 57.8 ns. Photons were collected up to a peak count of 100,000 corresponding in average to a total number of $30 \cdot 10^6$ photons. The instrument response function $IRF$ was collected under the same recording settings at the excitation wavelength of the sample without cutoff-filters using a scattering Ludox-dispersion which yielded a comparable count rate as the later on measured samples.

Before each measurement session a reference measurement with a continuous light signal was performed to account for the differential non-linearities of the counting electronics. The recorded uncorrelated photons yield a reference histogram that is ideally constant. After recording of this measurement the average number of photons in each time-bin is calculated. Next the measurement is smoothed by a window function using a Hanning-filter with a window-size of 17 bins. The smoothed decay histogram is normalized to the previously calculated average number of photons. Instead of correcting the experimental histogram the model function is multiplied by the smoothed and normalized reference histogram to preserve the Poissonian statistics of the measured fluorescence intensity histograms of interest.

## 4.2.1.7 Multiparameter Fluorescence Detection (MFD)

MFD for confocal single molecule Förster Resonance Energy Transfer (smFRET) measurements was done using a 485 nm diode laser (LDH-D-C 485 PicoQuant, Germany, operating at 64 MHz, power at objective 110 µW) exciting freely diffusing labeled T4L molecule that passed through a detection volume of the 60X, 1.2 NA collar (0.17) corrected Olympus objective. The emitted fluorescence signal was collected through the same objective and spatially filtered using a

100 μm pinhole, to define an effective confocal detection volume. Then, the signal was divided into parallel and perpendicular components at two different colors ("green" and "red") through band pass filters, HQ 520/35 and HQ 720/150, for green and red respectively, and split further with 50/50 beam splitters. In total eight photon-detectors are used- four for green (τ-SPAD, PicoQuant, Germany) and four for red channels (APD SPCM-AQR-14, Perkin Elmer, Germany). A time correlated single photon counting (TCSPC) module (HydraHarp 400, PicoQuant, Germany) was used for data registration.

For smFRET measurements samples were diluted (buffer used 50 mM sodium phosphate, pH 7.5, 150 mM NaCl, 40 μM TROLOX and 1 μM unlabeled T4L) to pM concentration assuring ~ 1 burst per second. Collection time varied from several minutes up to 10 hours. To avoid drying out of the immersion water during the long measurements an oil immersion liquid with refraction index of water was used (Immersol, Carl Zeiss Inc., Germany). NUNC chambers (Lab-Tek, Thermo Scientific, Germany) were used with 500 μL sample volume. Standard controls consisted of measuring water to determine the instrument response function (IRF), buffer for background subtraction and the nM concentration green and red standard dyes (Rh110 and Rh101) in water solutions for calibration of green and red channels, respectively. To calibrate the detection efficiencies we used a mixture solution of double labeled DNA oligonucleotides with known distance separation between donor and acceptor dyes.

## 4.2.1.8 MFD burst analysis: Multiparameter FRET histograms and FRET lines

Bursts were selected by $2\sigma$ criteria out of the mean background value with cut off times that vary from sample to sample with a minimum of 60 photons for each burst. Each burst was then processed and fitted using a maximum likelihood algorithm (Maus et al, 2001) using in house developed programs (LabVIEW, National Instruments Co.). Fluorescent bursts were plotted in 2D histograms (Origin 8.6, OriginLab Co).

The relationship between the ratio of the donor fluorescence over the acceptor fluorescence $F_D/F_A$ and the fluorescence weighted donor lifetime obtained in burst analysis $\langle\tau_{D(A)}\rangle_f$ depends on specific experimental parameters such as fluorescence quantum yields of the dyes ($\Phi_{FD(0)}$ and $\Phi_{FA}$ for donor and acceptor respectively), background ($\langle B_G\rangle$ and $\langle B_R\rangle$ for green and red channels), detection efficiencies ($g_G$ and $g_R$ for green and red respectively) and crosstalk ($\alpha$). In the $F_D/F_A$ vs. $\langle\tau_{D(A)}\rangle_f$ 2D representations it is useful to represent a static FRET line such as:

$$\left(\frac{F_D}{F_A}\right)_{static} = \frac{\Phi_{FD(0)}}{\Phi_{FA}} \cdot \left(\frac{\tau_{D(0)}}{\tau_{D(A)}} - 1\right)^{-1} \tag{4.S1}$$

In Eq. 4.S1, $\tau_{D(A)} = \langle \tau_{D(A)} \rangle_f$ is the fluorescence averaged lifetime obtained via the maximum likelihood estimator when fitting ~100 green photons per burst. $\tau_{D(0)}$ is the donor fluorescence lifetime in the absence of acceptor.

The corrected fluorescence ($F_D$ and $F_A$) depends on the detection efficiencies of green ($g_G$) and red ($g_R$) channels as follows:

$$F_D = \frac{S_G - \langle B_G \rangle}{g_G}, \tag{4.S2}$$

$$F_A = \frac{S_R - \alpha F_G - \langle B_R \rangle}{g_R}, \tag{4.S3}$$

where the total signal in green and red channels are $S_G$ and $S_R$, respectively. The ratio ($F_D/F_A$) is weighted by the species fractions.

To properly describe the FRET line, one needs to consider that fluorophores are moving entities coupled to specific places via flexible linkers. This in turns generates a distance distribution between two fluorophores governed by the linker dynamics. Additionally, one needs to consider that the measured lifetime per burst is the fluorescence weighted average lifetime $\langle \tau_{D(A)} \rangle_f$. Therefore Eq. 4.S1 is only valid in the ideal scenario.

The goal is to transform Eq. 4.S1 to include the linker dynamics which is slower than the fluorescence decay time.

$$\left(\frac{F_D}{F_A}\right)_{static,L} = \frac{\Phi_{FD(0)}}{\Phi_{FA}} \cdot \left(\frac{\tau_{D(0)}}{\langle \tau_{D(A)} \rangle_{x,L}} - 1\right)^{-1}. \tag{4.S4}$$

To include this correction, the first thing to consider is a distance distribution between two fluorophores. We assume a Gaussian probability distance distribution with standard deviation $\sigma_{DA}$ and mean value $\langle R_{DA} \rangle$ such as

$$p(R_{DA}) = \frac{1}{\sqrt{2\pi} \cdot \sigma_{DA}} \exp\left(\frac{-(R_{DA} - \langle R_{DA} \rangle)^2}{2\sigma_{DA}^2}\right). \tag{4.S5}$$

For each $R_{DA}$ one can calculate the corresponding species lifetime following to

$$\tau_{D(A)}(R_{DA}) = \tau_{D(0)} \cdot \left(1 + \left(R_0 \big/ R_{DA}\right)^6\right)^{-1}, \tag{4.S6}$$

where each species corresponds to a different distance between the two fluorophores. For simplicity the donor lifetime is treated as mono exponential decay and $\tau_{D(0)} = \langle \tau_{D(0)} \rangle_x = \langle \tau_{D(0)} \rangle_f$. Each $\tau_{D(A)}(R_{DA})$ has a probability defined by the corresponding distribution $p(\tau_{DA}(R_{DA}))=p(R_{DA})$. The average species lifetime, due to linker dynamics, can be defined in the continuous approximation as

$$\left\langle \tau_{D(A)} \right\rangle_{x,L} = \int \tau_{D(A)}(R_{DA}) p(R_{DA}) dR_{DA} \, , \tag{4.S7}$$

and the fluorescence average lifetime as

$$\left\langle \tau_{D(A)} \right\rangle_{f,L} = \left\langle \tau_{D(A)} \right\rangle_f = \frac{\int \left( \tau_{D(A)}(R_{DA}) \right)^2 p(R_{DA}) dR_{DA}}{\left\langle \tau_{D(A)} \right\rangle_{x,L}} \tag{4.S8}$$

Thus, we can set a pair of parametric relations with respect to $\langle R_{DA} \rangle$ corresponding species to the species and fluorescence average lifetime such as

$$\left\langle \tau_{D(A)}(R_{DA}) \right\rangle_{x,L} \text{ and } \left\langle \tau_{D(A)}(R_{DA}) \right\rangle_{f,L}. \tag{4.S9}$$

Generally it is analytically impossible to solve Eq. 4.S9 using Gaussian distributions. Thus, Eq. 4.S9 is solved numerically covering a range for $\langle R_{DA} \rangle = [1$ Å to $5 R_0]$ Å. From the numerical solution we can create an empirical relation between the species and fluorescence average lifetimes for the selected range of $\langle R_{DA} \rangle$'s using an $i^{\text{th}}$ order polynomial function of with coefficients $A_{i,L}$ like

$$\left\langle \tau_{D(A)} \right\rangle_{x,L} = \sum_{i=0}^{n} A_i \left( \left\langle \tau_{D(A)} \right\rangle_{f,L} \right)^i \tag{4.S10}$$

Finally, we introduce Eq. (4.S10) into Eq. (4.S4) and obtain the static FRET line corrected for dye linker movements as

$$\left( \frac{F_D}{F_A} \right)_{\text{static,L}} = \frac{\Phi_{FD(0)}}{\Phi_{FA}} \cdot \left( \frac{\tau_{D(0)}}{\sum_{i=0}^{3} A_i \left( \left\langle \tau_{D(A)} \right\rangle_{f,L} \right)^i} - 1 \right)^{-1} . \tag{4.S11}$$

Coefficients ($A$'s) vary by variant. Experimentally, there is no difference on the observable thus, unless otherwise specified, we use in all figures and captions the assumption that all measured average lifetimes include the linker effect or $\langle \tau_{D(A)} \rangle_f = \langle \tau_{D(A)} \rangle_{f,L}$.

In the case of transition between two different states, one can also get an equation for a dynamic FRET line. In this case, a mixed fluorescence species arises from the interconversion

between two conformational states. For the simplest case the dynamic FRET line can be analytically presented as (Sisamakis et al, 2010)

$$\left(\frac{F_D}{F_A}\right)_{dyn,L} = \frac{\Phi_{FD(0)}}{\Phi_{FA}} \cdot \frac{\langle\tau_1\rangle_f \cdot \langle\tau_2\rangle_f}{\tau_{D(0)} \cdot \left(\langle\tau_1\rangle_f + \langle\tau_2\rangle_f - \sum_{i=0}^{3} A'_i \left(\langle\tau_{D(A)}\rangle_{f,L}\right)^i\right) - \langle\tau_1\rangle_f \cdot \langle\tau_2\rangle_f}, \qquad (4.S12)$$

where $\langle\tau_{D(A)}\rangle_{f,L}$ is the mixed fluorescence lifetime and $\Phi_{FD(0)}$, $\Phi_{FA}$ are the quantum yields of the donor and acceptor dyes, respectively. $\langle\tau_1\rangle_f$ and $\langle\tau_2\rangle_f$ are two donor fluorescence lifetimes in presence of acceptor at the beginning and end points of the interconverting states. The $C_{i,L}$ coefficients are determined for each FRET pair and differ from the $A$ coefficients. The $L$ sub index notation is to identify and specify the linker effects.

## 4.2.1.9 Donor and Acceptor quantum yields

Depending on the labeling position, the donor and acceptor fluorescence quantum yields vary and have been estimated for each sample. We assumed that only dynamic quenching takes place and that $\Phi_{FD(0)}$, $\Phi_{FA}$ are proportional to the species-averaged fluorescence lifetime $\langle\tau\rangle_x$ of donor or acceptor, respectively. As reference samples we used Alexa488-labeled DNA $\langle\tau_{D(0)}\rangle_x = 4.0$ ns, $\Phi_{FD(0)} = 0.8$ and for the acceptor we used Cy5-labeled DNA with $\langle\tau_A\rangle_x = 1.17$ ns and $\Phi_{FA} = 0.32$ (Woźniak et al, 2008). The obtained donor and acceptor quantum yields are presented in Table 4.S1. This FRET pair has a Förster distance of 52 Å.

## 4.2.1.10 Guidelines for reading MFD histograms

Several guidelines are needed to properly read MFD histograms. A short list is presented here. *I)* Donor only population is shown at high $F_D/F_A$ ratio with lifetime ~ 4 ns (donor-only for Alexa 488). *II)* High FRET appears at shorter lifetimes when the fluorescence of acceptor is high ($F_D/F_A \to 0$). *III)* Static FRET states distributions have circular shapes in $F_D/F_A$ vs. $\langle\tau_{D(A)}\rangle_f$ (typical scaling), and their distribution widths are given by shot noise and acceptor photophysics (Kalinin et al, 2010a). *IV)* Static FRET states follow a theoretical line that accounts for dye linker mobility called "static FRET line" Eq. (S11) (Kalinin et al, 2010a). *V)* A molecule that exchanges conformations at timescales faster than the diffusion time emits a burst of photons whose mixed

fluorescence is characterized by the fluorescence average lifetime. Therefore, dynamics in sub-ms range is expected when the shape is not circular. For example, elongation of the islands and deviation from the static line represent slow processes on the hundreds of microseconds. Tilting is an indication of fast dynamics.

## 4.2.1.11 Time-resolved fluorescence decay analysis

### *Model*

The fluorescence decay of the donor in the absence of FRET was multi-exponential, most likely, due to local quenching. To account for these effects the donor only reference samples were fitted by a multi-exponential relaxation model.

$$F_{D(0)}(t) = \sum_i x_{D(0)}^{(i)} \exp(-t/\tau_{D(0)}^{(i)})$$

(4.S13)

Here $\tau_{D(0)}^{(i)}$ is the donor fluorescence lifetime and $x_{D(0)}^{(i)}$ are the pre-exponential factors.

These effects were considered in the analysis of the FRET samples by global analysis. Here, it is assumed that all donor species are quenched by the same FRET rate constant $k_{FRET}$. This is true if quenching does not change the donor radiative lifetime and the spectral overlap. Then, instead of an empirical multi exponential decay function the donor fluorescence decay in the presence of FRET $F_{D(A)}(t)$ can be approximated with discrete FRET rate constants. Assuming the orientation factor $\kappa^2$ can be approximated by 2/3 the FRET-rate constants can be converted to distances and the time-resolved fluorescence of the donor in presence of the acceptor $F_{D(A)}(t)$ is given by:

$$F_{D(A)}(t) = \sum_i \sum_j x_{D(0)}^{(i)} x_{D(A)}^{(j)} \exp\left(-t \cdot \left(\frac{1}{\tau_{D(0)}^{(i)}} + \frac{1}{\tau_0} \cdot \left(\frac{R_0}{R_{DA}^{(j)}}\right)^6\right)\right)$$

(4.S14)

Here, $R_0$ is the Förster radius and for the Alexa488 and Alexa647 FRET pair is 52 Å by assuming $\kappa^2$ is 2/3, which is justified by the anisotropy studies compiled in Tables 4.S5A-S5D, $\tau_0 = 1/(k_{ISC} + k_{IC} + k_F)$ is the lifetime of the donor that corresponds to the quantum yield of the donor employed in the calculation of the Förster-radius, $\tau_{D(0)}^{(i)}$ are the lifetimes of the donor in the absence of FRET, and $x_{D(0)}^{(i)}$ is the pre-exponential factor of the corresponding donor-lifetime and $x_{D(A)}^{(j)}$ are

the species fractions of the distances $R_{DA}^{(j)}$. When the sum in Eq. 4.S14 has two $x_{D(A)}^{(j)}$ components or j = 2, we say that this is the two-state model, and the three-state model corresponds to j = 3. In the FRET-samples molecules not performing FRET are considered as a fraction ($x_{DOnly}$) of Donor-only molecules.

### eTCSPC

Given the model functions the experimental fluorescence intensity decays were fitted using the iterative re-convolution approach. Here, the model-decay curves were convoluted with the experimental instrument response function (IRF). Furthermore a constant offset c of the fluorescence intensity is considered. Finally, to correct the instrumental differential non-linearity the response of uncorrelated light was recorded and considered in the fitting procedure by multiplying the model-function with the normalized/smoothed uncorrelated instrumental response (Becker, 2005). Given these corrections the experimental time-resolved fluorescence intensities of the FRET-sample and the donor reference sample are proportional to:

$$F_{\text{FRET}}(t) = \left(N_0 \cdot \left[(1 - x_{\text{DOnly}}) F_{\text{D(A)}}(t) + x_{\text{DOnly}} F_{\text{D(0)}}(t)\right] \otimes IRF + sc \cdot IRF + c\right) \cdot Lin$$
$$F_{\text{Ref}}(t) = \left(N_0 \cdot F_{\text{D(0)}}(t) \otimes IRF + sc \cdot IRF + c\right) \cdot Lin$$

(4.S15)

Here *sc* is due to scattered light of the sample. The scaling of the model fluorescence intensity histograms was performed given the experimental measured number of photons. Here the model function was scaled such that the number of photons corresponds to the experimental number of photons. This reduces the number of free fitting parameters as the initial amplitude $N_0$ is not fitted.

### seTCSPC

Due to the lower photon-count in the single-molecule as compared to the ensemble data the number of free parameters had to be further reduced. This was accomplished by buffer reference measurements which allow for the determination of the number of fluorescent photons $N_F$ and the number of background photons $N_{BG}$ as the measurement time of the background measurement and the single-molecule measurements are known. Given the number of fluorescence and background photons the constant offset and the fraction of scattered light does not have to be fitted and the fluorescence decays are given by:

$$F_{\text{FRET}}(t) = N_F \cdot \left[(1 - x_{\text{DOnly}})F_{\text{D(A)}}(t) + x_{\text{DOnly}}F_{\text{D(0)}}(t)\right] \otimes IRF + N_{BG} \cdot IRF$$

$$F_{\text{Ref}}(t) = N_F \cdot F_{\text{D(0)}}(t) \otimes IRF + N_{BG} \cdot IRF$$

(4.S16)

### *Fitting and error-estimation*

Functional variants were fitted globally, i. e. distances for states $C_1$ and $C_2$ were linked over all three variants used to mimic free enzyme E, enzyme-substrate complex ES and enzyme product complex EP while the distance for $C_3$ was only linked for E and ES to allow for the different (covalent) nature of this state in EP. The experimental fluorescence decays were fitted by the conventional Levenberg–Marquardt algorithm using custom software written in Python.

The statistical uncertainties of the fits were estimated by exploring the $\chi^2$-surface of the model function using the Metropolis-Hastings algorithm with at least 20 independent Markov-chains with 50000 steps each starting from the minimum fit-result $p_{\min}$ with a step-width of $w = p_{\min} \cdot r \cdot w_0$ (where r is a random sample from a uniform distribution over [-0.5, 0.5] and $w_0$=0.005) up to the maximum confidence level $conf_{\max} = 1\text{-}10^{-5}$. The error-margins of the individual fitting parameters are the projection to the individual parameter-dimension. The maximum allowed $\chi^2_{r,\max}$ for a given confidence-level (P; e.g. for 2σ P = 0.95) was calculated by:

$$\chi^2_{r,\max}(P) = \chi^2_{r,\min} \cdot \left[1 + n/\upsilon \cdot \text{cdf}^{-1}(F(n,\upsilon,P))\right],$$

(4.S17)

where $\text{cdf}^{-1}(F(n,\nu,P))$ is the inverse of the cumulative distribution function of the F-distribution for n number of free parameters, and with ν degrees of freedom. $\chi^2_{r,\min}$ is the minimum determined $\chi^2_r$ (Lakowicz, 2006).

## 4.2.1.12 Filtered Fluorescence Correlation Spectroscopy

In fluorescence correlation spectroscopy (FCS) information on fluctuating systems is obtained by calculation of the correlation function (Elson & Magde, 1974; Magde et al, 1974)

$$^{A,B}G(t_c) = 1 + \frac{\langle \delta^A S(t) \cdot \delta^B S(t + t_c) \rangle}{\langle ^A S(t) \rangle \cdot \langle ^B S(t) \rangle}.$$

(4.S18)

where $t_c$ is the correlation time, $^{A,B}S(t)$ represents the detected intensity signal (number of detected photons per time interval) at channels $A$ or $B$, and $\delta^{A;B}S(t)$ corresponds to the fluctuations from the time average signal denoted as $\left\langle ^{A;B}S(t)\right\rangle$. The auto-correlation function is defined when the detected channels are the same $A = B$ and it is called cross-correlation if $A \neq B$.

For a mixture of $n$ molecular species, ignoring fluctuations contributions from additional kinetics, an analytical form of the correlation function was derived (Kim et al, 2005; Schwille, 2001).

$$G(t_c) = 1 + \frac{1}{N} \cdot \frac{\sum\limits_{i}^{n} x^{(i)} \cdot (Q^{(i)})^2 \cdot G_{diff}^{(i)}(t_c)}{\left(\sum\limits_{i}^{n} x^{(i)} \cdot Q^{(i)}\right)^2}, \tag{4.S19}$$

where

$$G_{diff}^{(i)}(t_c) = \left(1 + \frac{t_c}{t_{diff}^{(i)}}\right)^{-1} \cdot \left(1 + \left(\frac{\omega_0}{z_0}\right)^2 \cdot \frac{t_c}{t_{diff}^{(i)}}\right)^{-\frac{1}{2}}. \tag{4.S20}$$

A 3-dimensional Gaussian shaped volume element with spatial distribution of the detection probabilities: $W(x,y,z) = \exp\left(-2\left(x^2 + y^2\right)/\omega_0^2\right) \cdot \exp\left(-2 z^2/z_0^2\right)$ is assumed. The $1/e^2$ radii in $x$ and $y$ or in $z$ direction are denoted by $\omega_0$ and $z_0$, respectively. The diffusion coefficient for each species $i$ is related to its characteristic diffusion time $t_{diff}^{(i)}$ as $t_{diff}^{(i)} = \omega_0^2/4D^{(i)}$. The amplitude of the correlation is scaled with the reciprocal of the average number of fluorescent particles $N$ in the confocal volume. Each molecular fraction $x^{(i)} = \dfrac{c^{(i)}}{\sum\limits_{i} c^{(i)}}$ has a concentration $c^{(i)}$, and brightness $Q^{(i)}$.

To separate species, we use filtered FCS (fFCS)(Böhmer et al, 2002; Felekyan et al, 2012). fFCS differs from standard FCS (Elson & Magde, 1974) and FRET-FCS (Felekyan et al, 2013) by interrogating the "species" (conformational states) fluctuations instead of photon count rates (Felekyan et al, 2013). We define the species cross- correlation function (sCCF) as

$$G^{(i,m)}(t_c) = \frac{\langle F^{(i)}(t) \cdot F^{(m)}(t+t_c) \rangle}{\langle F^{(i)}(t) \rangle \cdot \langle F^{(m)}(t+t_c) \rangle} = \frac{\left\langle \left( \sum_{j=1}^{d \cdot L} f_j^{(i)} \cdot S_j(t) \right) \cdot \left( \sum_{j=1}^{d \cdot L} f_j^{(m)} \cdot S_j(t+t_c) \right) \right\rangle}{\left\langle \sum_{j=1}^{d \cdot L} f_j^{(i)} \cdot S_j(t) \right\rangle \cdot \left\langle \sum_{j=1}^{d \cdot L} f_j^{(m)} \cdot S_j(t+t_c) \right\rangle} \qquad (4.S21)$$

where (*i*) and (*m*) are two selected "species" in a mixture. A set of filters $f_j^{(i)}$ that depend on the arrival time of each photon after each excitation pulse is used. The signal $S_j(t)$, obtained via pulsed excitation is recorded at each j = 1 ... *L* TCSPC-channel. The signal and filters per detector, *d*, are stacked in a single array with dimensions $d \cdot L$ for global minimization according to (Felekyan et al, 2012). Filters are defined in such a way that the relative "error" difference between the photon count per species ($w^{(i)}$) and the weighted histogram $f_j^{(i)} \cdot H_j$ is minimized as defined in Eq. (4.S22).

$$\left\langle \left( \sum_{j=1}^{d \cdot L} f_j^{(i)} \cdot H_j - w^{(i)} \right)^2 \right\rangle \to \min . \qquad (4.S22)$$

where brackets represent time averaging.

The requirement is that the decay histogram $H_j$ can be expressed as a linear combination of the conditional probability distributions $p_j^{(i)}$, such as $H_j = \sum_{i=1}^{n(=2)} w^{(i)} p_j^{(i)}$, with $\sum_{j=1}^{d \cdot L} p_j^{(i)} = 1$. Hence, the species cross-correlation $G^{(i,m)}(t_c)$ provides maximal contrast for intercrossing dynamics (Felekyan et al, 2012). One major advantage of sCCF is that, if photophysical properties are decoupled from species selection, the intercrossing dynamics (Felekyan et al, 2013) is recovered with great fidelity.

To properly fit the species cross-correlation function we used (Felekyan et al, 2012)

$$G(t_c) = 1 + \frac{1}{N} \cdot G_{diff}(t_c) \cdot [1 - G_K(t_c)], \qquad (4.S23)$$

where $G_K(t_c)$ is

$$G_K(t_c) = \sum_{t_{Ri}}^{t_{Rn}} A_{Ki} \exp(-t_c / t_{Ri}) \qquad . \qquad (4.S24)$$

In Eq. (4.S24) the summation is over *n* reaction times $t_{Rn}$.

The same 3-dimensional Gaussian shaped volume element is assumed. We assume that $G_{diff}(t_c) = G_{diff}^{(i)}(t_c) = G_{diff}^{(m)}(t_c)$ take the form of Eq. (4.S20). The normalized correlation function is presented as:

$$g(t_c) = N \cdot (G(t_c) - 1). \tag{4.S25}$$

Filtered FCS requires prior knowledge of the time-resolved fluorescence and polarization decays for each species. For a mixture of more than two species, we generated two decays corresponding to two "pseudo-species". Using the scatter profile as the excitation pulse the parallel and perpendicular decay components $F_{\parallel}(t)$ and $F_{\perp}(t)$ for each "pseudo-species" were modeled as

$$F_{\parallel}(t) = F(t) \cdot (1 + (2 - 3l_1) \cdot r(t)) / 3$$
$$F_{\perp}(t) = F(t) \cdot (1 - (1 - 3 \cdot l_2) \cdot r(t)) / 3 \tag{4.S26}$$

where $F(t)$ is the time-resolved fluorescence decay at magic angle (typically mono-exponential), and $l_1 = 0.01758$ and $l_2 = 0.0526$ are correction factors (Koshioka et al, 1995; Schaffer et al, 1999). The anisotropy decay $r(t)$ is given by

$$r(t) = r_{0,ov} \exp(-t / \rho_{overall}) + r_{0,ba} \exp(-t / \rho_{backbone}) + r_{0,li} \exp(-t / \rho_{linker}) \tag{4.S27}$$

Background signal consists of dark counts (uniformly distributed over TCSPC channels) and scatter contribution.

## 4.2.1.13 Accessible volume (AV) model and interdye distance

The accessible volume considers dyes as hard sphere models connected to the protein via flexible linkers (modeled as a flexible cylindrical pipe) (Sindbert et al, 2011a). The overall dimension (width and length) of the linker is based on their chemical structures. For Alexa 488 the five carbon linker length was set to 20 Å, the width of the linker is 4.5 Å and the dye radii $R_1 = 5$ Å, $R_2 = 4.5$ Å and $R_3 = 1.5$ Å. For Alexa 647 the dimensions used were: length = 22 Å, width = 4.5 Å and three dye radii $R_1 = 11$ Å, $R_2 = 3$ Å and $R_3 = 3.5$ Å. Similar approaches have been introduced before to predict possible positions for EPR and FRET labels (Cai et al, 2007; Kalinin et al, 2012; Muschielok et al, 2008).

To account for dye linker mobility we generated a series of AV's for donor and acceptor dyes attached to T4L placing the dyes at multiple separation distances. For each pair of AV's, we calculated the distance between dye mean positions ($R_{mp}$)

$$R_{mp} = \left| \left\langle \vec{R}_{D(i)} \right\rangle - \left\langle \vec{R}_{A(j)} \right\rangle \right| = \left| \frac{1}{n} \sum_{i=1}^{n} \vec{R}_{D(i)} - \frac{1}{m} \sum_{j=1}^{m} \vec{R}_{A(j)} \right| \tag{4.S28}$$

where $\vec{R}_{D(i)}$ and $\vec{R}_{A(i)}$ are all the possible positions that the donor fluorophore and the acceptor fluorophore can take. However, in ensemble TCSPC (eTCSPC) the mean donor-acceptor distance is observed:

$$\left\langle R_{DA} \right\rangle = \left| \left\langle \vec{R}_{D(i)} - \vec{R}_{A(j)} \right\rangle \right| = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \left| \vec{R}_{D(i)} - \vec{R}_{A(j)} \right| \tag{4.S29}$$

which can be modeled with the accessible volume calculation.

The relationship between $R_{mp}$ and $\langle R_{DA} \rangle$ can be derived empirically following a third order polynomial from many different simulations. The $\langle R_{DA} \rangle$ is not directly connected to the distance between atoms on the backbone (Cα-Cα).

## 4.2.1.14 Single molecule Brownian dynamics simulator

Simulations of single-molecule measurements were done via Brownian dynamics (Dix et al, 2006; Enderlein et al, 1997; Kask et al, 1999; Laurence et al, 2004). The spatial intensity distribution of the observation volume was assumed a 3D Gaussian. In contrast to other simulators, freely diffusing molecules in an "open" volume are used. Transition kinetics is modeled by allowing $i \rightarrow j$ transitions. The time that molecules spend in $i$ and $j$ states ($t_i$ and $t_j$, respectively) are exponentially distributed with

$$P(t_i) = k_i^{-1} \exp(-k_i t_i), \text{ and } P(t_j) = k_j^{-1} \exp(-k_j t_j) \tag{4.S30}$$

Simulated photon counts are saved in SPC-132 data format (Becker & Hickel GmbH, Berlin, Germany) and treated as experimental data.

## 4.2.1.15 FRET positioning and screening (FPS)

FPS is done in four steps, and its flow is based on the recent work by Kalinin *et al.* (Kalinin et al, 2012). In order to do our experimental design and a combinatorial screening of $C_1$ and $C_2$ using 409 PDB structures with respect to our FRET data, FPS calculates the donor and acceptor accessible volumes for each donor-acceptor labeling scheme. We then compute an error function for each conformation $j$

$$\chi^2_{r,FPS}{}^{(j)} = \frac{1}{N} \sum_{i=1}^{N} \frac{\left( \langle R_{DA} \rangle^{(i)}_{\text{experiment}} - \langle R_{DA} \rangle^{(i)}_{\text{model,j}} \right)^2}{\Delta R^2_{DA}\left( \kappa^2, k_{FRET} \right)^{(i)}} \tag{4.S31}$$

where $N = 24$ is the total number of FRET distances ($\langle R_{DA} \rangle$) and the overall uncertainty $\Delta R^2_{DA}\left( \kappa^2, k_{FRET} \right)$ is determined by the error-propagation rule of $\Delta R^2_{DA}\left( \kappa^2, k_{FRET} \right) = \Delta R^2_{DA}\left( \kappa^2 \right) + \Delta R^2_{DA}\left( k_{FRET} \right)$. $\Delta R^2_{DA}\left( \kappa^2 \right)$ is the uncertainty from the mutual orientation of donor and acceptor ($\kappa^2$ errors) and can be estimated by analyzing anisotropy decays accessible in MFD (Dale et al, 1979). We assume a weak dependence of $R_0$ on $\kappa^2$ (van der Meer et al, 1994) (Table 4.S5D). As a consequence, correcting $\langle R_{DA} \rangle$ values for the mean $\kappa^2$ for all allowed orientations of D and A yields almost the same final screened structures as with the assumption of $\langle \kappa^2 \rangle = 2/3$. $\Delta R^2_{DA}\left( k_{FRET} \right)^{(i)}$ corresponds to the statistical uncertainty on determining $R_{DA}$ as function of the $k_{FRET}$ used on the model-function to fit eTCSPC data (see Eq. 4.S14-S15). The error propagation was applied to all measured FRET distances (Table 4.S3).

## 4.2.2. Supporting Results

### 4.2.2.1 Fluorescence decay analysis of single and double labeled T4 Lysozyme

Selected mutants were labeled in two configurations (DA) and (AD), D for donor (Alexa 488) and A for acceptor fluorophore (Alexa 647). The order of the letters represents the position of the fluorophore. The first letter represents the label of the keto handle at the N-terminal domain and the second position the thiol reaction for labeling at the C-terminus, except for the double cysteine mutant.

Each sample was measured in eTCSPC as described in the materials and methods section and analyzed with various models including multi-exponential decay, one, two and three FRET- rates. The best consistent model based on our experimental data and statistical analysis is that three-FRET rates are needed to describe all T4L variants.

To reach that conclusion first we needed to characterize the donor and acceptor fluorescence quantum yield $\Phi_{FD(0)}$ and $\Phi_{FA}$, respectively. A summary table of these is shown in Table 4.S1A. A summary of the fits using the two and three state models are found in Table 4.S1B and 4.S1C, decays are shown in Figure 4.S2.

**Table 4.S1A.** Multi exponential fit of Donor-only labeled samples, fluorescence quantum yields of the donor (Alexa488) and acceptor (Alexa647) and the multi exponential model fits of ensemble fluorescence decays for all FRET variants Eq. (4.S13-S15). Fluorescence quantum yields are calculated from the species averaged lifetimes $\langle\tau\rangle_x = \sum_{i=0}^{n} x_i \tau^{(i)}$, where $x_i$'s are the species fractions.

| Samples | $\tau_1$ [ns] | $x_1$ | $\tau_2$ [ns] | $x_2$ | $\tau_3$ [ns] | $x_3$ | $\langle\tau\rangle_{x,D(0)}$ [ns] | $\Phi_{FD(0)}$ | $\langle\tau\rangle_{x,A}$ [ns] | $\Phi_{FA}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| R8pAcF/Q69C | 4.10 | 0.89 | 1.60 | 0.11 | | | 3.83 | 0.77 | 1.28 | 0.35 |
| R8pAcF/P86C | 4.16 | 0.92 | 1.89 | 0.08 | | | 3.98 | 0.80 | 1.25 | 0.34 |
| K19pAcF/Q69C | 4.25 | 0.78 | 2.22 | 0.13 | 0.65 | 0.09 | 3.66 | 0.73 | 1.25 | 0.34 |
| K19pAcF/P86C | 3.90 | 0.77 | 2.56 | 0.15 | 0.32 | 0.08 | 3.41 | 0.68 | 1.22 | 0.33 |
| K19pAcF/R119C | 3.91 | 0.74 | 2.51 | 0.15 | 0.52 | 0.11 | 3.33 | 0.67 | 1.26 | 0.35 |
| K19pAcF/N132C | 3.91 | 0.81 | 2.66 | 0.14 | 0.32 | 0.05 | 3.56 | 0.71 | 1.34 | 0.37 |
| S44pAcF/P86C | 4.32 | 0.94 | 1.75 | 0.06 | | | 4.17 | 0.83 | 1.23 | 0.34 |
| S44pAcF/R119C | 4.32 | 0.96 | 1.44 | 0.04 | | | 4.20 | 0.84 | 1.20 | 0.33 |

| Variant | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| S44pAcF/N132C | 4.32 | 0.96 | 1.44 | 0.04 | | | 4.20 | 0.84 | 1.30 | 0.35 |
| S44pAcF/I150C | 4.32 | 0.96 | 1.44 | 0.04 | | | 4.20 | 0.84 | 1.35 | 0.37 |
| N55pAcF/Q69C | 4.14 | 0.92 | 1.48 | 0.08 | | | 3.93 | 0.79 | 1.26 | 0.34 |
| N55pAcF/R119C | 4.412 | 0.84 | 2.14 | 0.10 | 0.22 | 0.06 | 3.93 | 0.79 | 1.26 | 0.34 |
| N55pAcF/N132C | 4.28 | 0.94 | 1.49 | 0.06 | | | 4.11 | 0.82 | 1.26 | 0.34 |
| N55pAcF/I150C | 4.32 | 0.69 | 3.08 | 0.25 | 0.72 | 0.06 | 3.79 | 0.76 | 1.27 | 0.35 |
| K60pAcF/P86C | 4.12 | 0.94 | 2.07 | 0.06 | | | 4.00 | 0.80 | 1.24 | 0.34 |
| K60pAcF/R119C | 4.26 | 0.91 | 1.81 | 0.09 | | | 4.04 | 0.81 | | |
| K60pAcF/N132C | 4.08 | 0.89 | 1.78 | 0.11 | | | 3.83 | 0.77 | | |
| K60pAcF/I150C | 4.09 | 0.88 | 1.76 | 0.12 | | | 3.81 | 0.76 | | |
| Q69pAcF/P86C | 4.20 | 0.94 | 1.52 | 0.06 | | | 4.04 | 0.81 | 1.26 | 0.34 |
| Q69pAcF/R119C | 4.20 | 0.88 | 1.64 | 0.12 | | | 3.89 | 0.78 | 1.25 | 0.34 |
| Q69pAcF/N132C | 4.20 | 0.89 | 1.47 | 0.11 | | | 3.90 | 0.78 | 1.33 | 0.36 |
| Q690pAcF/I150C | 4.09 | 0.89 | 1.88 | 0.11 | | | 3.85 | 0.77 | 1.29 | 0.35 |
| D70pAcF/R119C | 4.14 | 0.68 | 2.61 | 0.23 | 0.82 | 0.09 | 3.49 | 0.70 | 1.26 | 0.34 |
| D70pAcF/N132C | 4.08 | 0.88 | 1.12 | 0.12 | | | 3.72 | 0.74 | 1.26 | 0.34 |

**Table 4.S1B.** Table of determined inter-fluorophore distances for one and two FRET rates fit (Eq. 4.S14, S15)

| Measurement number | Variant | 1 state model | | | 2 state model | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\langle R_{DA} \rangle$ [Å] | $x_{DOnly}$ | $\chi_r^2$ | $\langle R_{DA1} \rangle$ [Å] | $x_1$ | $\langle R_{DA2} \rangle$ [Å] | $x_2$ | $x_{DOnly}$ | $\chi_r^2$ |
| *1* | R8pAcF/Q69C-(DA) | 41.0 | 0.30 | 4.81 | 37.1 | 0.74 | 47.1 | 0.26 | 0.24 | 1.23 |
| *2* | R8pAcF/P86C-(DA) | 45.3 | 0.28 | 7.55 | 38.2 | 0.54 | 49.7 | 0.46 | 0.21 | 1.15 |
| *3* | K19pAcF/Q69C-(DA) | 48.8 | 0.58 | 8.82 | 38.5 | 0.50 | 64.0 | 0.50 | 0.36 | 1.15 |
| *4* | K19pAcF/P86C-(DA) | 49.8 | 0.15 | 6.55 | 53.0 | 0.67 | 40.5 | 0.33 | 0.10 | 1.21 |
| *5* | K19pAcF/R119C-(DA) | 49.9 | 0.16 | 5.20 | 52.5 | 0.71 | 40.5 | 0.29 | 0.11 | 1.23 |
| *6* | K19pAcF/N132C-(DA) | 46.9 | 0.15 | 12.44 | 50.8 | 0.52 | 38.2 | 0.48 | 0.10 | 1.31 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *7* | S44pAcF/P86C-(DA) | 49.3 | 0.31 | 5.03 | 52.9 | 0.64 | 40.3 | 0.36 | 0.25 | 1.19 |
| *8* | S44pAcF/R119C-(DA) | 51.7 | 0.33 | 5.01 | 55.6 | 0.69 | 41.3 | 0.31 | 0.26 | 1.11 |
| *9* | S44pAcF/N132C-(DA) | 52.2 | 0.52 | 3.73 | 57.7 | 0.68 | 40.9 | 0.32 | 0.43 | 1.21 |
| *10* | S44pAcF/I150C-(DA) | 52.0 | 0.12 | 8.12 | 55.4 | 0.70 | 42.0 | 0.30 | 0.06 | 1.11 |
| *11* | N55pAcF/Q69C-(DA) | 40.4 | 0.33 | 3.57 | 36.8 | 0.74 | 46.2 | 0.26 | 0.29 | 1.24 |
| *12* | N55pAcF/R119C-(DA) | 52.0 | 0.51 | 3.36 | 57.0 | 0.68 | 41.0 | 0.32 | 0.42 | 1.26 |
| *13* | N55pAcF/N132C-(DA) | 50.0 | 0.12 | 10.93 | 53.7 | 0.63 | 41.1 | 0.37 | 0.07 | 1.26 |
| *14* | N55pAcF/I150C-(DA) | 50.6 | 0.62 | 1.57 | 53.4 | 0.76 | 42.4 | 0.24 | 0.59 | 1.19 |
| *15* | K60pAcF/P86C-(DA) | 47.9 | 0.47 | 2.61 | 50.7 | 0.66 | 38.8 | 0.34 | 0.41 | 1.12 |
| *16* | K60pAcF/R119C-(DA) | 50.2 | 0.14 | 10.09 | 53.6 | 0.65 | 40.7 | 0.35 | 0.09 | 1.23 |
| *17* | K60pAcF/N132C-(DA) | 45.6 | 0.27 | 6.25 | 49.3 | 0.53 | 38.0 | 0.47 | 0.21 | 1.05 |
| *18* | K60pAcF/I150C-(DA) | 44.4 | 0.31 | 5.20 | 37.4 | 0.52 | 48.2 | 0.48 | 0.25 | 1.08 |
| *19* | Q69pAcF/P86C-(DA) | 41.1 | 0.55 | 1.94 | 36.8 | 0.63 | 47.4 | 0.37 | 0.47 | 1.10 |
| *20* | Q69pAcF/R119C-(DA) | 43.9 | 0.40 | 4.32 | 38.1 | 0.59 | 48.9 | 0.41 | 0.34 | 1.12 |
| *21* | Q69pAcF/N132C-(DA) | 43.7 | 0.36 | 3.79 | 37.7 | 0.54 | 48.8 | 0.46 | 0.27 | 1.15 |
| *22* | Q690pAcF/I150C-(DA) | 45.6 | 0.53 | 2.25 | 48.7 | 0.59 | 37.8 | 0.41 | 0.47 | 1.11 |
| *23* | D70pAcF/R119C-(DA) | 42.7 | 0.21 | 7.51 | 36.1 | 0.61 | 46.5 | 0.39 | 0.15 | 1.15 |
| *24* | D70pAcF/N132C-(DA) | 42.8 | 0.26 | 8.11 | 36.3 | 0.64 | 47.5 | 0.36 | 0.18 | 1.20 |

**Table 4.S1C.** Table of determined inter-fluorophore distances including statistical uncertainties as described in sections 4.2.1.11, 4.2.1.13 and 4.2.1.15. Distances were used for FPS. We present the weighted residuals for the lowest $\chi_r^2$ for each conformation $C_1$, $C_2$ and $C_3$ and the Cα- Cα distances from the best model structure.

| Measurement number | Conf. state | Cα-Cα X-ray* [Å] | $\langle R_{DA} \rangle_{X\text{-ray}}$ [Å] | Experiment | | | | | Screening w. res. ** [Å/Å] |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $\langle R_{DA} \rangle$ [Å] | $x_i$ | $\pm\Delta R$ $\langle \kappa^2 \rangle$ [Å] | $\pm\Delta R$ $\langle \tau_{DA} \rangle$ [Å] | $\chi_r^2$ $(x_{DOnly})$ | |
| **R8pAcF/Q69C-(DA)** | | | | | | | | | |
| 1 | $C_1$ | 15.1 | 33.2 | 39.7 | 0.51 | 4.1 | 6.4 | 1.03 (0.19) | -0.16 |
| | $C_2$ | 13.7 | 34.8 | 31.0 | 0.39 | 3.2 | 5.4 | | 0.04 |
| | $C_3$ | 13.9 | 32.2 | 51.2 | 0.10 | 5.3 | 8.1 | | -0.29 |
| **R8pAcF/P86C-(DA)** | | | | | | | | | |
| 2 | $C_1$ | 26.8 | 44.3 | 51.9 | 0.29 | 5.6 | 8.1 | 1.10 (0.20) | -0.12 |
| | $C_2$ | 26.0 | 46.8 | 36.0 | 0.36 | 3.9 | 6.1 | | 0.22 |
| | $C_3$ | 26.4 | 44.5 | 43.4 | 0.35 | 4.6 | 7.6 | | 0.02 |
| **K19pAcF/Q69C-(DA)** | | | | | | | | | |
| 3 | $C_1$ | 20.2 | 38.8 | 33.7 | 0.26 | 3.5 | 5.3 | 1.06 (0.24) | 0.18 |
| | $C_2$ | 20.7 | 41.8 | 41.7 | 0.26 | 4.3 | 7.0 | | -0.04 |
| | $C_3$ | 20.6 | 40.1 | 68.7 | 0.47 | 7.1 | 11.4 | | -0.22 |
| **K19pAcF/P86C-(DA)** | | | | | | | | | |
| 4 | $C_1$ | 35.9 | 53.9 | 55.4 | 0.46 | 5.4 | 8.1 | 1.14 (0.08) | -0.02 |
| | $C_2$ | 34.2 | 53.7 | 46.2 | 0.36 | 4.5 | 7.2 | | 0.11 |
| | $C_3$ | 34.0 | 51.5 | 36.2 | 0.17 | 3.6 | 6.2 | | 0.40 |
| **K19pAcF/R119C-(DA)** | | | | | | | | | |
| 5 | $C_1$ | 37.9 | 51.8 | 55.0 | 0.46 | 5.9 | 8.4 | 1.13 (0.10) | -0.05 |
| | $C_2$ | 33.5 | 52.1 | 46.3 | 0.39 | 5.0 | 7.4 | | 0.00 |
| | $C_3$ | 33.5 | 47.8 | 35.2 | 0.15 | 3.8 | 6.1 | | 0.34 |
| **K19pAcF/N132C-(DA)** | | | | | | | | | |
| 6 | $C_1$ | 35.3 | 49.9 | 54.9 | 0.26 | 7.6 | 9.7 | 1.14 (0.08) | -0.05 |
| | $C_2$ | 28.2 | 44.8 | 45.4 | 0.39 | 6.3 | 8.6 | | -0.05 |
| | $C_3$ | 29.0 | 42.6 | 35.7 | 0.34 | 4.9 | 7.0 | | 0.14 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **S44pAcF/P86C-(DA)** | | | | | | | | | |
| *7* | $C_1$ | 39.3 | 53.0 | 61.0 | 0.19 | 6.3 | 9.2 | 1.16 (0.23) | -0.09 |
| | $C_2$ | 37.4 | 50.2 | 49.9 | 0.54 | 5.1 | 8.1 | | 0.01 |
| | $C_3$ | 37.4 | 49.8 | 38.8 | 0.27 | 4.0 | 6.0 | | 0.31 |
| **S44pAcF/R119C-(DA)** | | | | | | | | | |
| *8\*\*\** | $C_1$ | 44.2 | 58.8 | 55.6 | 0.69 | 4.6 | 8.0 | 1.11 (0.00) | 0.06 |
| | $C_2$ | 40.3 | 52.6 | | | | | | -0.10 |
| | $C_3$ | 40.3 | 53.3 | 41.3 | 0.31 | 3.4 | 6.2 | | 0.31 |
| **S44pAcF/N132C-(DA)** | | | | | | | | | |
| *9\*\*\** | $C_1$ | 45.2 | 63.2 | 57.7 | 0.68 | 6.7 | 9.5 | 1.21 (0.12) | 0.07 |
| | $C_2$ | 39.6 | 51.5 | | | | | | -0.04 |
| | $C_3$ | 39.4 | 53.6 | 40.9 | 0.32 | 4.7 | 7.0 | | 0.26 |
| **S44pAcF/I150C-(DA)** | | | | | | | | | |
| *10* | $C_1$ | 38.9 | 59.5 | 65.5 | 0.20 | 6.3 | 9.4 | 1.07 (0.02) | -0.07 |
| | $C_2$ | 35.4 | 52.3 | 52.8 | 0.56 | 5.1 | 8.0 | | -0.01 |
| | $C_3$ | 35.7 | 52.3 | 40.9 | 0.23 | 3.9 | 6.4 | | 0.29 |
| **N55pAcF/Q69C-(DA)** | | | | | | | | | |
| *11* | $C_1$ | 20.9 | 35.9 | 32.7 | 0.32 | 3.9 | 5.6 | 1.17 (0.27) | 0.10 |
| | $C_2$ | 21.2 | 35.9 | 39.4 | 0.56 | 4.7 | 7.1 | | -0.11 |
| | $C_3$ | 21.3 | 33.7 | 49.4 | 0.23 | 5.9 | 8.5 | | -0.22 |
| **N55pAcF/R119C-(DA)** | | | | | | | | | |
| *12* | $C_1$ | 46.5 | 63.8 | 62.4 | 0.45 | 5.1 | 8.4 | 1.24 (0.38) | 0.02 |
| | $C_2$ | 44.8 | 62.9 | 49.6 | 0.34 | 4.1 | 7.3 | | 0.18 |
| | $C_3$ | 45.0 | 38.3 | 38.3 | 0.21 | 3.1 | 6.4 | | 0.58 |
| **N55pAcF/N132C-(DA)** | | | | | | | | | |
| *13* | $C_1$ | 46.2 | 65.1 | 64.7 | 0.18 | 6.5 | 9.5 | 1.18 (0.03) | 0.00 |
| | $C_2$ | 43.1 | 60.4 | 50.8 | 0.55 | 5.1 | 7.9 | | 0.15 |
| | $C_3$ | 43.0 | 61.0 | 39.4 | 0.27 | 3.9 | 6.3 | | 0.54 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **N55pAcF/I150C-(DA)** | | | | | | | | | |
| 14*** | $C_1$ | 38.7 | 56.0 | 53.4 | 0.76 | 5.8 | 8.5 | 1.19 (0.47) | 0.04 |
| | $C_2$ | 37.3 | 53.1 | | | | | | 0.00 |
| | $C_3$ | 37.7 | 42.4 | 42.4 | 0.24 | 4.6 | 6.8 | | 0.25 |
| **K60pAcF/P86C-(DA)** | | | | | | | | | |
| 15*** | $C_1$ | 32.7 | 49.7 | 50.7 | 0.66 | 5.6 | 8.2 | 1.12 (0.26) | -0.02 |
| | $C_2$ | 35.1 | 52.5 | | | | | | 0.02 |
| | $C_3$ | 35.2 | 52.6 | 38.8 | 0.34 | 4.3 | 6.6 | | 0.32 |
| **K60pAcF/R119C-(DA)** | | | | | | | | | |
| 16 | $C_1$ | 36.4 | 53.8 | 46.4 | 0.33 | 5.1 | 7.2 | 1.14 (0.08) | 0.14 |
| | $C_2$ | 38.6 | 57.3 | 55.7 | 0.47 | 6.1 | 9.0 | | -0.01 |
| | $C_3$ | 38.7 | 56.6 | 37.1 | 0.20 | 4.0 | 6.4 | | 0.48 |
| **K60pAcF/N132C-(DA)** | | | | | | | | | |
| 17 | $C_1$ | 35.9 | 52.8 | 49.8 | 0.47 | 5.8 | 8.1 | 1.04 (0.21) | 0.05 |
| | $C_2$ | 38.5 | 55.9 | 41.4 | 0.22 | 4.8 | 7.2 | | 0.26 |
| | $C_3$ | 37.3 | 54.8 | 36.5 | 0.31 | 4.3 | 6.6 | | 0.42 |
| **K60pAcF/I150C-(DA)** | | | | | | | | | |
| 18 | $C_1$ | 27.1 | 41.1 | 49.3 | 0.35 | 5.6 | 8.1 | 1.07 (0.25) | -0.13 |
| | $C_2$ | 29.2 | 44.7 | 43.3 | 0.22 | 4.9 | 7.2 | | 0.01 |
| | $C_3$ | 29.4 | 44.0 | 36.5 | 0.43 | 4.2 | 6.5 | | 0.18 |
| **Q69pAcF/P86C-(DA)** | | | | | | | | | |
| 19 | $C_1$ | 22.0 | 33.5 | 37.7 | 0.51 | 3.8 | 6.0 | 1.10 (0.47) | -0.12 |
| | $C_2$ | 22.9 | 35.4 | 33.6 | 0.15 | 3.4 | 5.6 | | 0.03 |
| | $C_3$ | 23.2 | 35.3 | 47.7 | 0.34 | 4.8 | 7.2 | | -0.24 |
| **Q69pAcF/R119C-(DA)** | | | | | | | | | |
| 20 | $C_1$ | 27.9 | 41.8 | 36.9 | 0.47 | 4.3 | 6.2 | 1.10 (0.33) | 0.13 |
| | $C_2$ | 28.5 | 42.6 | 45.1 | 0.38 | 5.2 | 7.7 | | -0.10 |
| | $C_3$ | 28.7 | 43.1 | 53.6 | 0.15 | 6.2 | 8.6 | | -0.14 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Q69pAcF/N132C-(DA)** | | | | | | | | | |
| *21\*\*\** | $C_1$ | 31.0 | 48.2 | 48.8 | 0.54 | 5.8 | 8.4 | 1.15 (0.27) | -0.01 |
| | $C_2$ | 30.8 | 45.9 | | | | | | -0.02 |
| | $C_3$ | 30.7 | 46.9 | 37.7 | 0.44 | 4.5 | 6.2 | | 0.24 |
| **Q69pAcF/I150C-(DA)** | | | | | | | | | |
| *22\*\*\** | $C_1$ | 24.7 | 47.3 | 48.7 | 0.58 | 5.4 | 8.0 | 1.12 (0.47) | -0.02 |
| | $C_2$ | 24.3 | 47.0 | | | | | | -0.04 |
| | $C_3$ | 24.8 | 46.5 | 37.9 | 0.42 | 4.2 | 6.6 | | 0.20 |
| **D70pAcF/R119C-(DA)** | | | | | | | | | |
| *23* | $C_1$ | 25.8 | 37.1 | 38.9 | 0.40 | 4.2 | 6.4 | 1.07 (0.13) | -0.04 |
| | $C_2$ | 25.7 | 37.9 | 32.0 | 0.33 | 3.4 | 5.6 | | 0.06 |
| | $C_3$ | 25.9 | 38.4 | 47.7 | 0.27 | 5.1 | 7.2 | | -0.18 |
| **D70pAcF/N132C-(DA)** | | | | | | | | | |
| *24* | $C_1$ | 28.2 | 41.5 | 39.2 | 0.40 | 4.5 | 6.7 | 1.10 (0.16) | 0.05 |
| | $C_2$ | 27.3 | 39.3 | 32.0 | 0.36 | 3.7 | 5.8 | | 0.26 |
| | $C_3$ | 27.2 | 40.8 | 48.8 | 0.24 | 5.7 | 7.7 | | -0.13 |

\* Cα- Cα distance and $\langle R_{DA} \rangle$ of best fit model (172L for $C_1$, 1C67 for $C_2$, 3C7Y for $C_3$)

\*\* definition of $w.res = \dfrac{\langle R_{DA} \rangle_{model} - \langle R_{DA} \rangle_{experiment}}{\Delta R_{DA}^2 \left( \kappa^2, k_{FRET} \right)}$

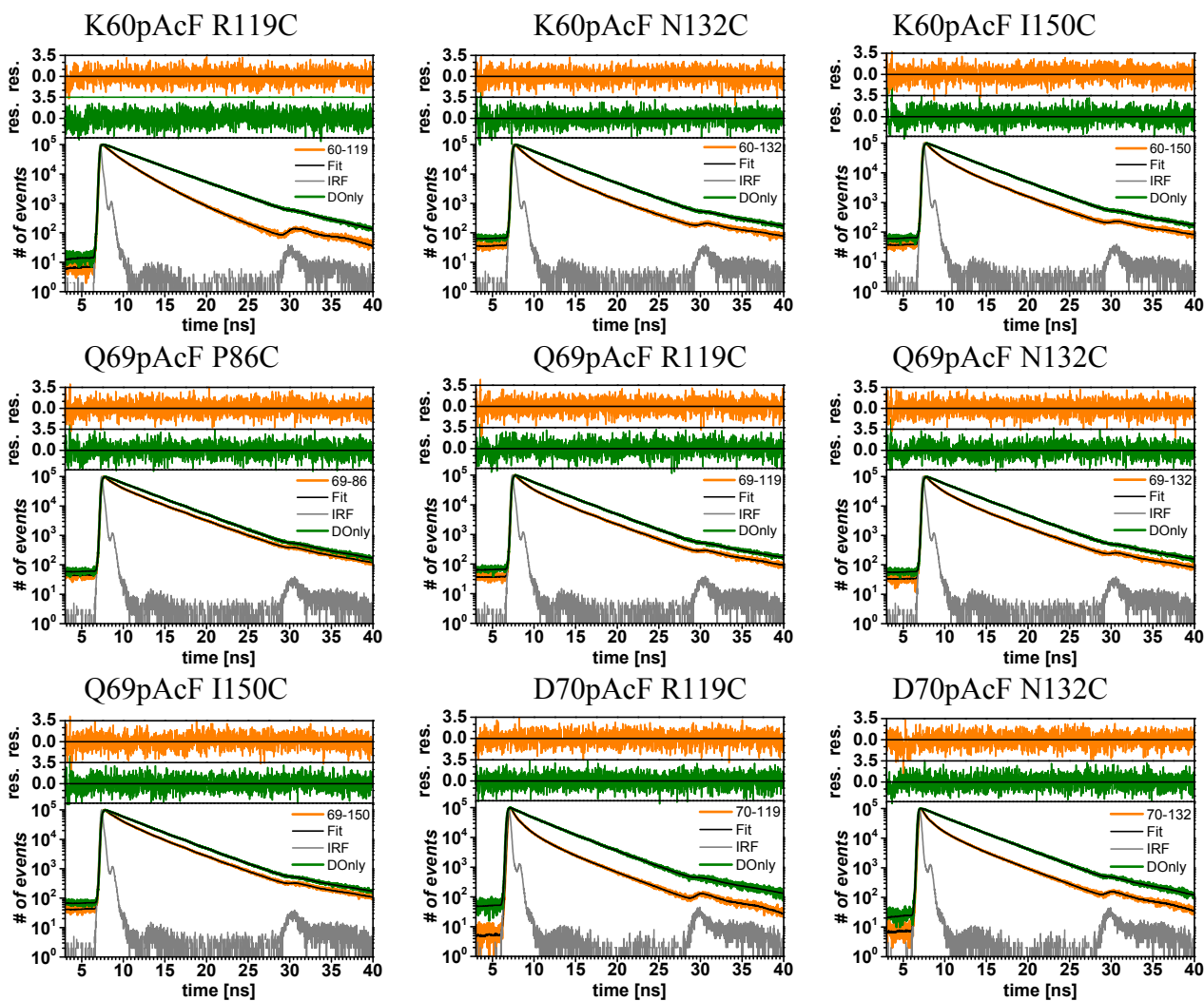\*\*\* fitting with the three states model was not successful.
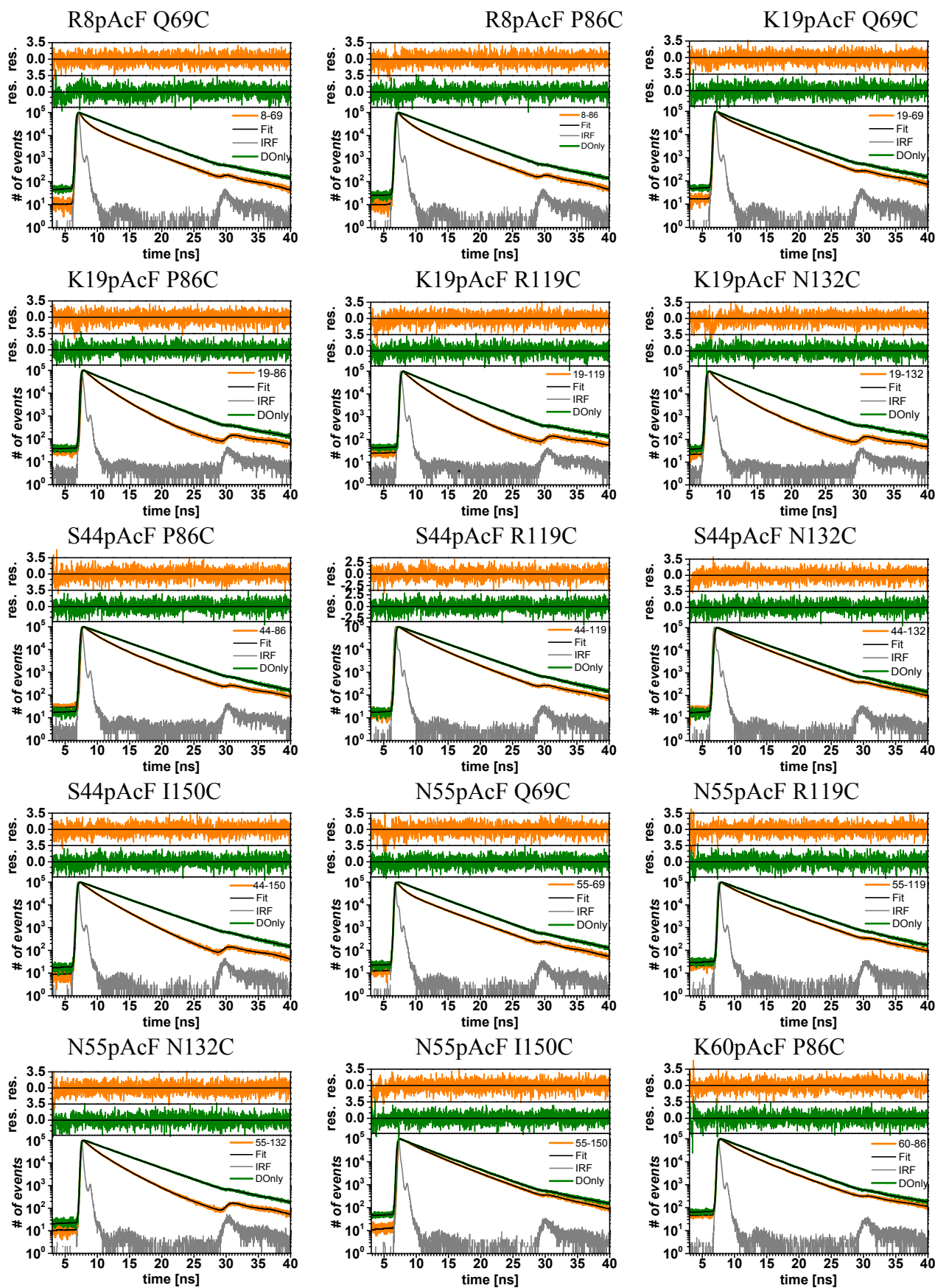
**Figure 4.S2. eTCSPC results.** Fit (black) of the experimental data of double-labeled sample (orange) and respective Donor-only labeled sample (green), weighted residuals are shown on top. Fit parameter are given in Table 4.S1A (Donly) and 4.S1C (double-labeled sample). Instrument response function (IRF) is shown in gray.

Using only the two states model and comparing the modeled distances using PDBID 172L and 148L for the two states showed that our data cannot be correlated with the structural information from the two crystallographic structures (Fig. 4.S3).
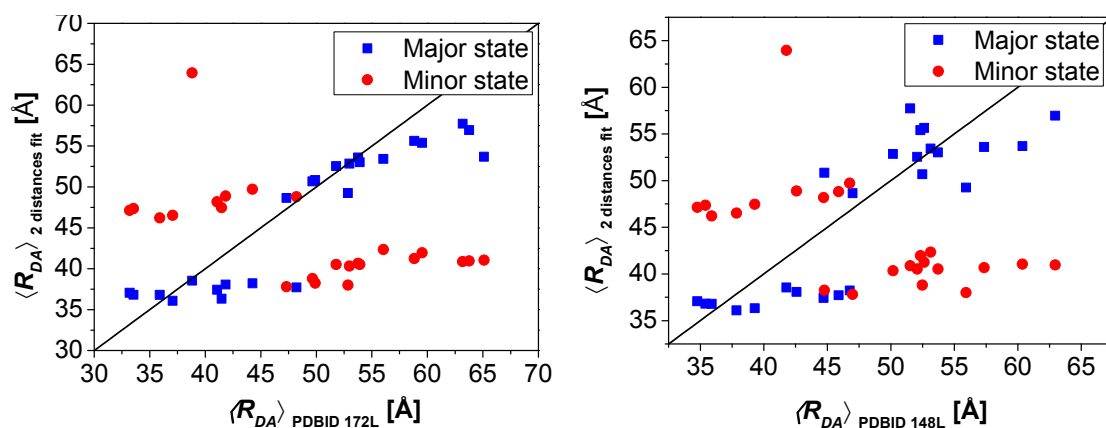


**Figure 4.S3. T4 lysozyme fit with two states compared to simulated model distances. (A)** Fitted distances of two statess fit plotted versus the distances calculated for the model X-ray structure of the open state (PDBID: 172L). "Major state" is the distance having the higher amplitude in fraction, while "minor state" is the distance with the lower fraction **(B)** Same as (A), only for the model X-ray structure of the closed state (PDBID 148L).

Additionally, to evaluate the difference in $\chi^2_r$ between the two states and three states fit statistically, we firstly calculated the ratio $\chi^2_{r-rel} = \chi^2_r(3\,states)/\chi^2_r(2\,states)$ (*F-value*). Second, we calculated the probability *p* which determines the odds with which the three states model describes the data better than the two states model, by calculating cumulative distribution function (CDF) for each single data set by Excel (Microsoft):

$$p = F.DIST = (\text{F-value, dof (2 states), dof (3 states), TRUE}) \tag{4.S32}$$

The degrees of freedom (dof) per data set are three (two states fit) and five (three states fit), respectively (Table 4.S2).

**Table 4.S2** Values used and obtained for the statistical evaluation

| Fit-Model | $\chi^2_{r,avg}$ | curves | Parameters/curve | total # of parameters | # of channels | total # of channels | dof |
|---|---|---|---|---|---|---|---|
| **3 states** | 1.123 | 24 | 5 | 120 | 1800 | 43200 | 43080 |
| **2 states** | 1.174 | 24 | 3 | 72 | 1800 | 43200 | 43128 |
| **F-Value** | 1.046 | | | | | | |
| *p* | 0.999998 | | | | | | |

In a second step, we treated the whole data as a single set and used the global, average $\chi^2_r$ to determine the confidence level with which the two states model has to be refused. The values are summarized in Table 4.S2. To be able to refuse the hypothesis "A two state model is sufficient to describe the data" with a confidence level of 99 %, the F-value of our data should not exceed 1.046 as calculated by Excel (Microsoft):

F.INV = [0.99, dof (2 states), dof (3 states)] = 1.046                    (4.S33)


## 4.2.2.2 Species Cross Correlation Function -(DA) and –(AD) labeled samples.

Theoretically, the species cross-correlation function (sCCF), as defined in Eq. (4.S20), can be extended to more than two species in solution. Practically this suffers of technical limitations. The more species one has in solution the more photons are required to differentiate between them. Therefore, we selected two pseudo-species that represent mixtures of the states found in solution. In addition, we added a third pseudo species that takes into consideration the contribution of scatter photons (Felekyan et al, 2012). In this approach the meaning of specific amplitudes and their relationships is lost; however sCCF can extract the relaxation times as kinetic signatures of conformational transitions between all possible states.

For all data presented, we generated two pseudo-species, plus the addition of the scatter-filter. Decays were generated accordingly to Eq. (4.S26)-(4.S27), based on sub-ensemble burst analysis and eTCSPC data. In some cases, lifetimes of the pseudo-species were adjusted by 100's of ps to properly cross the y-axis of the correlation at a predetermined time for visual comparison. This procedure does not affect the recovered relaxation times.

Considering the case of the double labeled mutants S44pAcF/I150C-(DA) and –(AD), the patterns $p_j^{(i)}$ that correspond to the normalized probability distributions for the –(DA) and –(AD) samples are shown on Fig. 4.S4A,C. The parameters used on the decay generation are shown in the caption. From these patterns, the filters $f_j^{(i)}$ (Eq. (4.S21)) were calculated. These are shown in panels B and D of Fig. 4.S4. These filters are then used to compute the sCCF by multiplying each photon and weighting its contribution to each state as in Eq. (4.S20). The patterns that are shown in Fig. 4.S4 correspond to only to half of the detectors. The other half shows similar patterns. The need of another set of detectors with similar patterns and decays is to increase the amount of pair correlations and to exclude detector after-pulsing related artifacts from calculations. Finally, a full correlation containing all relaxation times and the characteristic diffusion time can be extracted. The reproducibility of the methodology is observed by the overlap of the two species cross-correlations

(Fig. 4.3A, main text), even with the fact that different parameters were used on the generation of the filters. Similar overlap is shown for the mutant S44pAcF/I150C-(DA) and –(AD) at pH 3.0 (Fig. 4.S4E). For the functional mutants (E11A, T26E, R137E) and the variant G77A we show the sCCF (Fig. 4.S4F-I).

### 4.2.2.3 Additional SMD and fFCS

To test for possible influences of the dyes on the protein and the obtained smFRET-/sCCF-results two distinct labeling configurations (DA) and (AD) were prepared as previously described. In the sub μs to ms range the dynamics of T4L is independent of the labeling-configuration. However, we can see some small differences in the two samples. For example, the species fractions in eTCSPC for S44pAcPh/I150C-(DA) and –(AD) are not identical, although, one can clearly identify the same conformers corresponding to the states $C_1$, $C_2$, and $C_3$.

Some slight differences were observed when comparing experiments for -(DA) and –(AD) at the single-molecule level. When comparing the mutant S44pAcF/I150C-(DA) shown in Fig. 4.S7A to the –(AD) labeling scheme showing in Fig. 4.S5, we observe some differences: *i*) There is more "donor-only" fraction in the -(AD) labeling scheme than in the -(DA), this is part of the variability in labeling. *ii*) There is no accumulation of a high FRET state in the –(AD) scheme. However, in this situation the elongation toward higher FRET or state $C_3$ is slightly more pronounced. This elongation is also present in the T26E/S44pAcF/I150C-(AD) mutant (Fig. 4.S5B). This resembles the accumulation found in the sample T26E/S44pAcF/I150C–(DA) (Fig. 4.4G, main text). Regardless of these differences, the 2D histograms and eTCSPC show similar states. This is clear evidence that the three conformational states are present independent of the fluorophores.

In summary, the kinetic scheme might change slightly but not significantly given the conserved effect on the sCCF curves (Fig. 4.3A, main text). The sCCF shows unequivocally that the transition times are present in both labeling schemes. Therefore, the specific dye-protein interactions are not responsible for the transition times between sub μs and ms.

The major difference between the –(DA) and –(AD) is the state $C_{3t}$. This state seems to accumulate for the –(DA) configuration. However, at low pH the –(AD) shows a similar elongation towards the $C_3$ state similar as T26E/S44pAcF/I150C–(AD). Also consistent with the data presented for the S44pAcF/I150C– (DA) at low pH.

S44pAcF/I150C
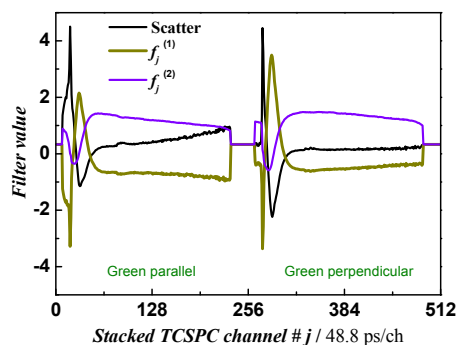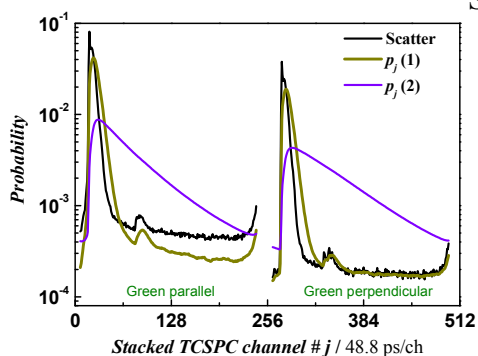
S44pAcF/I150C
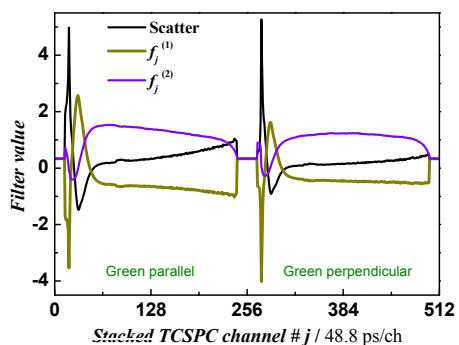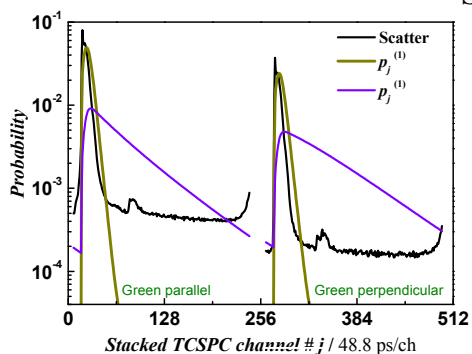
S44pAcF/I150C
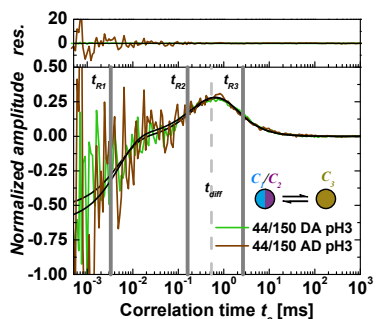
T26E/S44pAcF/I
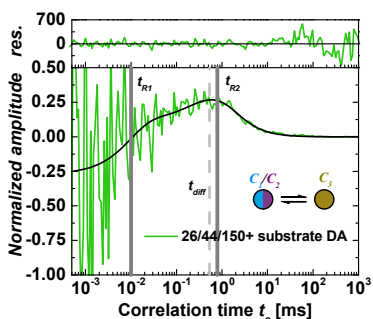
S44pAcF/R119C
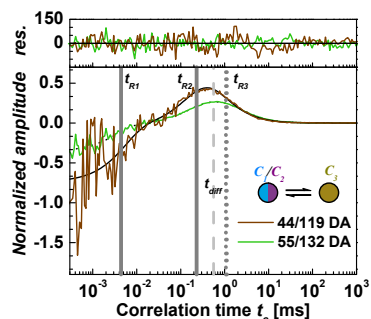
R137E/S44pAcF/I
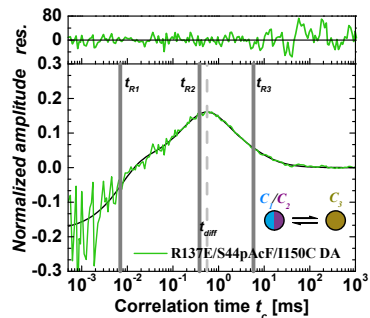
G77A/S44pAcF/I150C-
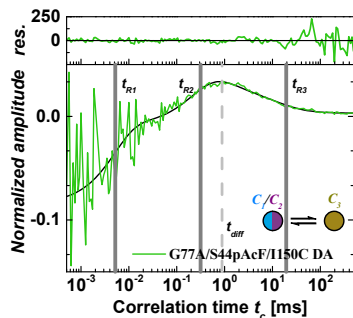
Figure 4.S4 **fFCS results. (A)** Generated decays for two pseudo-species presented for the S44pAcF/I150C-DA in addition to the scatter profile. The parameters of the decay generation for the first pseudo-species were $\tau_1$ = 0.25 ns lifetime, and rotational correlation of $\rho_1$ = 3.3 ns. The second pseudo-species had a lifetime of $\tau_2$ = 2.97 ns and the same rotational correlation $\rho_1$ = 3.3 ns. **(B)** Filters $f_j^{(i)}$ were calculated according to Eq. (S21) using the decays from graph (A). A similar process was done for each fFCS data set. First the decays are generated, than filters are calculated to compute the species cross-correlation function. **(C,D)** Show the decay patterns and the corresponding filters for the S44pAcF/I150C-AD with first pseudo species lifetime $\tau_1$ = 0.25 ns, and rotational correlation $\rho_1$ = 3.3 ns. The second pseudo species was generated with $\tau_2$ = 3.25 ns and same rotational correlation time. Note that each set is treated independently and at the end the sCCF is completely reproducible (Fig. 3A, main text). **(E)** Species cross-correlation between the mix $C_1/C_2$ and the $C_3$ mix at pH 3.0 for the two configurations of labeling –(DA) and –(AD). The fit with Eqs. (S22-S23) required three relaxation times. The diffusion time was fixed to $t_{diff}$ = 0.54 ms. **(F)** Species cross-correlation between the mix $C_1/C_2$ and the $C_3$ mix for T26E/S44pAcF/I150C-(DA) incubated with substrate. Two relaxation times are found ($t_{R1}$= 10 μs, and $t_{R2}$ = 0.790 ms) **(G)** Overlay of the normalized sCCF of S44pAcF/R119C-DA and N55pAcF/N132C-(DA). Global fit shows two common relaxation times ($t_{R1}$ = 4 ± 2.4 μs, $t_{R2}$ = 230 ± 28 μs). The variant N55pAcF/N132C-(DA) requires an additional rate $t_{R3}$ ~ 1.1 ms. **(H)** sCCF for variant R137E/S44pAcF/I150C-(DA). Three relaxation times were needed to fit the curve ($t_{R1}$ = 7 μs, $t_{R2}$ = 0.38 ms and $t_{R3}$ = 5.84 ms). The diffusion time was fixed to $t_{diff}$ = 0.54 ms. **(I)** sCCF for variant G77A/S44pAcF/I150C-(DA). Three relaxation times were needed to fit the curve ($t_{R1}$ = 5 μs, $t_{R2}$ = 0.34 ms and $t_{R3}$ = 11.5 ms). The diffusion time was $t_{diff}$= 0.86 ms.

Additional MFD histograms for other mutants mentioned in the main text are shown in Fig. 4.S5D-J. A summary of the ensemble or subensemble fits for these mutants is shown on Fig. 4.S5K.

## 4.2.2.4 Catalytic activity of T4L mutants

The ability to process the selected substrate (peptidoglycan from *Micrococcus luteus*) of the mutants was monitored by reverse phase chromatography. Double-labeled mutants (1 μM) were incubated with 3mg/ml of substrate and allowed to react for several hours in 50 mM sodium phosphate buffer, 150 mM NaCl at pH 7.5. Samples at different times were monitored under a reverse phase HPLC at 495 nm. In this way we can identify the labeled lysozyme. Typical examples for the processing of substrate are shown in Figure 4.S5. Fig. 4.S5A shows the elution profile of the peptidoglycan monitored at 215 nm. Multiple peaks from 10 to 14 min appear. In the same panel the elution of the T26E/S44pAcF/I150C-(AD) monitored at 215 nm is shown as incubated with the peptidoglycan. For better contrast of the shift in populations the absorbance was measured at the maximum for the AlexaFluor488 (495 nm). This is shown in Fig. 4.S6B. After 260 minutes this mutant is fully saturated with the substrate.
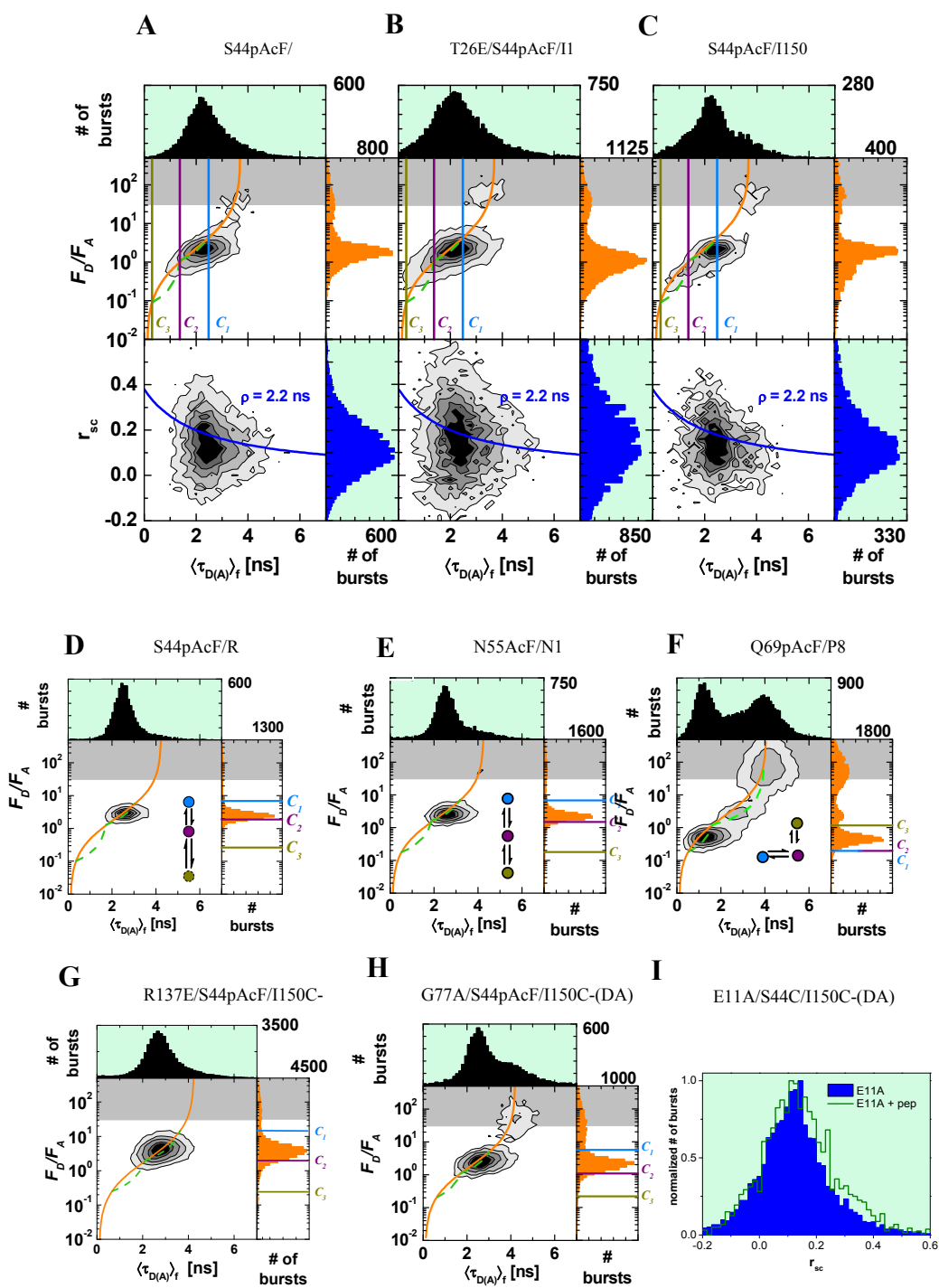
**K**

| Samples (showing aa #'s) | $\langle R_{DA1}\rangle$ [Å] | $x_1{}^\dagger$ | ± err. (%) | $\langle R_{DA2}\rangle$ [Å] | $x_2{}^\dagger$ | ± err. (%) | $\langle R_{DA3}\rangle$ [Å] | $x_3{}^\dagger$ | ± err. (%) | $x_{D(0)}{}^*$ | $\chi^2_r$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 44-150-(DA)(-) | 65.1 | 0.25 | 8.0 | 51.7 | 0.55 | 8.0 | 38.8 | 0.20 | 3.0 | 0.01 | 1.06 |
| 44-150-(DA)(+)*** | 65.1 | 0.25 | 12.3 | 51.7 | 0.57 | 12.3 | 38.8 | 0.18 | 3.9 | - | 1.52** |
| 26/44/150-(DA)(-)*** | 65.1 | 0.37 | 4.7 | 51.7 | 0.35 | 3.0 | 34.9 | 0.28 | 9.0 | 0.62 | 1.21 |
| 26/44/150-(DA)(+)*** | 65.1 | 0.20 | 1.9 | 51.7 | 0.28 | 1.0 | 34.9 | 0.52 | 13.0 | 0.74 | 1.08 |
| 11/44/150-(DA/AD)(-)*** | 65.1 | 0.75 | 7.8 | 51.7 | 0.12 | 7.8 | 38.8 | 0.13 | 11.8 | - | 1.98** |
| 11/44/150-(DA/AD)(+)*** | 65.1 | 0.56 | 4.9 | 51.7 | 0.27 | 4.9 | 38.8 | 0.17 | 17.4 | - | 2.00** |
| 137/44/150-(DA) | 59.3 | 0.52 | 7.3 | 49.3 | 0.37 | 2.6 | 36.2 | 0.11 | 9.1 | 0.24 | 1.07 |
| 77/44/150-(DA) | 65.1 | 0.35 | 7.9 | 51.7 | 0.37 | 11.0 | 42.6 | 0.29 | 2.5 | 0.56 | 1.19 |

Figure S5 **MFD analysis of selected samples. (A)** MFD histogram of S44pAcF/I150C-(AD) labeled T4L. Two dimensional histogram $F_D/F_A$ vs. lifetime of donor in the presence of acceptor $\langle\tau_{D(A)}\rangle_f$, and anisotropy vs. $\langle\tau_{D(A)}\rangle_f$ for S44pAcF/I150C-(AD). One dimensional projections for $F_D/F_A$, $\langle\tau_{D(A)}\rangle_f$ and anisotropy are also shown. FRET line (Eq. (S11)) is shown in orange. Pure donor and acceptor fluorescence ($F_D$ and $F_A$) are corrected for background ($\langle B_G\rangle$ = 1.8 kHz, $\langle B_R\rangle$ = 0.7 kHz), spectral cross-talk ($\alpha$ = 1.3%) and detection efficiency ratio ($g_G/g_R$ = 0.77). Shaded area in gray is the region of donor only. On the anisotropy vs. $\langle\tau_{D(A)}\rangle_f$ histograms the Perrin's equation with rotational correlation $\rho$ = 2.2 ns is shown as blue line. Vertical guidelines for states $C_1$, $C_2$, and $C_3$ according to the eTCSPC results of the same sample are added as references. Ignoring the donor only population a single unimodal distribution is observed in all $F_D/F_A$ vs. $\langle\tau_{D(A)}\rangle_f$, similarly to what was observed in the –(DA) sample. Two slight differences can be observed: the tilt towards the state $C_3$ is more evident and the accumulation of the $C_{3t}$ is not visible. **(B)** MFD histograms for the variant T26E/S44pAcF/I150C-(AD) without substrate. We observe a more pronounced tilt towards the $C_3$. **(C)** At pH 3.0, the MFD histograms for the S44pAcF/I150C-(AD) show very similar characteristics as the variant T26E. **(D-F) Examples results for further variants of T4L.** MFD histograms for S44pAcF/R119C-(DA), N55pAcF/N132C-(DA) and Q69pAcF/P86C-(DA). For variant Q69pAcF/P86C-(DA). This can be expected when the acceptor is no longer functional. **(G-H) Functional mutants of T4L.** MFD histograms for R137E/S44pAcPh/I150C-(DA) and G77A/S44pAcPh/I150C-(DA). **(I) Effect of substrate on E11A/S44C/I150C.** Upon addition of substrate we observe a higher anisotropy (green line). All samples were corrected for background, cross talk, and detection efficiencies according to experimentally determined parameters. **(K) Fit results for functional variants.** Globally fitted values are shown in gray cells. $^\dagger$ Values for the FRET populations are normalized such that $x_1+x_2+x_3$ =1. *Donor decay was fixed and $x_{D(0)}$ represents the fraction of donor only from the total. ** Data from single molecule experiments shows higher $\chi^2_r$ when compared to eTCSPC, due to low photon statistics. ***Sub-ensemble fit from burst analysis. For E11A/S44C/I150C-(DA), it was not possible to measure in eTCSPC due to high donor-only (double Cys variant).
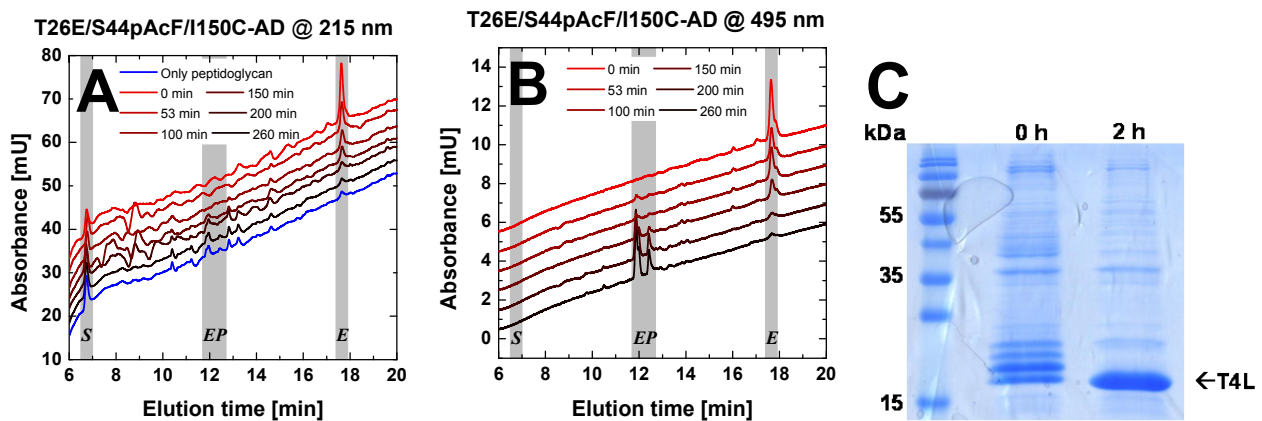
Figure 4.S6 **T4L binding to peptidoglycan as observed by reverse phase HPLC and cleavage at low pH. (A)** The elution profile measured at 215 nm of reverse phase chromatography with a C-18 column for the T26E/S44pAcF/I150C-(AD). Samples were taken at different times during the incubation with peptidoglycan. First line shows only the elution of peptidoglycan. Note different peaks of the heterogeneity on the peptidoglycan. Offset between lines was added for clarity. Gray lines represent the free enzyme population (*E*), the product bound enzyme (*EP*) and the substrate alone (*S*). **(B)** Elution of the same sample as in **(A)** but monitored at 495 nm, which corresponds to the absorbance of Alexa Fluor 488. Saturation of the T26E/S44pAcF/I150C-(AD) is reached at ~ 4 hours of incubation. **(C)** Purification of T26E/S44pAcF/I150C from the *E. coli* cell pellet yielded a mixture of free and to cell wall pieces of different sizes bound protein. After incubation for 2 hrs at pH 3, nearly all bound peptidoglycan had been cleaved and the free enzyme could now be used for labeling and further experiments after adjusting the conditions to neutral pH again.

## 4.2.2.5 Consolidated model of T4L

To construct the best kinetic model that describes the free enzyme in solution let us consider the experimental facts: *i*) eTCSPC resolves three different FRET states. *ii*) fFCS shows two transition times faster than 10 ms. *iii*) smFRET diagrams are better described by a unimodal distribution mixed with a small population (~ 5-10%) with very high FRET only for the S44pAcF/I150C-(DA) variant.

Unimodal distributions in single-molecule experiments can occur due to time-averaging. Ignoring the donor only population, the free enzyme (S44pAcF/I150C-(DA)) samples four conformational states ($C_1$, $C_2$, $C_{3d}$ and $C_{3t}$), where the $C_{3t}$ is a static population at very high FRET, and $C_1$, $C_2$, $C_{3d}$ mix at the observed times of ~4 μs and ~230 μs.

Putting aside the state $C_{3t}$, the simplest model of conformational transitions that one can build from experimental observables is

$$[C_1] \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} [C_2] \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} [C_3]$$

(4.S32)

where $C_1$ corresponds to the most open conformer, $C_2$ is similar to the substrate-enzyme complex and $C_3$ has an interdye distance much shorter than $C_2$. With this in mind we disregarded the cyclic model

$$[C_1] \rightleftharpoons [C_2] \quad \text{with} \quad [C_3]$$

(4.S33)

due to the sequential closing of the enzyme or microscopic reversibility. This limits the return process of $[C_1] \rightleftharpoons [C_3]$.

Our goal is to extract the reaction rates ($k_{12}$, $k_{21}$, $k_{23}$, $k_{32}$) from our experimental observables. To solve this, first we need to write the rate matrix $K$ for the system described in Eq. (4.S32).

$$K = \begin{pmatrix} -k_{12} & k_{21} & 0 \\ k_{12} & -(k_{21}+k_{23}) & k_{32} \\ 0 & k_{23} & -k_{32} \end{pmatrix}$$

(4.S34)

The two eigenvalues of $K$ correspond to the two observables measured by fFCS.

$$\frac{1}{t_{R1,2}} = \frac{1}{2} \cdot (k_{12}+k_{21}+k_{23}+k_{32}) \pm \sqrt{(k_{12}+k_{21}+k_{23}+k_{32})^2 - 4 \cdot (k_{12}k_{23}+k_{12}k_{32}+k_{21}k_{32})}$$

(4.S35)

The time evolution of the system on Eq. (4.S32) is defined by

$$\frac{d}{dt}[C_i](t) = K_{ij}[C_i](t),$$

(4.S36)

Equation (4.S35) has an analytical solution on the form of

$$[C_i](t) = \exp(K_{ij}t).$$

(4.S37)

At equilibrium, or $t \rightarrow \infty$, the equilibrium fractions for each conformer can be obtained analytically and are given by

$$[C_i] = \begin{pmatrix} \dfrac{k_{21} \cdot k_{32}}{k_{21} \cdot k_{32} + k_{12} \cdot (k_{23} + k_{32})} \\[2mm] \dfrac{k_{12} \cdot k_{32}}{k_{21} \cdot k_{32} + k_{12} \cdot (k_{23} + k_{32})} \\[2mm] \dfrac{k_{12} \cdot k_{23}}{k_{21} \cdot k_{32} + k_{12} \cdot (k_{23} + k_{32})} \end{pmatrix}. \tag{4.S38}$$

Note that $[C_3] = 1 - ([C_1] + [C_2])$. These fractions are obtained by fluorescence decay analysis as done in Section 4.2.1.11. The reaction rates ($k_{12}$, $k_{21}$, $k_{23}$, $k_{32}$) can be expressed in terms of the equilibrium fractions ($x_1 = [C_1]$, $x_2 = [C_2]$, $x_3 = [C_3]$) and the relaxation times $t_{R1}$ and $t_{R2}$.

The analytical solution of this system has two solutions:

$$k_{12}{}^{(\pm)} = \frac{[C_2] \cdot \left( \dfrac{1}{t_{R1}} + \dfrac{1}{t_{R2}} \right) \pm \left( [C_2] \cdot \left( \dfrac{4 \cdot ([C_1] - 1) \cdot ([C_1] + [C_2])}{t_{R1} \cdot t_{R2}} + [C_2] \cdot \left( \dfrac{1}{t_{R1}} + \dfrac{1}{t_{R2}} \right)^2 \right) \right)^{1/2}}{2 \cdot ([C_1] + [C_2])},$$

$$k_{21}{}^{(\pm)} = \frac{[C_1] \cdot \left( [C_2] \cdot \left( \dfrac{1}{t_{R1}} + \dfrac{1}{t_{R2}} \right) \pm \left( [C_2] \cdot \left( \dfrac{4 \cdot ([C_1] - 1) \cdot ([C_1] + [C_2])}{t_{R1} \cdot t_{R2}} + [C_2] \cdot \left( \dfrac{1}{t_{R1}} + \dfrac{1}{t_{R2}} \right)^2 \right) \right)^{1/2} \right)}{2 \cdot [C_2] \cdot ([C_1] + [C_2])},$$

$$k_{23}{}^{(\pm)} = \frac{([C_1] + [C_2] - 1) \cdot \left( [C_2] \cdot \left( \dfrac{1}{t_{R1}} + \dfrac{1}{t_{R2}} \right) \mp \left( [C_2] \cdot \left( \dfrac{4 \cdot ([C_1] - 1) \cdot ([C_1] + [C_2])}{t_{R1} \cdot t_{R2}} + [C_2] \cdot \left( \dfrac{1}{t_{R1}} + \dfrac{1}{t_{R2}} \right)^2 \right) \right)^{1/2} \right)}{2 \cdot [C_2] \cdot ([C_1] - 1)},$$

$$k_{32}{}^{(\pm)} = \frac{-[C_2] \cdot \left( \dfrac{1}{t_{R1}} + \dfrac{1}{t_{R2}} \right) \pm \left( [C_2] \cdot \left( \dfrac{4 \cdot ([C_1] - 1) \cdot ([C_1] + [C_2])}{t_{R1} \cdot t_{R2}} + [C_2] \cdot \left( \dfrac{1}{t_{R1}} + \dfrac{1}{t_{R2}} \right)^2 \right) \right)^{1/2}}{2 \cdot ([C_1] - 1)} \tag{4.S39}$$

To complete the model we need to add the static fraction of ~ 5-10%. We assigned this static fraction to conformer $C_{3t}$, which is identical in FRET to the state $C_{3d}$. We split the fraction of $C_3$ into these two populations. The final reaction model can be expressed as

$$[C_1] \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} [C_2] \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} [C_{3d}] \underset{k_{43}}{\overset{k_{34}}{\rightleftharpoons}} [C_{3t}] \tag{4.S40}$$

Where $k_{34} = 0.003$ ms$^{-1}$ and $k_{43} = 0.008$ ms$^{-1}$, were empirically determined, but satisfy the condition that they have to be smaller than 0.01 ms$^{-1}$.

With all the determined rates we did Brownian dynamics simulations as described in main text. The single-molecule MFD histograms for the data presented in Fig 4.3 is shown in Fig. 4.S7.

## 4.2.2.5.1 Simulation of the FRET data in complex kinetic schemes

To describe the experimental 2D histogram a four-state scheme was used. As the first step, we calculated FRET histograms (Gopich & Szabo, 2012a). A kinetic model with discrete conformations was assumed, with the transition between the states described by rate equations. The probability for the system to be in state $i$ at time $t$, $p_i(t)$, satisfies a set of rate equations, which can be written in matrix notation as:

$$\frac{d\boldsymbol{p}}{dt} = \boldsymbol{K} \cdot \boldsymbol{p} \tag{4.S41}$$

where $\boldsymbol{p}$ is a column vector with the components $p_i(t)$ and $\boldsymbol{K}$ is a transition rate matrix representing the rate constants for the transitions between states $i$ and $j$. At long times, $\boldsymbol{p}(t)$ approaches its equilibrium value, $\boldsymbol{p}_{eq}$. The vector of the equilibrium populations $\boldsymbol{p}_{eq}$ is normalized to 1 and satisfies $\boldsymbol{K} \cdot \boldsymbol{p}_{eq} = 0$. For each burst mean averaged efficiency $\langle E \rangle$ and average fluorescence weighted lifetime $\langle \tau_{D(A)} \rangle_f$ can be calculated by:

$$\langle E \rangle = \frac{\sum t_i \cdot E_i}{t_{burst}} \tag{4.S42}$$

and

$$\left\langle \tau_{D(A)} \right\rangle_f = \frac{\sum t_i(\boldsymbol{K}) \cdot \tau_i^2}{\sum t_i(\boldsymbol{K}) \cdot \tau_i} \tag{4.S43}$$

where $t_i(\boldsymbol{K})$ is time spent by a molecule in state $i$ within the duration of the burst, which depends on the transition rate matrix $\boldsymbol{K}$; $E_i$ is the FRET efficiency of the $i$-th state; $t_{burst}$ is the duration of the burst and $\tau_i$ is the fluorescence lifetime of the $i$-th state. Practically, each burst has certain duration and number of photons, which were chosen arbitrary from experimentally measured $t_{burst}$ (duration time) vs. $N$ (number of photons) 2D histogram. The residence times spent by the molecule in

different states were calculated using Gillespie algorithm for continuous-time Markov Chain. Then, average fluorescence lifetime $\langle \tau_{D(A)} \rangle_f$ for each burst was calculated by Monte-Carlo simulation of fluorescence emission given FRET efficiencies of each state, stationary (equilibrium) populations of states were obtained by solving interstate transition dynamics matrix and the residence times obtained on previous step. The descriptions for the vector $\boldsymbol{p}$ and the rate matrix $\boldsymbol{K}$ (resulting into the equilibrium fractions for the state i, $\boldsymbol{p_{eq,i}}$) and the experimental observables, $E$ and $\tau$, used in the simulations to recapture the experimental 2D histogram are shown in the Table 4.S2. $E$ was in the following converted to $F_D/F_A$ ratio. The simulation procedure was repeated for a high number of bursts to generate $F_D/F_A$ vs $\langle \tau_{D(A)} \rangle_f$ 2D histogram (Fig. 4.S7).
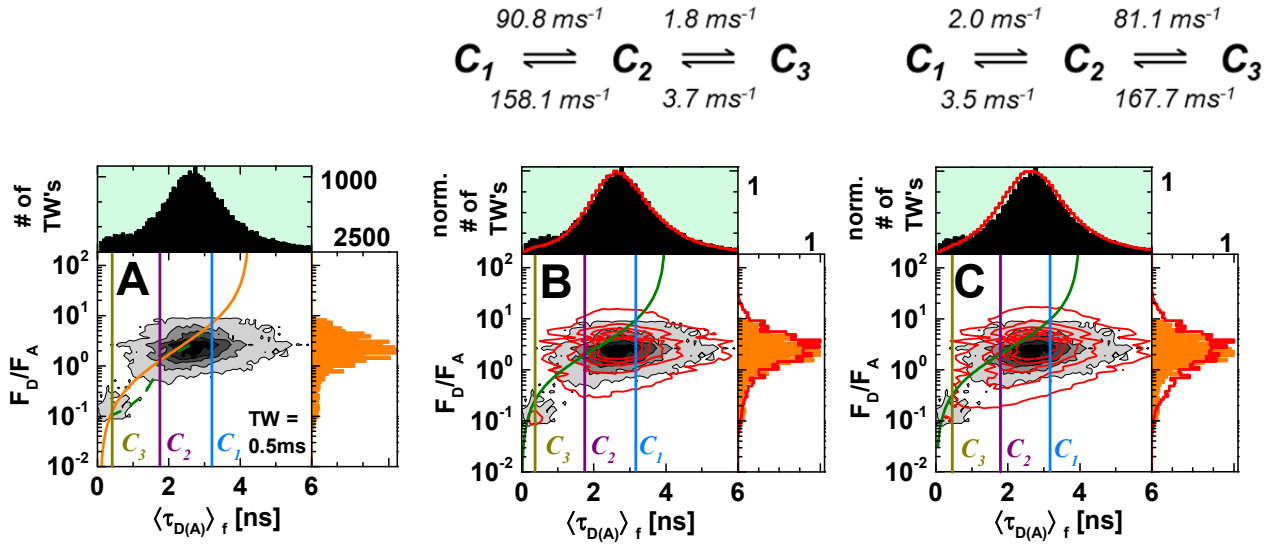
Figure 4.S7 **Brownian Dynamic Simulations of S44pAcF/I150C-(DA).** (**A**) Two dimensional histograms from smFRET analysis ($F_D/F_A$ vs. fluorescence lifetime ($\langle \tau_{D(A)} \rangle_f$) and anisotropy $r_{sc}$ vs. $\langle \tau_{D(A)} \rangle_f$) of the raw data from the S44pAcF/I150-(DA) variant at pH 7.5 selected with 0.5 ms time-windows (TW's). FRET lines, static and dynamic are shown as green dashed lines. $\langle B_G \rangle = 1.6$ kHz, $\langle B_R \rangle = 0.8$ kHz, spectral crosstalk $\alpha = 1.2\%$ and ratio of green and red detection efficiencies $g_G/g_R = 0.77$ are used for corrections. (**B,C**) Brownian dynamics simulation using the rates from Fig. 4.3B,C was processed as the experimental data. Simulated parameters ($\langle B_G \rangle$, $\langle B_R \rangle$, $\alpha$, $g_G/g_R$) were the same as in the experiment. In addition we considered a rotational correlation of $\rho = 2.2$ ns for conformational state. Analysis results of simulated data are presented in the same fashion as in panel (**A**).

Table 4.S2 **States description for the vector *p*, equilibrium faction vector $p_{eq}$ and the rate matrix *K*.**

| Kinetic state i | State name | Efficiency $E$ | Fluorescence lifetime $\tau$, ns [a] | Equilibrium fraction $p_{eq,i}$ |
|---|---|---|---|---|
| 1 | $C_1$ | 0.2 | 3.2 | 0.51 |
| 2 | $C_2$ | 0.5 | 2.0 | 0.29 |
| 3 | $C_3$ | 0.9 | 0.5 | 0.14 |
| 4 | $C_{3d}$ | 0.9 | 0.5 | 0.06 |

$$K_{fastslow} = \begin{pmatrix} -158.1 & 90.8 & 0 & 0 \\ 158.1 & -94.5 & 1.8 & 0 \\ 0 & 3.7 & -1.808 & 0.003 \\ 0 & 0 & 0.008 & -0.003 \end{pmatrix} \quad K_{slowfast} = \begin{pmatrix} -3.5 & 2 & 0 & 0 \\ 3.5 & -169.7 & 81.1 & 0 \\ 0 & 167.7 & -81.108 & 0.003 \\ 0 & 0 & 0.008 & -0.003 \end{pmatrix}$$

To estimate our errors on determining the rates we considered the 2σ confidence interval in determining the population fractions (Fig 4.S5I) and the 2σ confidence interval in determining the relaxation times by fFCS. Taking those extremes we estimated the error and computed the reaction rate constant for Fig. 4.6 in main text according to Eq. 4.S39.

Table 4.S3 **Calculated reaction rates for several variants using Eq. (4.S39)**. Confidence intervals ($2\sigma$) are shown in squared brackets and the corresponding renormalized fractions shown below $x_1+x_2+x_{3d}=1$*.

| Samples | $k_{12}$ [ms$^{-1}$] | $k_{21}$ [ms$^{-1}$] | $k_{23}$ [ms$^{-1}$] | $k_{32}$ [ms$^{-1}$] |
|---|---|---|---|---|
| 44/150-(DA) | 90.8 [45.5-175.6] | 158.1 [74.1-312.7] | 1.8 [1.5-2.3] | 3.7 [3.1-4.3] |
| 11/44/150-(DA)+pept | 217 [134.8-535.4] | 33 [19.0-86.38] | 0.8 [0.6-1.1] | 3.7 [3.1-4.3] |
| 26/44/150-(DA) + pept | 79 [49.2-193.3] | 21 [13.0-52.6] | 0.5 [0.4-0.6] | 0.9 [0.8-1.0] |
| The relaxations times used were: $t_{R1} = 4 \pm 2.3\ \mu s$; $t_{R2} = 230 \pm 28.4\mu s$ for 44/150-(DA) and 11/44/150-(DA)+pept. $t_{R1} = 10\ \mu s$; $t_{R2} = 790\ \mu s$ (Fig. S4F) was used for 26/44/150-(DA) + pept. | | | | |

| Chemical State | Samples | $x_1$ | $x_2$ | $x_{3d}$ |
|---|---|---|---|---|
| E | 44/150-(DA) | 0.54 | 0.30 | 0.16 |
| ES | 11/44/150-(DA)+pept | 0.30 | 0.54 | 0.16 |
| EP | 26/44/150-(DA) +pept | 0.35 | 0.29 | 0.36 |
| * Rounded to 2 digits. Renormalized fractions based on the relative changes observed in all states in the presence of substrate (Fig. 5E). Only the amino acid number of the mutagenesis is shown. | | | | |

## 4.2.2.6 Challenges of smFRET measurements and their solutions

Most of the potential problems with smFRET come from the complexities associated with the labels. Here we list the solution for potential label artifacts, and how our approaches and considerations allow us to draw conclusion, artifact free, of our data.

1) Labeling influence on enzymatic work:

HPLC on the T26E/S44pAcF/150C-(DA) and –(AD) mutants show that they can process the peptidoglycan to keep the substrate bound. Non-functional mutants stayed non-functional after labeling (E11A/S44C/I150C-(DA) and R137E/S44pAcF/I150C-(DA)).

2) Local quenching of Donor:

eTCSPC: In ensemble measurements local quenching is observed by changes in the average lifetime of the donor. The multi-exponential fluorescence decays of the donor only labeled variants reflect the presence of quenched states. At these states the fluorophore senses a different environment. Most likely these differences represent various conformations of the protein.

smFRET: Donor quenching, as in the case of eTCSPC, shifts the average donor lifetime towards shorter lifetimes. FRET lines are corrected for the multi-exponential properties of the donor decay.

fFCS: We use the multi exponential time-resolved fluorescence decay information to generate different filters to calculate the species cross-correlation. Although, protein dynamics can be extracted from single label variants, the structural information is lost. This is only possible from the FRET labeled samples.

3) Triplet-state of Donor:

eTCSPC: Triplet state is long lived compared to the fluorescence lifetime. Therefore, on ensemble time-resolved fluorescence decays this effect is not visible.

smFRET: Triplet or dark states kinetics are short lived compared to the burst duration.

fFCS: In a classical FCS experiment triplet or dark states appear as a bunching term in the correlation function. In fFCS we do not correlate fluctuations on signal but rather we correlate fluctuations of species. In our case they correspond to different conformations of T4L. We assume that triplet/dark states are not coupled to the conformations or the selected pseudo species. In other words, the photo-physics of the dye is independent of the conformation in which the molecule is. With this in mind, the sCCF will have positive and negative contributions from each species resulting in the fact that the bunching term is not present. We know that increasing the power can increase the triplet amplitude. To test this, we measured the sCCF of T4L-(DA) at different powers at objective and we did not observe any major differences in the relaxation times $t_{R1}$ and $t_{R2}$ (Fig. 4.S8) or shape of the sCCF. We also tested the addition of triplet quenchers Cycloocta-1,3,5,7-tetraenecarboxylic acid (COTc) but did not observe major deviations (Fig. 4.S8A-D).

4) Acceptor cis-trans isomerization:

eTCSPC: If FRET to *cis* and *trans* is different the donor decay would reflect the *cis-trans* population. We assume that this effect is small therefore not visible.

smFRET: This effect can be observed as acceptor quenching. The reason is that the *cis* state is dark. Spending more time in the *cis* state will reduce the overall counts observed from the acceptor. This effect can be seen in the two dimensional histograms as a vertical shift of the islands position on $F_D/F_A$ vs. lifetime $\langle \tau_{D(A)} \rangle_f$ representation.

fFCS: For fFCS we correlate only photons emitted by the donor fluorophore. Changes in the brightness of the acceptor are not correlated. However something that can happen is that the absorption of the energy transferred from the donor can be different for *cis* and *trans* states. This is something that was not tested. But as in the case of the donor triplet we assume that, even in the case in which this occurs, the photophysics dynamics of the acceptor dye is decoupled from the conformational dynamics of the molecule.

5) Dye mobility:

eTCSPC: Dye mobility occurs at slower timescales than the time-resolved fluorescence decay of the fluorophore. For this reason it is better to consider FRET due to all configurations of fluorophore positions during time-resolved fluorescence decays. We take into consideration this by having a distribution of lifetimes instead of single lifetimes to identify each conformational state. These are included in the treatment of the FRET lines. In order to do so, ensemble time resolved anisotropy decays were measured. We assumed that fluorophore mobility follows the "wobble in a cone" model (Dale et al, 1979). Table S5A-C summarizes the residual anisotropies ($r_\infty$) of D - donor, A - acceptor and A(D) - the sensitized by FRET emission of acceptor that were used to calculate dye order parameters and $\kappa^2$ distributions (Table 4.S5D) according to refs. (Dale et al, 1979; Sindbert et al, 2011a) (Eq. 9 and 10 in Sindbert *et al.*). The assumption is that fluorophores move according to the "wobble in a cone" model. According to all distributions the assumption of $\kappa^2 = 2/3$ is very well justified.

smFRET: In smFRET one can inspect the anisotropy $r_{sc}$ vs. lifetime $\langle \tau_{D(A)} \rangle_f$ histograms. If anisotropy is too high then one would expect that the dye can have restricted mobility.

fFCS: The mobility of the dye alone is better resolved using a complete FCS technique (Felekyan et al, 2005).

Table 4.S4A **Analysis of time-resolved fluorescence anisotropies *r(t)* for donor only labeled samples [a] obtained by ensemble time-resolved fluorescence decays as described in Section 4.2.1.6.**

| Samples (showing aa #'s) | $r_{1, D}$ | $\rho_{1,D}$ | $r_{2, D}$ | $\rho_{2,D}$ | $\Sigma r$ |
|---|---|---|---|---|---|
| 8/69-(D(0)) | 0.08 | 0.62 | 0.25 | 8.80 | 0.33 |
| 8/86-(D(0)) | 0.08 | 0.78 | 0.24 | 8.66 | 0.32 |
| 19/69-(D(0)) | 0.14 | 0.60 | 0.19 | 6.23 | 0.34 |
| 19/86-(D(0)) | 0.14 | 0.60 | 0.19 | 6.23 | 0.34 |
| 19/119-(D(0)) | 0.14 | 0.60 | 0.19 | 6.23 | 0.34 |
| 19/132-(D(0)) | 0.09 | 0.52 | 0.25 | 9.63 | 0.34 |
| 44/86-(D(0)) | 0.14 | 0.64 | 0.18 | 6.68 | 0.32 |
| 44/119-(D(0)) | 0.14 | 0.64 | 0.18 | 6.68 | 0.32 |
| 44/132-(D(0)) | 0.14 | 0.64 | 0.18 | 6.68 | 0.32 |
| 44/150-(D(0)) | 0.11 | 0.59 | 0.21 | 7.93 | 0.32 |
| 55/69-(D(0)) | 0.15 | 0.48 | 0.17 | 8.03 | 0.32 |
| 55/119-(D(0)) | 0.15 | 0.48 | 0.17 | 8.03 | 0.32 |
| 55/132-(D(0)) | 0.15 | 0.48 | 0.17 | 8.03 | 0.32 |
| 55/150-(D(0)) | 0.15 | 0.48 | 0.17 | 8.03 | 0.32 |
| 60/86-(D(0)) | 0.11 | 0.58 | 0.22 | 7.42 | 0.33 |
| 60/119-(D(0)) | 0.11 | 0.58 | 0.22 | 7.42 | 0.33 |
| 60/132-(D(0)) | 0.11 | 0.58 | 0.22 | 7.42 | 0.33 |
| 60/150-(D(0)) | 0.11 | 0.58 | 0.22 | 7.42 | 0.33 |
| 69/86-(D(0)) | 0.14 | 0.54 | 0.18 | 7.39 | 0.32 |
| 69/119-(D(0)) | 0.14 | 0.54 | 0.18 | 7.39 | 0.32 |
| 69/132-(D(0)) | 0.14 | 0.54 | 0.18 | 7.39 | 0.32 |
| 69/150-(D(0)) | 0.14 | 0.54 | 0.18 | 7.39 | 0.32 |
| 70/119-(D(0)) | 0.13 | 0.62 | 0.20 | 7.21 | 0.33 |
| 70/132-(D(0)) | 0.13 | 0.62 | 0.20 | 7.21 | 0.33 |

[a]: The fluorescence anisotropy decay *r(t)* can be described as a sum of two exponentials:

$$r(t) = r_1 \exp(-t / \rho_1) + r_2 \exp(-t / \rho_2) \text{ with } r_1 + r_2 \leq r_0$$

For Alexa488-hydroxylamide the fundamental anisotropy $r_{0,D}$ is 0.375, for Alexa647-maleimide $r_{0,A}$ is 0.39 and for FRET-sensitized anisotropy decay $r_{0,A(D)}$ is 0.38.

Table 4.S4B **Analysis of time-resolved fluorescence anisotropies r(t) for direct acceptor excitation of double labeled samples obtained by ensemble time-resolved fluorescence decays as described in Section 4.2.1.6.**

| Samples (showing aa #'s) | $r_{1, A}$ | $\rho_{1,A}$ | $r_{2, A}$ | $\rho_{2,A}$ | $\Sigma r$ |
|---|---|---|---|---|---|
| 8/69-(DA) | 0.13 | 0.66 | 0.22 | 11.28 | 0.36 |
| 8/86-(DA) | 0.17 | 0.62 | 0.20 | 10.06 | 0.37 |

| | | | | | |
|---|---|---|---|---|---|
| 19/69-(DA) | 0.16 | 0.49 | 0.23 | 13.47 | 0.39 |
| 19/86-(DA) | 0.17 | 0.90 | 0.18 | 17.53 | 0.35 |
| 19/119-(DA) | 0.15 | 0.76 | 0.19 | 11.58 | 0.34 |
| 19/132-(DA) | 0.10 | 0.52 | 0.29 | 8.85 | 0.39 |
| 44/86-(DA) | 0.17 | 0.84 | 0.17 | 21.90 | 0.35 |
| 44/119-(DA) | 0.16 | 0.93 | 0.17 | 15.29 | 0.33 |
| 44/132-(DA) | 0.10 | 0.48 | 0.29 | 10.37 | 0.38 |
| 44/150-(DA) | 0.17 | 0.63 | 0.21 | 9.78 | 0.38 |
| 55/69-(DA) | 0.17 | 0.58 | 0.22 | 12.69 | 0.39 |
| 55/119-(DA) | 0.21 | 0.24 | 0.18 | 9.91 | 0.39 |
| 55/132-(DA) | 0.11 | 0.45 | 0.24 | 10.38 | 0.35 |
| 55/150-(DA) | 0.16 | 0.71 | 0.23 | 25.05 | 0.39 |
| 60/86-(DA) | 0.17 | 0.66 | 0.22 | 11.79 | 0.39 |
| 60/119-(DA) | 0.10 | 0.48 | 0.22 | 9.94 | 0.38 |
| 60/132-(DA) | 0.10 | 0.47 | 0.28 | 9.86 | 0.37 |
| 60/150-(DA) | 0.17 | 0.68 | 0.28 | 11.49 | 0.39 |
| 69/86-(DA) | 0.17 | 0.61 | 0.22 | 11.49 | 0.39 |
| 69/119-(DA) | 0.15 | 0.52 | 0.24 | 10.80 | 0.39 |
| 69/132-(DA) | 0.08 | 0.49 | 0.31 | 11.03 | 0.38 |
| 69/150-(DA) | 0.10 | 0.59 | 0.27 | 8.91 | 0.37 |
| 70/119-(DA) | 0.08 | 0.62 | 0.30 | 10.41 | 0.38 |
| 70/132-(DA) | 0.08 | 0.53 | 0.29 | 9.63 | 0.38 |

Table 4.S5C **Analysis of time-resolved fluorescence anisotropies *r*(*t*) for FRET-sensitized emission of acceptor of double labeled samples obtained by ensemble time-resolved fluorescence decays as described in Section 1.6 except for [1] and [2].**
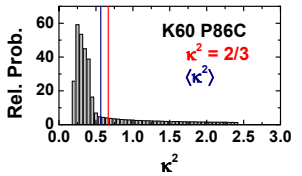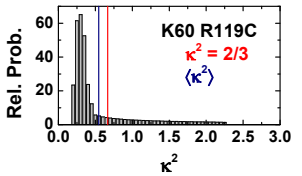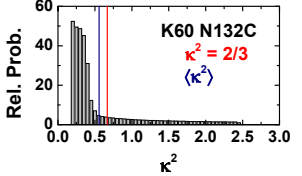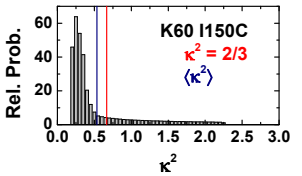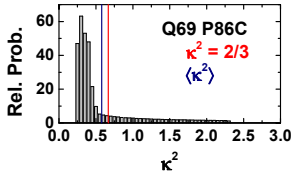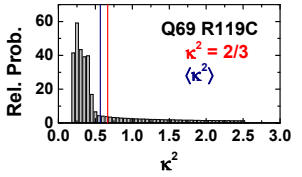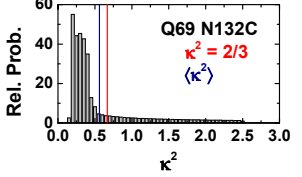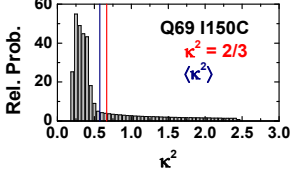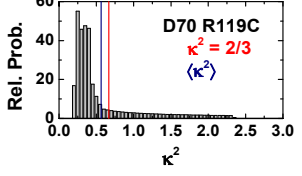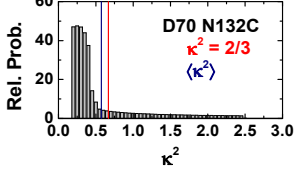
| Samples (showing aa #'s) | $r_{1, A(D)}$ | $\rho_{1,A(D)}$ | $r_{2, A(D)}$ | $\rho_{2,A(D)}$ | $\Sigma r$ |
|---|---|---|---|---|---|
| 8/69-(DA) | 0.025 | 0.509 | 0.041 | $\infty$ | 0.066 |
| 8/86-(DA) | 0.032 | 0.438 | 0.049 | 380 | 0.081 |
| 19/69-(DA) | 0.081 | 0.398 | 0.105 | 48.063 | 0.186 |
| 19/86-(DA)[1] | 0.209 | 0.756 | 0.0561 | 19.901 | 0.2651 |
| 19/119-(DA) | 0.041 | 0.512 | 0.091 | 202 | 0.132 |
| 19/132-(DA) | 0.1 | 0.373 | 0.112 | 88.561 | 0.212 |
| 44/86-(DA)[2] | <0.06 | | | | |
| 44/119-(DA)[2] | <0.09 | | | | |
| 44/132-(DA) | 0.054 | 0.246 | 0.115 | 23.934 | 0.169 |
| 44/150-(DA) | 0.087 | 0.563 | 0.048 | 101.937 | 0.135 |
| 55/69-(DA) | 0.036 | 0.405 | 0.069 | 63.43 | 0.105 |
| 55/119-(DA) | 0.067 | 1.31 | 0.089 | 136.651 | 0.156 |
| 55/132-(DA) | 0.064 | 1.039 | 0.016 | 14.346 | 0.08 |
| 55/150-(DA) | 0.065 | 0.512 | 0.061 | 150.739 | 0.126 |
| 60/86-(DA) | 0.103 | 0.483 | 0.104 | 127.327 | 0.207 |
| 60/119-(DA) | 0.079 | 0.501 | 0.086 | 114.851 | 0.165 |
| 60/132-(DA) | 0.054 | 1.035 | 0.058 | 74.739 | 0.112 |
| 60/150-(DA) | 0.038 | 1.102 | 0.067 | 77.378 | 0.105 |
| 69/86-(DA) | 0.038 | 0.604 | 0.073 | $\infty$ | 0.111 |
| 69/119-(DA) | 0.045 | 0.603 | 0.059 | 84.864 | 0.104 |
| 69/132-(DA) | 0.039 | 0.294 | 0.049 | 72.456 | 0.088 |
| 69/150-(DA) | 0.049 | 0.595 | 0.048 | 210.295 | 0.097 |
| 70/119-(DA)[2] | <0.04 | 0.2416 | | | |
| 70/132-(DA)[2] | <0.04 | 0.2471 | | | |

[1] eTCSPC data not available. Fluorescence anisotropy decay was fitted from subensemble single-molecule MFD data of the FRET population.

[2] eTCSPC data not available. Considering variants with a very high FRET efficiency, no satisfactory anisotropy decays from subensemble single-molecule MFD data were obtainable due the short donor fluorescence lifetime. Here, steady-values anisotropies were taken as upper limit from single-molecule MFD measurements.

Table 4.S5D **κ² distributions for the 24 DA samples.** Donor positions are labeled on green and acceptor positions on red. The mean $\kappa^2$ ($\langle\kappa^2\rangle$)) is shown as a solid bar in blue, and $\kappa^2$ = 2/3 is shown in red. Therefore, the assumption of $\kappa^2$ = 2/3 is justified. Nevertheless, the $\kappa^2$ distribution adds to the uncertainty on our distances, which is considered as previously described in Eq. S30.

| Acceptor / Donor | Q69C | P86C | R119C | N132C | I150C |
|---|---|---|---|---|---|
| **R8pAcF** | R8 Q69C | R8 P86C | | | |
| **K19pAcF** | K19 Q69C | K19 P86C | K19 R119C | K19 N132C | |
| **S44pAcF** | | S44 P86C | S44 R119C | S44 N132C | S44 I150C |
| **N55pAcF** | N55 Q69C | | N55 R119C | N55 N132C | N55 I150C |

| | | | | | |
|---|---|---|---|---|---|
| **K60pAcF** | | K60 P86C $\kappa^2 = 2/3$ $\langle\kappa^2\rangle$ | K60 R119C $\kappa^2 = 2/3$ $\langle\kappa^2\rangle$ | K60 N132C $\kappa^2 = 2/3$ $\langle\kappa^2\rangle$ | K60 I150C $\kappa^2 = 2/3$ $\langle\kappa^2\rangle$ |
| **Q69pAcF** | | Q69 P86C $\kappa^2 = 2/3$ $\langle\kappa^2\rangle$ | Q69 R119C $\kappa^2 = 2/3$ $\langle\kappa^2\rangle$ | Q69 N132C $\kappa^2 = 2/3$ $\langle\kappa^2\rangle$ | Q69 I150C $\kappa^2 = 2/3$ $\langle\kappa^2\rangle$ |
| **D70pAcF** | | | D70 R119C $\kappa^2 = 2/3$ $\langle\kappa^2\rangle$ | D70 N132C $\kappa^2 = 2/3$ $\langle\kappa^2\rangle$ | |
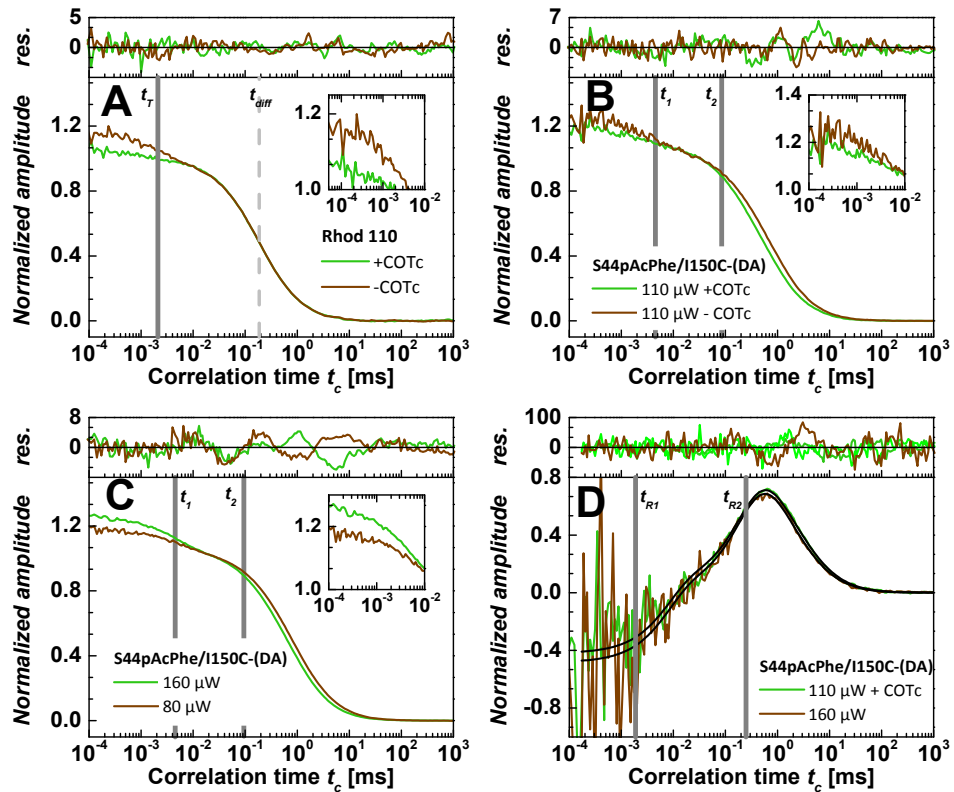
Figure 4.S8 **Triplet or dark states do not influence the sCCF on the variant S44pAcF/I150C-(DA).** (**A**) The addition of the triplet quencher COTc into Rhod110 solution significantly reduces triplet fraction (see in inset). (**B**) Overlay of the standard auto/cross-correlation curves from signals in the green channels for the variant S44pAcF/I150C-(DA) without (-COTc) and with (+COTc) triplet quencher COTc in solution. Inset shows the regime where triple kinetics is observed. (**C**) Overlay of standard auto/cross-correlation of the green signals at 80 µW and at 160 µW power at objective. Two bunching terms are needed to fit the data ($t_T$ = 4.5 µs, and $t_b$ = 60 µs). The triplet fraction changes from 10% at 80 µW to 15% at 160 µW power at objective. Also changes in diffusion times are observed from 0.8 ms at 80 µW to 0.6 ms at 160 µW power at objective. Photobleaching can account for this change. Inset shows the reduction of the triplet fraction by COTc quencher. (**D**) sCCF of the variant S44pAcF/I150C-(DA) between pseudo-species $C_1/C_2$ and $C_3$ at different power at 80 µW and at 160 µW power at the objective. The relaxation times fitted globally are $t_{R1}$ = 6 µs and $t_{R2}$ = 240 µs, that are within the errors presented on Table S4. Note that the amplitudes do not change as in the case of the standard auto-correlation.

6) The existence and the population fraction of the new conformational state $C_3$ is consistent across our library of mutants (Table 4.S1) $x_3$ = 27 ± 10 %. The statistical uncertainty of this species fraction of all FRET wt* variants determined from fluorescence decay analysis is about 2.2 %. We conclude that the variability of population fractions arises from mutagenesis and not on the ability of FRET to distinguish such fractions.

7) FPS provides also a consistent view of the conformational states of T4L. Each distinct set of conformer specific FRET restrains are within the expected uncertainty of our tools. In

117

addition, the kinetics found in all our variants are consistent with two global relaxation time ($t_{R1}$ = ~4 µs, $t_{R2}$ = ~ 230 µs) and the expected three conformational states.

8) Oversampling of number of FRET restraints $n_{restraints}$ ($n_{restraints} \geq 3(N_{NM}-1)$), where $N_{NM}$ are the normal modes, assures that there is enough structural information and reduces uncertainty introduced by each mutation. Based on normal mode analysis 5 modes are needed ($N_{NM} = 5$) (Data not shown). In total we have 24 restraints for each conformational state compared to the minimal 12 restraints.

9) Thermodynamic stability and proper folding of our mutants were verified by chemical denaturation using urea.

10) Fluorescence intensity decay where fit with various models and gave a consistent view of three FRET induced donor lifetimes or two FRET induced donor lifetimes where only one would be expected if the conformer $C_3$ did not exist.

# 5. Initiation factor 3 manuscript

This chapter corresponds to an almost finished manuscript "Intrinsic flexibility of the translation factor IF3 as a sensor of the conformational landscape of the 30S ribosomal subunit", which is very close to submission.

The manuscript is based on collaborative work in a joint project of following persons: Dmytro Rodnin, Pohl Milón, Suren Felekyan, Ingo Wohlgemuth, Marina V. Rodnina and Claus A.M. Seidel.

The contributions to this manuscript are as follows: P.M. and D.R. designed the studies, prepared the samples and did the measurement on the confocal setup; S.F. and D.R. performed analysis and simulations; S.F. developed the software, I.W. prepared the samples for the mass spectroscopy and performed the corresponding measurement; D.R., P.M., M.V.R. and C.A.M.S. wrote the manuscript; C.A.M.S. and M.V.R. supervised the project

DR: 50%

PM: 20%

SF: 5%

IW: 5%

MVR: 10%

CAMS: 10%

5.1 Initiation factor 3 manuscript main text

# Intrinsic flexibility of the translation factor IF3 as a sensor of the conformational landscape of the 30S ribosomal subunit

Dmytro Rodnin[1], Pohl Milón[2,3], Suren Felekyan[1], Marina V. Rodnina[2], Claus A.M. Seidel[1]

[1]Chair for Molecular Physical Chemistry, Heinrich-Heine University, 40225 Düsseldorf, Germany.

[2] Department of Physical Biochemistry, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany

[3] School of Medicine, Universidad Peruana de Ciencias Aplicadas - UPC, Lima, Perú

Correspondence should be addressed to M. V. R. (rodnina@mpibpc.mpg.de) or to C.A.M.S (cseidel@hhu.de)

Initiation factor 3 (IF3) is a protein with flexibly linked N- and C-terminal domains that binds to the 30S ribosomal subunit and increases the fidelity of translation initiation. Here we study structural dynamics of IF3 free in solution and on the ribosome at different stages of translation initiation over four orders of magnitude in time using a toolkit for high-precision single-molecule FRET analysis. We identify eight distinct interconverting states of free IF3 that differ in domain assembly and linker structure and three static or dynamic IF3 states in complexes with the 30S subunit depending on the nature of the mRNA and start codon recognition. Our data provide insights into the richness of the conformational landscape of a translation factor and suggest how this flexibility can be utilized to control the progression of initiation complexes towards translation.

During initiation of protein synthesis, the ribosome binds the mRNA and selects the reading frame for translation by recruiting the initiator tRNA to the start codon. In bacteria, initiation requires three initiation factors (IFs), IF1, IF2, and IF3. Translation initiation proceeds through a number of intermediate steps (Fig. 5.1a). The mRNA, IFs, and initiator fMet-tRNA$^{fMet}$ are first recruited to the small ribosomal subunit (30S) to form a labile 30S pre-initiation complex (30S PIC). The recognition of the start codon by the anticodon of fMet-tRNA$^{fMet}$ converts the 30S PIC to a mature 30S initiation complex (30S IC), which provides a docking interface for the large ribosomal subunit (50S). Finally, 50S subunit joining and dissociation of IFs result in formation of the 70S initiation complex (70S IC) which is ready to enter the elongation phase of protein synthesis. The interplay between the 30S subunit, mRNA, fMet-tRNA$^{fMet}$, and IFs controls the efficiency of mRNA recruitment for translation, thereby determining the composition of the cellular proteome and controlling the response and adaptation to environmental stimuli (Allen & Frank, 2007; Gualerzi et al, 2001; Laursen et al, 2005; Milon & Rodnina, 2012; Myasnikov et al, 2009).

IF3 is a crucial factor that controls the fidelity of fMet-tRNA$^{fMet}$ and start codon selection (Fabbretti et al, 2007; Petrelli et al, 2001). The factor takes part in screening the translation initiation region of mRNA (Antoun et al, 2006; Milon et al, 2008), promotes the movement of the 30S-bound mRNA from the 'stand-by' to the 'P-decoding' site (La Teana et al, 1993; Yusupova et al, 2006), and defines the timing and stability of the 50S joining (Grunberg-Manago et al, 1975; MacDougall & Gonzalez, 2014; Milon et al, 2008), thereby acting as a key regulator of the mRNA selection. Furthermore, IF3 accelerates the dissociation of the 30S-bound tRNA, thus contributing to the selection of the correct initiator tRNA (Antoun et al, 2006; Milon et al, 2010; Wintermeyer & Gualerzi, 1983). IF3 specifically inhibits translation of mRNAs with AUU as a start codon when IF3 is abundant, providing an autoregulation loop for its own synthesis, as the initiation codon of the mRNA coding for IF3 is AUU (La Teana et al, 1993; Sussman et al, 1996).
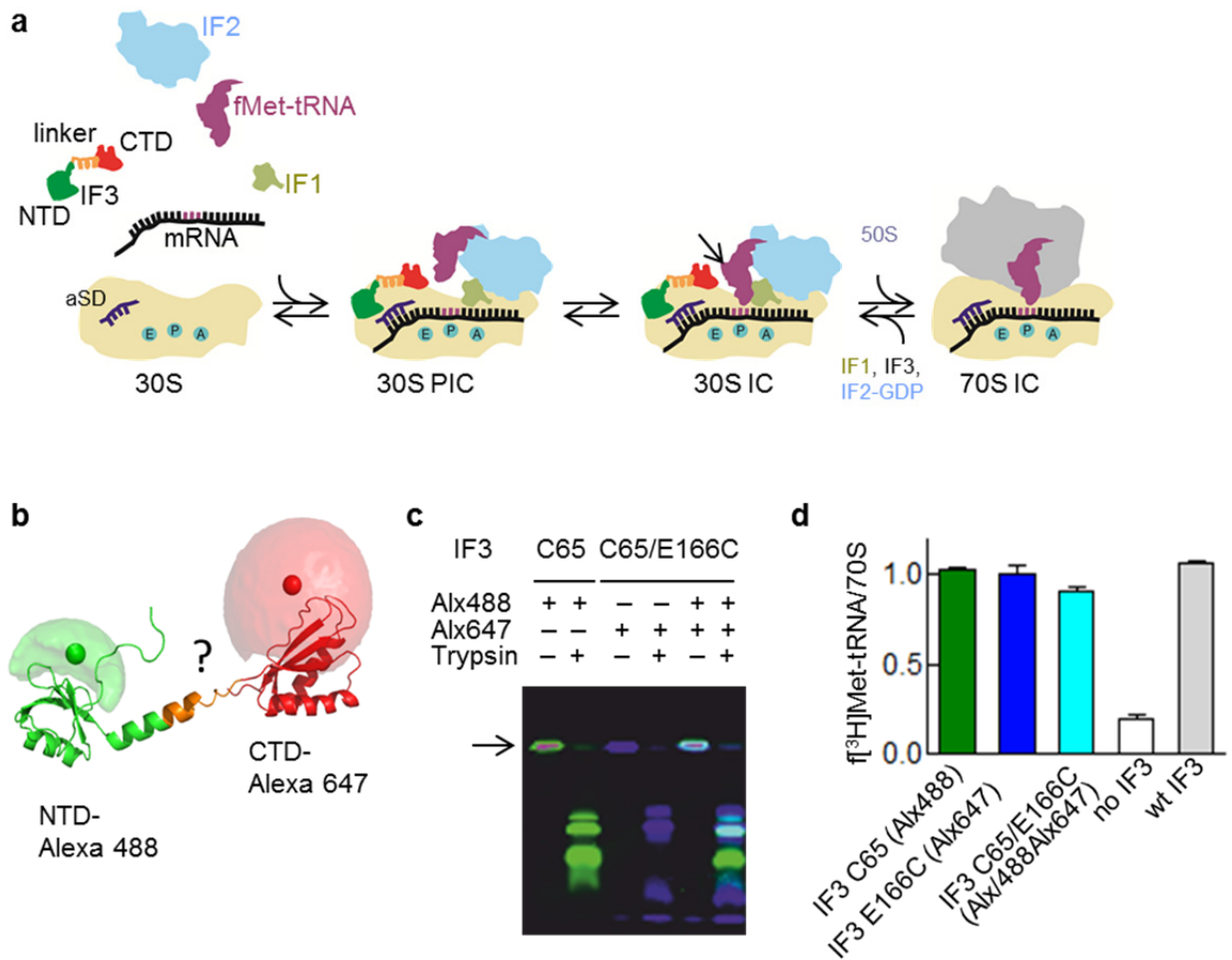
Figure 5.1 **Biochemical setup.** (**a**) Schematic of IF3 interactions during translation initiation. The NTD of IF3 (green) was labeled with Alexa 488 (donor), the CTD of IF3 (red) with Alexa 647 (acceptor). IF3 is the first IF to be recruited to the 30S subunit (yellow), followed by other IFs (IF2, blue; IF1, olive) and fMet-tRNA$^{fMet}$ (purple) to form the 30S PIC(Milon et al, 2012). Codon recognition (arrow) converts the 30S PIC into 30S IC. After the 50S subunit (gray) joining, the dissociation of IFs leads to formation of the 70S IC which is ready to enter the elongation phase of protein synthesis(Milon & Rodnina, 2012). (**b**) Model of IF3 based on the structure of individual domains(Biou et al, 1995; Moreau et al, 1997). N-terminal domain is shown in green and C-terminal domain in red. The amino acids in the linker region (orange) for which no reliable structure is available were modeled for visualization using PyMol. Accessible volumes (AV) for the dye Alexa 488 (pale green surface) and Alexa 647 (pale red surface) were simulated (Sindbert et al, 2011a), with mean dye positions depicted as green and red spheres, respectively. (**c**) Specific labeling of the NTD and CTD of IF3. Fluorescent scans of IF3 fragments separated on a 15% SDS-PAGE. Alexa 488 (green) or Alexa 647 (blue) dyes were exited at 473 nm and 635 nm, and the fluorescence was measured using a 520 or 670 nm filter, respectively, Full length IF3 (arrow) with single (IF3 wt, C65 at the NTD) or two cysteines (IF3 E166C, with C65 at the NTD and C166 at the CTD) was reacted with Alexa 647 first, followed by Alexa 488 dye (Online Methods section 1). Isolated IF3 domains were prepared by mild trypsin treatment. (**d**)

123

Biochemical activity of labeled IF3. Binding of f[$^3$H]Met-tRNA$^{fMet}$ to the 70S IC measured by nitrocellulose filter binding followed by liquid scintillation counting(Milon et al, 2010).

IF3 consists of two domains (Fig. 5.1b), the N-terminal domain (NTD), the function of which is not known and the location on the ribosome uncertain, and the C-terminal domain (CTD), which appears to carry out most of the IF3 activities in translation (Kycia et al, 1995; Moreau et al, 1997; Petrelli et al, 2001) and binds to the shoulder of the 30S subunit (Dallas & Noller, 2001; Fabbretti et al, 2007; Julian et al, 2011; McCutcheon et al, 1999). IF3 can bind to the 30S subunit through either NTD or CTD, whereby the affinity of the CTD is higher than that of the NTD (Fabbretti et al, 2007; Petrelli et al, 2001). The structures of the individual domains are known (Biou et al, 1995; Garcia et al, 1995a; Garcia et al, 1995b); however the structure of the full-length IF3 is not available. The domains are connected by a linker, which allows for different arrangements of the NTD and CTD with distance ranging from 45Å to 60Å between the centers of mass of the two domains (Moreau et al, 1997). The length of the linker may vary between just four residues (such as in IF3 from *Geobacillus stearothermophilus* (Kycia et al, 1995)) to 12 amino acids as found in *E. coli* and chloroplast IF3 (Hua & Raleigh, 1998; Moreau et al, 1997; Yu & Spremulli, 1997). Furthermore, the linker region can adopt different conformations: it can fold to an α-helix, as in the crystal structure of IF3 from *G. stearothermophilus* (Biou et al, 1995), or remain unstructured, as in the NMR structure of the *E. coli* IF3 (Moreau et al, 1997). NMR studies indicated that the linker is highly flexible and the two domains of IF3 can move largely independently of one another (Moreau et al, 1997). The functional relevance of switching between the structured and unstructured folds of the inter-domain linker, and the prevalence of different IF3 domain arrangements in solution and on the ribosome are poorly understood. In the cryo-EM reconstruction of the 30S IC, IF3 was bound in an open, extended conformation (Julian et al, 2011; McCutcheon et al, 1999). Single-molecule FRET study in a TIRF setup suggested that 30S IC-bound IF3 can assume several conformations (Elvekrog & Gonzalez, 2013) depending on the presence of initiator tRNA and proper

anticodon-codon interaction with the start codon within a completely assembled 30S IC. However, the link between the dynamics of IF3 in solution and the conformations of the factor on the 30S subunit during initiation remained unclear.

To understand the biological relevance of IF3 dynamics in solution and on the ribosome, we used single-molecule multiparameter fluorescence detection (smMFD) in combination with high-precision FRET (hpFRET) (Kalinin et al, 2012) and structural simulations. Because IF3 could switch between different domain arrangements, we first explored the conformational landscape of free IF3 using the full range of parameters provided by smMFD and subensemble time-correlated single-photon counting (seTCSPC). We obtained simultaneously several fluorescence parameters (anisotropy, lifetime, intensity, excitation and emission spectra, and quantum yield) together with the macrotime information and the calculated distance between fluorophores; structural simulations (Rothwell et al, 2003; Widengren et al, 2006) provided additional constraints for the domain arrangement. These techniques allowed us to identify a number of as yet unidentified IF3 conformations and map the network of the factor's fluctuations in solution and on the ribosome. This represents the first high-resolved functional time-space map for a translational factor.

## 5.1.1 Results

**IF3 dynamics in solution**

To probe the arrangement of IF3 domains in solution, we attached a FRET donor(D)-acceptor(A) pair through flexible linkers to two cysteines, Alexa488(D) to the native C65 in the NTD of IF3 and Alexa647(A) to an engineered E166C in the CTD (Fig. 5.1b). The position-specific labeling was feasible due to the very low accessibility of C65 in IF3 to modification at non-denaturing conditions, whereas modification of C166 with Alexa647 was very rapid (Online methods, section 5.2.1). To obtain IF3 labeled with two different dyes, Cys166 was labeled first, the unreacted dye removed, and then Cys65 labeled under denaturing conditions. This resulted in IF3 preparations which had both fluorophores at defined position (Fig. 5.1c) and were active in biochemical assays (Fig. 5.1d).

For the hpFRET experiments on free IF3, the protein was diluted to a final concentration of ~100 pM and the fluorescence bursts of freely diffusing molecules were collected for >1 h. We first considered two FRET indicators, the fluorescence weighted mean donor lifetime in the presence of the acceptor ($\langle\tau_{D(A)}\rangle_f$) and the intensity-based FRET efficiency E. The two FRET indicators for each burst were plotted on a two-dimensional frequency histogram (Fig. 5.2a), where the number of bursts in each bin is color-coded from white (lowest) to black (highest). We find an extremely broad population with the majority of E at 0.70 and a small population (20%) with E ~ 0 reflecting IF3 species without FRET (D-only). The anisotropy of the donor fluorophore was relatively low, with a mean rotational correlation time $\rho = 3$ ns (indicated by the blue line based on the Perrin equation). This allowed us to safely assume that the orientation factor $\kappa^2$ is sufficiently close to 2/3 (Dale et al, 1979) for the calculation of FRET-based distances.
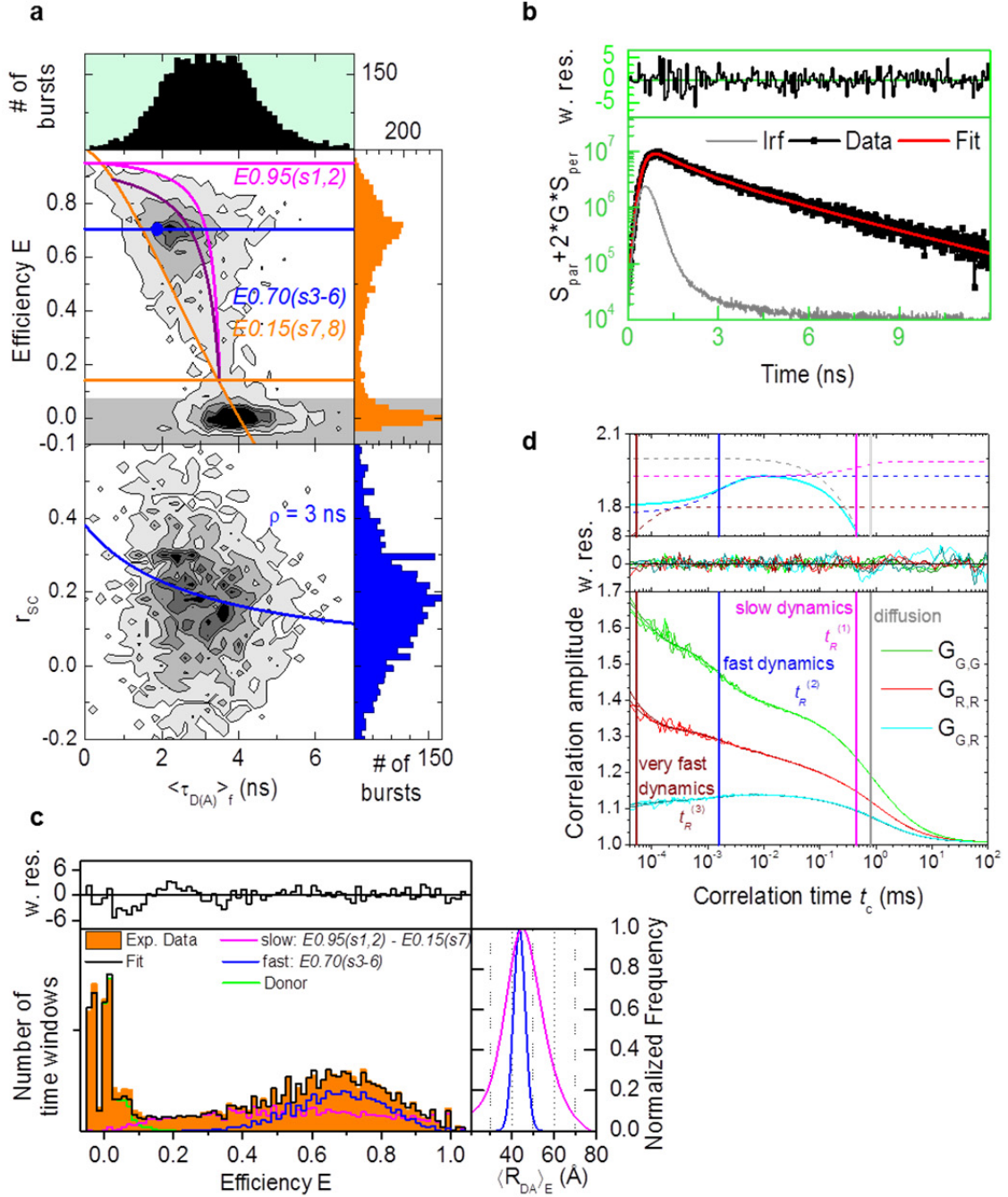
| Parameter \ Species | I<br>E0.95(s1,2) | II<br>E0.70(s3-6) | III<br>E0.15(s7,8) | Type of analysis | Figure |
|---|---|---|---|---|---|
| **Mean species lifetime, ns** | 0.2 | 1 | 3.3 | seTCSPC | 2b |
| **Fluorescence weighted average lifetime, ns [a]** | 0.4 | 1.3 | 3.5 | seTCSPC | 2b |
| **Fluorescence weighted average lifetime, ns** | 0.4; 0.7 | 2.3 | 3.5 | MFD | 2a |
| **Efficiency E [b]** | 0.95 | 0.76 | 0.19 | seTCSPC | 2b |
| **Efficiency E [c]** | 0.96 | 0.7 | 0.15 | PDA | 2c |
| **⟨R$_{DA}$⟩, Å [d]** | 32 | 43 | 66 | seTCSPC | 2b |
| **⟨R$_{DA}$⟩$_E$, Å** | 30 | 45 | 69 | PDA | 2c |

Figure 5.2. **Dynamics of IF3 in solution**. (**a**) Two-dimensional smFRET analysis of the FRET efficiency (E) and anisotropy ($r_{SC}$) vs. fluorescence lifetime ($\langle\tau_{D(A)}\rangle_f$). Theoretical static FRET line(Kalinin et al, 2010a)(orange) and dynamic FRET lines (Sisamakis et al, 2010) (purple and magenta) are shown as visual guides (Online Methods, section 5.2.3). One-dimensional projection for $\langle\tau_{D(A)}\rangle_f$ is shown on top. One dimensional projection of E and $r_{SC}$ are shown on the right. Horizontal reference lines show the E values corresponding to *E0.95(s1,2)* (magenta, from PDA), *E0.70(s3-6)* (blue, from PDA), and *E0.15(s7,8)* (orange, from PDA) as visual help. (**b**) Fluorescence lifetime analysis of the main population from the 2D smFRET plot. Bursts were selected and all donor photons were accumulated in a decay histogram for seTCSPC analysis using a fit with three lifetimes (additional analysis is shown Suppl. Fig. 5.S1c). Experimental data of total signal S is shown in black, the instrument response function in grey, and the fit in red. The fit yielded $\tau_1 = 0.18$ ns ($x_1 = 0.44$), $\tau_2 = 1.02$ ns ($x_2 = 0.19$), $\tau_3 = 3.27$ ns ($x_3 = 0.37$), G-Factor = 1.046. (**c**) PDA analysis of the E values in a time window of 1 ms. *E* values were obtained by correcting as described in Online Methods, section 5.2.4. The measured green and red signal for background were $\langle B_G \rangle = 0.633$ kHz and $\langle B_R \rangle = 0.358$ kHz. The black line represents the fit containing two dynamic and one donor-only population. Magenta line represents the dynamic population with the end distances of 30 Å and 69 Å and $k_{12} = 2.08$ ms$^{-1}$ and $k_{21} = 1.76$ ms$^{-1}$; blue line represents the dynamic population with fast and very fast interconversion rates resulting in a quasi-static population Gaussian distance distribution with $\langle R_{D(A)} \rangle_E = 45$ Å and w = 8 Å. The panel on the right side indicates the distributions of the relevant FRET populations – *E0.95(s1,2)* to *E0.15(s7,8)* in magenta and *E0.70(s3-6)* in blue. (**d**) FCS analysis of the free IF3. Six correlations were calculated, two green auto-correlations $G_{G,G}$ (green lines), two red autocorrelations $G_{R,R}$ (red lines) and two cross-correlations $G_{G,R}$ green to red and red to green (cyan lines). The curves were globally fitted with a model containing six bunching terms for the auto-correlation functions and two bunching and four anti-bunching terms for the cross-correlation functions, respective fits are shown in a darker shade of the respective color. The fits yield: $t^{(1)}_R = 470$ µs (magenta vertical line); $t^{(2)}_R = 1.6$ µs (blue vertical line); and $t^{(3)}_R = 60$ ns (dark red vertical line); diffusion time is 800 µs (gray vertical line). Respective residuals are shown on top. The top panel indicates the contribution of different dynamic terms to the overall shape of the cross-correlation curve – the very fast dynamic term in dark red, the fast dynamic term in blue, the slow dynamic term in magenta, and the diffusion term in gray. The resulting cross-correlation term is shown in cyan. (**e**) Table of the different species measured for free IF3. The values are taken from the lifetime analysis and PDA and converted as described in the Online Methods.

[a] Converted from mean species lifetimes, see Online Methods, section 5.2.3.

[b] Calculated from mean species lifetime using Eq. 5.2

[c] Calculated from $\langle R_{DA} \rangle_E$ values using Eq. 5.6

[d] Calculated from mean species lifetime using Eq. 5.7

We next compared the measured E vs. $\langle\tau_{D(A)}\rangle_f$ distribution to a line reflecting all static FRET states (Fig. 5.2a, orange line)(Online Methods, section 5.2.3, eq. 5.3)(Sisamakis et al,

2010). If IF3 conformations corresponding to different FRET values were static during the duration of the burst, they should be distributed along this line. This is clearly not the case; rather, FRET distribution is shifted away from the static FRET line, which is a hallmark of a dynamic behavior.

**Analysis of dynamics.** Fluorescence analysis of the population dynamics of IF3 was carried out in four steps (Felekyan et al, 2013). To resolve short-lived states, we first performed sub-ensemble time correlated single-photon counting (seTCSPC) by selecting all bursts of the FRET population from Fig. 5.2a (Suppl. Fig. 5.S1a,b, green box for selection). The analysis of the combined decay histogram of the donor fluorescence (Kalinin et al, 2010b) showed that the FRET population was heterogeneous, because a single lifetime was clearly not sufficient to describe the experimental data (Fig. 5.2b). The simplest model which yielded satisfactory fits required three FRET states with mean species lifetimes (species fractions) $\tau_I = 0.2$ ns (44%), $\tau_{II} = 1.0$ ns (19%) and $\tau_{III} = 3.3$ ns (37%) representing distinct groups of kinetic species. For the comparison with the smMFD plots (see below), we converted these lifetimes to fluorescence weighted average lifetimes $\langle\tau_{D(A)}\rangle_f$ (Online Methods, section 5.2.3) (Fig. 5.2e).

The broadening of the FRET populations in the MFD diagram (Sisamakis et al, 2010) entails a unique footprint which allowed us to gain the first insight in the underlying protein dynamics in the second step of analysis (Fig. 5.2a). We used the lifetime information (Fig. 5.2b,e) to explain the extreme broadening of both FRET indicators in the MFD diagram in terms of exchange dynamics between several FRET populations. Using the MFD fluorescence weighted average lifetimes, we modeled two dynamic FRET lines (Sisamakis et al, 2010) (Online Methods section 5.2.3, eq. 5.4) to describe the averaging between high-FRET states with $\langle\tau_{D(A)}\rangle_{f-I} = 0.4$ ns and $\langle\tau_{D(A)}\rangle_{f-I} = 0.7$ ns, respectively, and a low-FRET state $\langle\tau_{D(A)}\rangle_{f-III} = 3.5$ ns (Fig. 5.2a, magenta and purple line, respectively). While these fluctuations described the bursts at the periphery of the distribution, an additional FRET population with E $\approx$ 0.70 was found between the static and

dynamic FRET lines (Fig. 5.2a), indicative for a fast exchange dynamics (relaxation time < 20 μs) between multiple FRET species (Gopich & Szabo, 2012b). This agrees with the presence of the third fluorescence lifetime $\langle \tau_{D(A)} \rangle_{f-II} \approx 2.3$ ns (Fig. 5.2e). Modeling the position of this averaged population in the MFD diagram (Fig. 5.2a, blue dot) (Online Methods, section 5.2.3) required a Gaussian FRET efficiency distribution with a very broad width w= 21.5 Å (see below).

In the third step, we analyzed the FRET efficiency histogram by photon distribution analysis (PDA) (Antonik et al, 2006) to estimate the time-scales of interconversion between the states (Kalinin et al, 2010b) (Fig. 5.2c). PDA in different sampling time windows showed that the distribution was broader than expected from the shot noise and further broadened with decreasing time window, which provides an additional evidence for the dynamic behavior. The best fit of the data was obtained using a model with two species: (i) a sub-population *I-slow* (59%) and ((ii) a sub-population *II-fast* (41%). The sub-population *I-slow* had a broad FRET efficiency distribution, which described slow exchange between the high-FRET states (E ≈ 0.96) and the low-FRET state (E = 0.15). The relaxation time was $t_R^{(1)} \sim 250$ μs was calculated from nearly equal rate constants for the forward and backward reaction $k_{fw} \approx k_{bw} \sim 2$ ms$^{-1}$; the actual value for $t_R^{(1)}$ was actually ~ 400 μs, because diffusion effects must be taken account (eq. 11 in (Kalinin et al, 2010b)). The second sub-population *II-fast* (41%) had a narrow effective FRET efficiency distribution indicative of rapid exchange with an upper limit for the relaxation time $t_R^{(2)} < 20$ μs (Fig. 5.2c). Overall, three FRET species with $E_I \approx 0.95$, $E_{II} = 0.70$ and $E_{III} \approx 0.15$ were consistently found by seTCSPC, PDA and smMFD (Fig. 5.2e). The mean FRET averaged interdye distances $\langle R_{DA} \rangle_E$ determined by PDA and smMFD, agree well with the mean interdye distance $\langle R_{DA} \rangle$ from donor lifetime analysis (Fig. 5.2e).

In the final step of analysis, we performed fluorescence correlation spectroscopy (FCS) to further increase the time resolution (Fig. 5.2d). We calculated the D and A autocorrelation ($G_{G,G}$,

$G_{R,R}$) and the DA cross-correlation curve ($G_{G,R}$) using the donor and acceptor signal in the green and red spectral detection window, respectively. A global fit of the three curves to eq. 5.9 (Online methods section 5.2.5) yields three relaxation times $t_R^{(1)} = 470$ µs, $t_R^{(2)} = 1.6$ µs, and $t_R^{(3)} = 60$ ns for the FRET-related anti-correlation terms, which confirms the presence of multiple exchange processes.

**Analysis of structure.** To relate FRET distances to the conformational states of IF3, we performed Markov Chain Monte-Carlo sampling and "accessible volume" (AV) simulations of the dyes to compute the effective interdye distances (Sindbert et al, 2011a) (Online method, section 5.2.6). The linker region of IF3 (amino acids 78-90) can adopt different conformations, e.g. α-helical up to the residue 86 (PDB: 1TIF) (Biou et al, 1995), or disordered (Moreau et al, 1997). The linker contains six lysines which convey a high mean net charge and low mean hydrophobicity which favor intrinsically disordered states (Muller-Spath et al, 2010; Uversky et al, 2000) (Suppl. Fig. 5.S2a,b). To consider all potential IF3 domain arrangements, we simulated two scenarios: (i) the structured linker is restricted to a helical conformation; or (ii) the unstructured flexible linker behaves as "random chain" (Suppl. Fig. 5.S2c). In the simulation with a structured linker, the mean dye distance $\langle R_{D(A)} \rangle$ varied from 48 to 75 Å (Fig. 5.3a, orange), similar to the earlier structural models for IF3 dynamics (Moreau et al, 1997). With an unstructured linker the sampled structural space was much broader, with $\langle R_{D(A)} \rangle$ ranging from 27 to 73 Å (Fig. 5.3a, blue). Thus, in the IF3 molecules with $\langle R_{D(A)} \rangle < 48$ Å the linker region is likely to be unstructured, whereas those with larger distances may be either in a structured or unstructured conformation. When the linker in unstructured, the NTD and CTD may adopt two alternative types of arrangements, (i) an extended arrangement in which the domains are spatially separated or (ii) a compact arrangement with domains contacting each other in some way; the latter we denoted as "domains associated" (Fig. 5.3a). The analysis of obtained structural models with "domains associated" showed that the distance between the respective centers of mass (amino acids 25 and 126 in the NTD and CTD, respectively) was smaller than

131
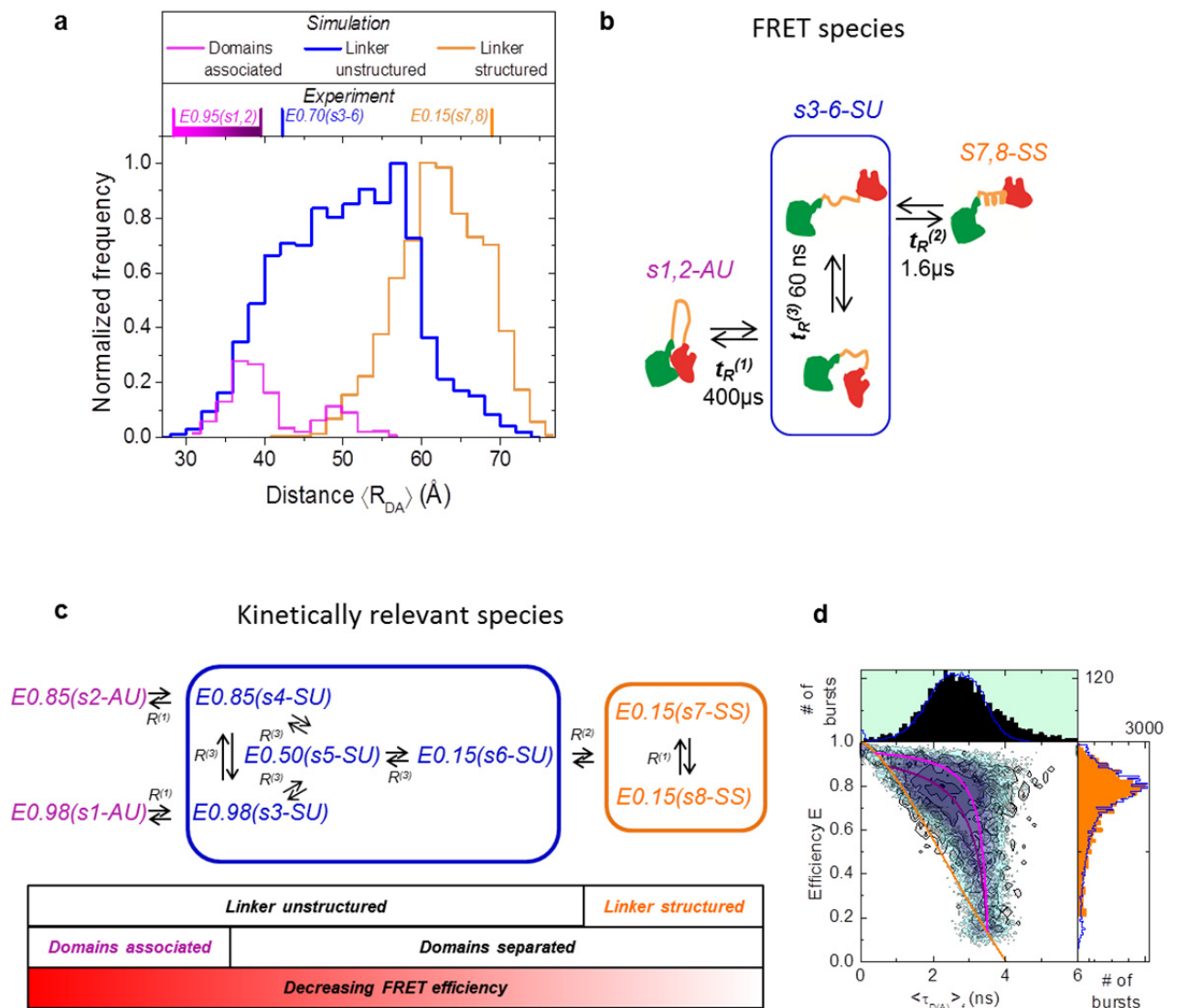
Figure 5.3 **Analysis of conformational states of free IF3.** (**a**) Analysis of the dye distance distribution by Markov Chain Monte-Carlo simulations. The histogram shows obtained distance distribution for the structures with structured linker (<u>s</u>tructured linker, <u>s</u>eparated domains, *SS*, orange), unstructured linker (<u>s</u>eparated domains, <u>u</u>nstructured linker, *SU*, blue), or with unstructured linker for those structures where the distance between centers of mass (as defined by distance between amino acids 25 and 126) was <25 Å (<u>a</u>ssociated domains, <u>u</u>nstructured linker, *AU*, magenta). (**b**) Assignment of the relaxation rates to the transitions between FRET populations based on PDA, FCS, and structural analysis. The transition between *AU* and *SU* is the slowest at ~ 400 µs. Interconversion within the *AU* population is 60 ns, governed by chain diffusion dynamics. The conversion from *SU* to *SS* populations requires folding of the linker region with the relaxation time of 1.6 µs. (**c**) Complete kinetic scheme for dynamic fluctuation of free IF3. The minimum of eight different species was required to fit the experimental data. Two states with associated domains, *E0.98(s1-AU)* and *E0.85(s2-AU),* are required to account for the particularly broad distribution in 2D histograms (Fig. 5.3d) and supported by a broad distribution of the *AU* population (Fig. 5.3a). These two states are interconnected to two states *E0.98(s3-SU)* and

*E0.85(s4-SU)* with the same FRET efficiency, but with separated domains. In turn, these two states are connected to the state *E0.75(s5-SU)* and to *E0.15(s6-SU)*, comprising the *SU* group of populations. Upon the formation of α-helical structure of the linker, the protein assumes the conformations *E0.15(s7-SS)* and *E0.15(s8-SS)*. (**d**) Comparison of the measured and simulated 2D histograms for the free IF3. Simulation was performed with eight kinetic states and rates constants given in the Online Methods, section 5.2.7. The resulting 1D and 2D-histograms (blue) were overlaid on top of the experimental data (grey, black and orange respectively). The difference in the very short lifetime is due to limitations of the measurement setup and analysis routine to detect molecules with a very short donor fluorescence lifetime.

25 Å. For this conformation the $\langle R_{D(A)} \rangle$ histogram revealed two populations with a major peak at 38 Å (width of 10 Å), which likely has an unstructured linker (see above), and a minor peak at 50 Å (Fig. 5.3a, magenta). Therefore we refer to this species with associated domains and unstructured linker as *AU*. For the low-FRET species ($E_{III} = 0.15$) (Fig. 5.2e), the arrangement with the structured linker is more likely (Fig. 5.3a).

**Analysis of the kinetic network.** Next, we made an initial structural assignment of the states with three relaxation times obtained by FCS and PDA. While rapid interconversion between different conformations may be feasible for the molecules in an open conformation, the closed conformations are unlikely to interchange rapidly, because the domains bound to one another have to unbind first, resulting in a rate-limiting step. We assign the limiting states for slow exchange dynamics with a relaxation time $t_R^{(1)} \sim 400$ µs to transitions between the high-FRET state with associated domains and unstructured linker, *AU*, and other states with separated domains and unstructured linker (referred to as *SU*). These transitions are slow because the domain dissociation is assumed to be rate limiting (Fig. 5.3b). In comparison, characteristic times for helix folding are in the order of a few microseconds (Fierz et al, 2009), suggesting that transitions between conformers with separated domains and unstructured linker (*SU*) and separated domains and structured linker (*SS*) may reflect the 1.6 µs relaxation time $t_R^{(2)}$. Rearrangements within the pool of *SU* conformations are expected to be very rapid, because characteristic times for chain diffusion are in the order of 50-100 ns (Krieger et al, 2004; Krieger et al, 2003), which agrees very well with $t_R^{(3)} = 60$ ns resolved by FCS.

133

Using this assignment, we simulated the kinetic scheme of transitions between different IF3 structures. The minimum model required eight distinct states (Fig. 5.3c) to obtain the rate-limiting steps equivalent to the experimentally observed FRET efficiencies and relaxation times (Online methods, section 5.2.7). At least two *AU* species were needed to reproduce the experimental 2D plots (Fig. 5.3d). Both of them contribute to the high-FRET population (Fig. 5.2a,e) and are thus denoted *E0.98(s1-AU)* and *E0.85(s2-AU)* (the nomenclature describes the FRET efficiency, state number, domain and linker conformation). The two states *s7* and *s8* have a low FRET efficiency, separated domains and structured linker, with the state *E0.15(s7,8-SS)* for the $E_{III}$ population (Fig. 5.3c). For the rapidly exchanging, mixed population *SU*, a minimum of four discrete kinetically relevant species was needed with different FRET efficiencies ranging from high to low. Due to a very fast chain dynamics ($t_R^{(3)}$ = 60 ns) of the unstructured linker, this ensemble *E0.70(s3-6-SU)* reflects the narrow average intermediate FRET efficiency distribution $E_{II}$ (Fig. 5.2c). Note that the broad distribution towards the end states *E0.98(s1-AU)*, *E0.85(s2-AU),* and *E0.15(s8-SS)* necessitates the use of slow exchange equilibria to *E0.70(s3-6-SU)* with $t_R^{(1)} \approx 400$ µs via their corresponding gating states. Thus, IF3 in solution is an extremely dynamic molecule that fluctuates between multiple structural states over 4 orders of magnitude in time. The agreement between the kinetic scheme with 8 relevant species (Fig. 5.3c) and all experimental results is demonstrated in Fig. 5.3d by simulating an E vs. $\langle \tau_{D(A)} \rangle_f$ MFD plot.

## IF3–30S subunit complex

In the complex with the 30S subunits, IF3 adopted a conformation with $\langle\tau_{D(A)}\rangle_f = 3.1$ ns and $E = 0.22$ on the static FRET line; the state has a low FRET efficiency and is referred to as *E0.22(sIC)*. The relative weight of this new static population of bound IF3 increased with the 30S subunit concentration. Because of a large difference in the molecular weight between IF3 and the 30S subunit, free IF3 has a much higher diffusion coefficient than the IF3–30S subunit complex, which allowed us to measure complex formation by FCS (Fig. 5.4c). To determine the degree of binding, we used a model with two diffusion terms and a bunching term for a triplet fraction (Online Methods, section 5.2.5). We obtained the diffusion times of free IF3 $t_{d\text{-}free} = 0.6$ ms and of the IF3–30S subunit complex $t_{d\text{-}complex} = 1.9$ ms (Fig. 5.4c), as well as the dissociation constant $K_d = 4.6$ nM for the 1:1 complex (Fig. 5.4d). This value agrees with that measured by fluorescence titrations or calculated from the association and dissociation rate constants of a two-step binding reaction (Milon et al, 2012; Weiel & Hershey, 1978), $K_d = 3$ nM.
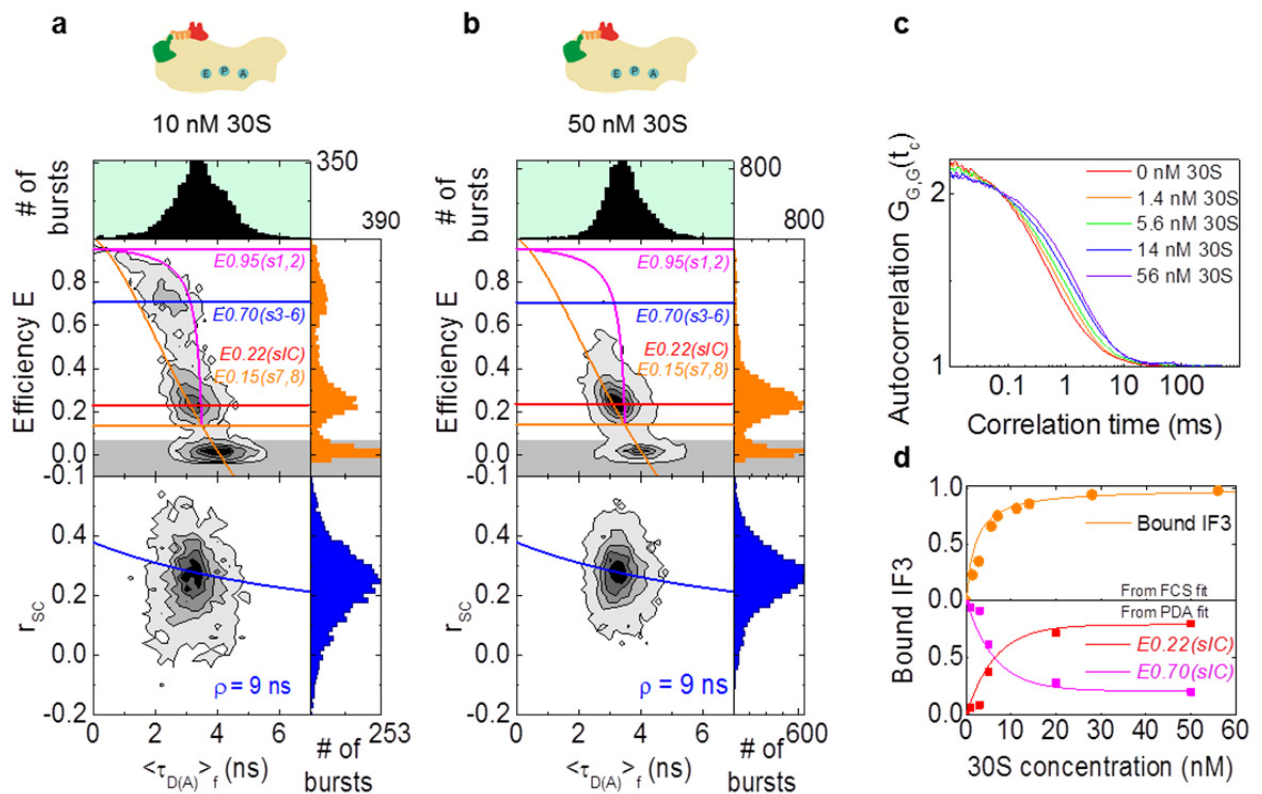
Figure 5.4 **Conformations of IF3 bound to the 30S subunit. (a,b)** Two-dimensional smFRET analysis of FRET efficiency (E) and anisotropy ($r_{SC}$) vs. fluorescence lifetime ($\langle\tau_{D(A)}\rangle_f$ ) of IF3 in the presence of 10 nM 30S subunits (**a**) and 50 nM 30S subunits (**b**). Horizontal reference lines show E values for *E0.95(s1,2)* (magenta), *E0.70(s3-6)* (blue), *E0.15(s7,8)* (orange) and *E0.22(sIC)* (red), all from PDA. The functional state of the ribosome is indicated by a cartoon at the top of the respective 2D plot. (**c**) FCS curves obtained with different 30S subunits concentrations, as indicated. (**d**) Titration of the IF3 with 30S subunits analyzed by PDA (Online Methods, section 5.2.4). Upper panel shows data analysis derived from the FCS analysis and fitted with a hyperbolic function (fraction bound IF3) = [30S] /($K_d$+[30S]), where $K_d$ = 4.6 ± 0.8 nM. Lower panel shows data derived from PDA. The size of the symbols represents the highest error bar for a single point.

Even though FCS indicated complete binding of IF3 at saturation concentration of 30S subunits (Fig. 5.4d, top panel), the FRET population in the MFD plot is broadened beyond the shot noise limit (Fig. 5.4b, right panel, orange bars), indicating residual dynamics of bound IF3. PDA analysis suggested that the major IF3 peak is represented by the static population *E0.22(sIC)*, with a Gaussian distribution of interdye distances $\langle R_{D(A)}\rangle_E$ = 64 Å with a width w = 9 Å (E = 0.22). The w values are characteristic for FRET broadening by brightness effects of the acceptor dye Alexa 647 (Kalinin et al, 2010a); however, within the resolution of FRET the major conformation of IF3 is static and homogeneous. In addition to this state, PDA revealed the presence of a second 30S-bound IF3 population with a relaxation time $t_R$ ≈ 200 μs, which exhibited dynamic behavior similar to the *E0.95(s1,2)-E0.15(s7,8)* population in free IF3 (Fig. 5.4d, bottom panel). Combined information from FCS and PDA suggested that this population (denoted as *E0.70(sIC)*) represents IF3 bound to the 30S subunit with only one of its two domains. When IF3 was displaced from the 30S subunit upon addition of unlabeled IF3, it re-gained its conformational dynamics characteristic for the free IF3, suggesting that the conversion between the bound and free IF3 states was reversible (see section with Fig. 5.6b).

## Conformation of IF3 in the 30S PIC and 30S IC

IF3 binding to the 30S subunit is stabilized by IF1 (Milon et al, 2012). In fact, at sub-saturating concentrations of 30S subunits the addition of IF1 resulted in a further shift from the dynamic to the static IF3 state (Fig 5.5a). The interdye distance in the static population of IF3 bound to the 30S subunit was the same in the absence and presence of IF1, suggesting that the domain arrangement of IF3 was not changed by IF1.
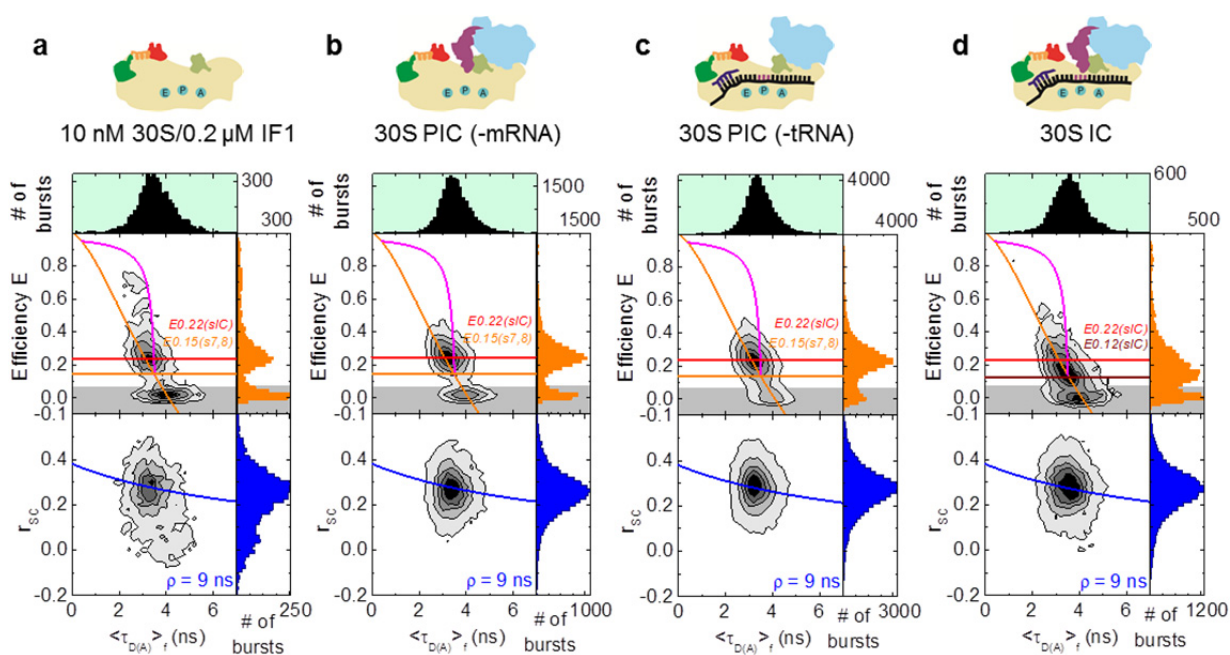


Figure 5.5 **Conformations of IF3 on 30S PIC and 30S IC**. (**a,b,c,d**) Two-dimensional smFRET analyses of FRET efficiency (E) and anisotropy ($r_{SC}$) vs. fluorescence lifetime ($\langle\tau_{D(A)}\rangle_f$) of IF3 in the presence of 30S subunits (10 nM) and IF1 (0.2 µM) (**a**), 30S subunits (50 nM), IF1 (0.2 µM), IF2 (0.2 µM) with GTP (0.25 mM), and fMet-tRNA$^{fMet}$ (0.3 µM) (**b**); with 30S subunits (50 nM), IF1 (0.2 µM), IF2 (0.2 µM) with GTP (0.25 mM), and m002 mRNA (0.3 µM) (**c**); or with 30S subunits (50 nM), IF1 (0.2 µM), IF2 (0.2 µM) with GTP (0.25 mM), fMet-tRNA$^{fMet}$ (0.3 µM), and m002 mRNA (0.3 µM) (**d**). Horizontal reference lines show E values for *E0.15(s7,8)* (orange), *E0.22(sIC)* (red), and *E0.12(sIC)* (dark red), all from PDA. The functional state of the ribosome is indicated by a cartoon at the top of the respective 2D plot.

Next, we examined whether binding to the 30S subunit of further initiation components, i.e. IF2, fMet-tRNA$^{fMet}$, or mRNAs, changed IF3 conformation. In all complexes which represent sub-states of the 30S PIC, i.e. the state before the start codon selection, IF3 was found in the same static conformation *E0.22(sIC)* with $\langle R_{D(A)}\rangle_E \sim 64$ Å (Fig. 5.5b,c), suggesting that the

structure of IF3 bound to the 30S subunit is not changed by other components of the initiation machinery. Only when all initiation components were present, including an m002-mRNA with the start codon AUG allowing for codon-anticodon recognition and formation of the 30S IC, we observed an additional FRET population of IF3 molecules with $\langle\tau_{D(A)}\rangle_f = 3.75$ ns and E = 0.12 (Fig. 5.5d), referred to as *E0.12(sIC)*. PDA showed that IF3 in these complexes was static with $\langle R_{D(A)}\rangle_E \sim 72$ Å. Thus, the two domains of IF3 moved further apart upon the formation of the 30S IC and start codon recognition. Notably, although 30S IC was expected to be biochemically homogeneous, i.e. the majority of 30S subunits should have a full complement of initiation components, only a portion (~60%) of the 30S IC was converted into the *E0.12* state whereas remaining complexes remained in the *E0.22* state.

The experiments described above were carried out using m002 mRNA with an enhanced translation initiation region. This contains a strong Shine-Dalgarno (SD) sequence with 8 nucleotides (nt) that can potentially base pair to the anti-SD (ASD) sequence in 16S rRNA and a short 4-nt spacer between the SD sequence and the start codon AUG (see schematics in Fig 5.6 and sequences in Suppl. Fig. 5.S4). To examine whether the structure of IF3 is sensitive to the elements in the translation initiation region, we performed analogous experiments using eight different mRNAs. In addition to m002 mRNA, we used model m022 mRNAs with a weaker SD-ASD complementarity and a longer spacer, m003 mRNAs which does not have an SD sequence(Calogero et al, 1988), and a natural mRNA infC, which has a 7-nt SD sequence followed by a 7-nt linker and a non-canonical AUU codon (infC AUU) (Suppl. Fig. 5.S4). For all mRNAs, PDA identified three populations of IF3: a small portion of dynamic molecules, *E0.70(sIC)* (7-20%) that displayed a broad distance distribution, and two static sub-populations, *E0.22(sIC)* and *E0.12(sIC)* (Fig. 5.6). The *E0.24(sIC)* and *E0.12(sIC)* conformations represent two discrete states of IF3 on the ribosome, open and extended, which differ in the arrangement of the NTD and CTD. The switch between the two conformations is specifically defined by the
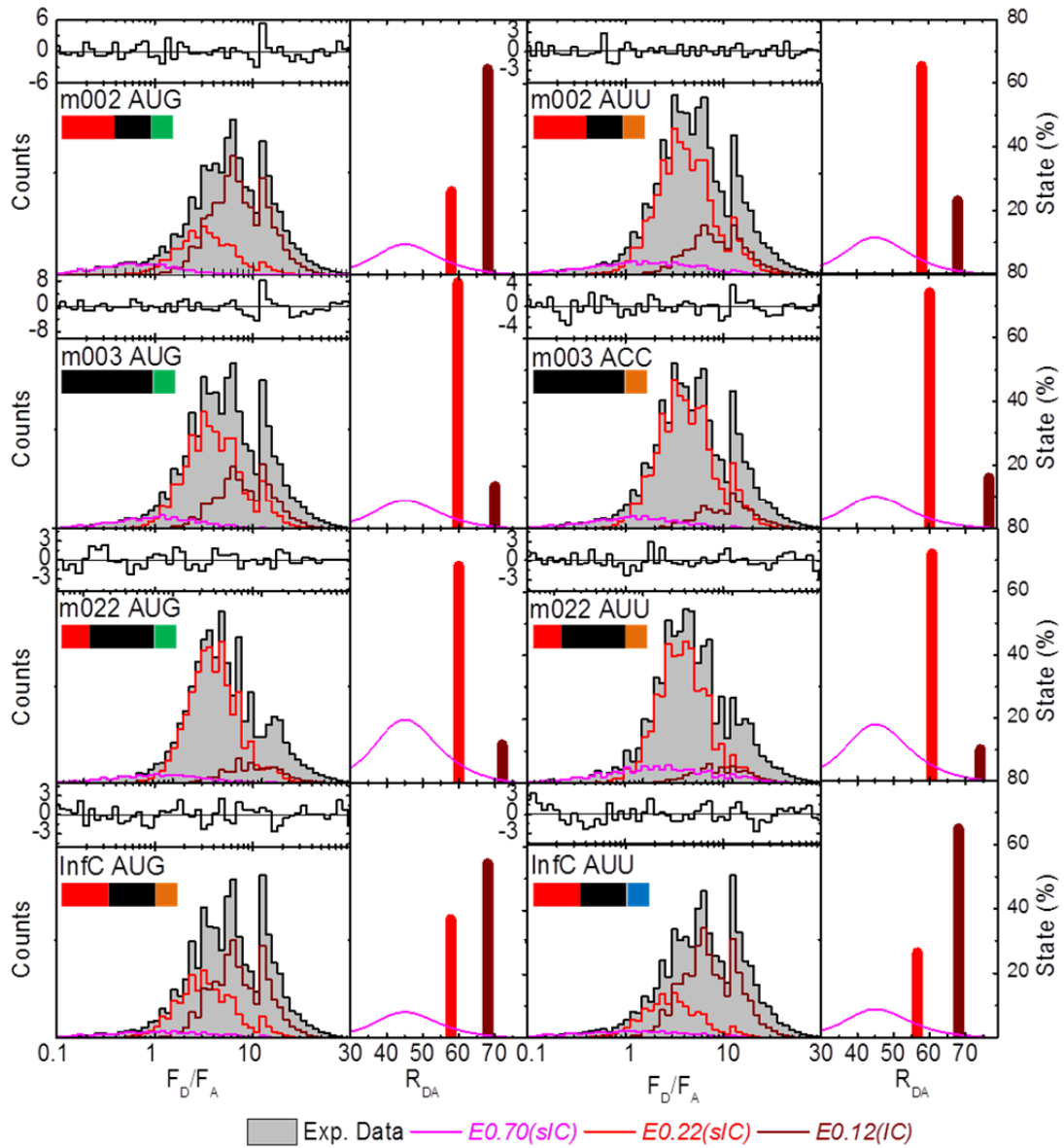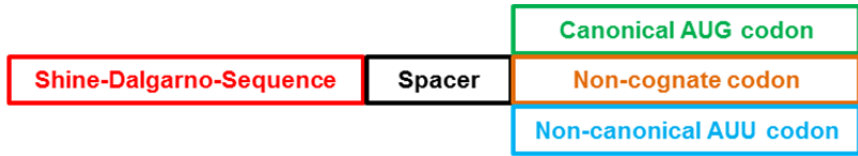
Figure 5.6 **Effect of different mRNAs on 30S IC**. The 1D histograms of the donor to acceptor fluorescence intensity ratio $F_D/F_A$ were obtained by correcting the measured green and red signal for background ($<B_G>$ = 1.022 kHz, $<B_R>$ = 0.745 kHz), using time window of 1 ms (Online Methods, section 5.2.4). Black line represents the fit containing one dynamic (purple), two static (red and dark red) and one donor-only population (not shown). Purple line represents the dynamic population with the end-distances of 25 Å and 69 Å and varying $k_{12}$ and $k_{21}$, red line represents static population with a Gaussian distribution with varying $R_{DA}$ between 60 Å and 64 Å and 8 Å width, dark red line represents static population with a Gaussian distribution with varying $R_{DA}$ between 73 Å and 77 Å and 8 Å width. Donor-only population fraction was between 10 and 20% in all samples. On the right side the described populations are shown as either Gaussian

139

distribution (purple) or discrete distances (red, dark red). Their height corresponds to the abundance of the certain state (for a Gaussian distribution only the height in the highest point is relevant).

sequence or structure of the mRNA. In addition to the two static IF3 populations, 10-20% of the molecules were found in a dynamic *E0.70(sIC)* state representing IF3 attached to the 30S subunit by only one of its two domains. For the mRNAs used, the proportion of dynamic vs. static molecules did not correlate appreciably with the properties of the ribosome binding site. Notably, the portion of the free IF3 molecules did not increase in any of the 30S ICs tested, as compared to the 30S–IF3 complex or to the various 30S PICs, which suggests that formation of the 30S PIC or 30S IC does not cause appreciable dissociation of IF3 from the 30S subunit.

**Dissociation of IF3**

The last step in the initiation process is the binding of the 50S subunit and the release of IF3 and other IFs from the 70S IC. To follow the fate of IF3 upon 70S IC formation, we added excess of 50S subunits to 30S ICs with different mRNAs (Suppl. Fig. 5.4); data collection was started about 2 min after 50S addition and continued for 1 h. The conformational distribution in IF3 population changed upon addition of the 50S subunit. The extended *E0.12(sIC)* population was no longer present, a portion of IF3 molecules (60-30%) were found in the *E0.22(sIC)* state, whereas the remaining IF3 fraction re-gained dynamic properties characteristic for the free IF3 (Fig. 5.7a). There was no obvious systematic variation depending on the strength of the SD-ASD interaction or the presence of the AUG start codon. The complete chase of IF3 from the 30S IC could be achieved by addition of excess unlabeled IF3 (Fig. 5.7b; note the similarity of the MFD plot to Fig. 5.2a), indicating that all double-labeled IF3 molecules were bound to the ribosome in a productive, potentially reversible manner. Thus, the incomplete dissociation of IF3 from the 70S IC represents conformational or compositional heterogeneity of 30S IC/PIC.
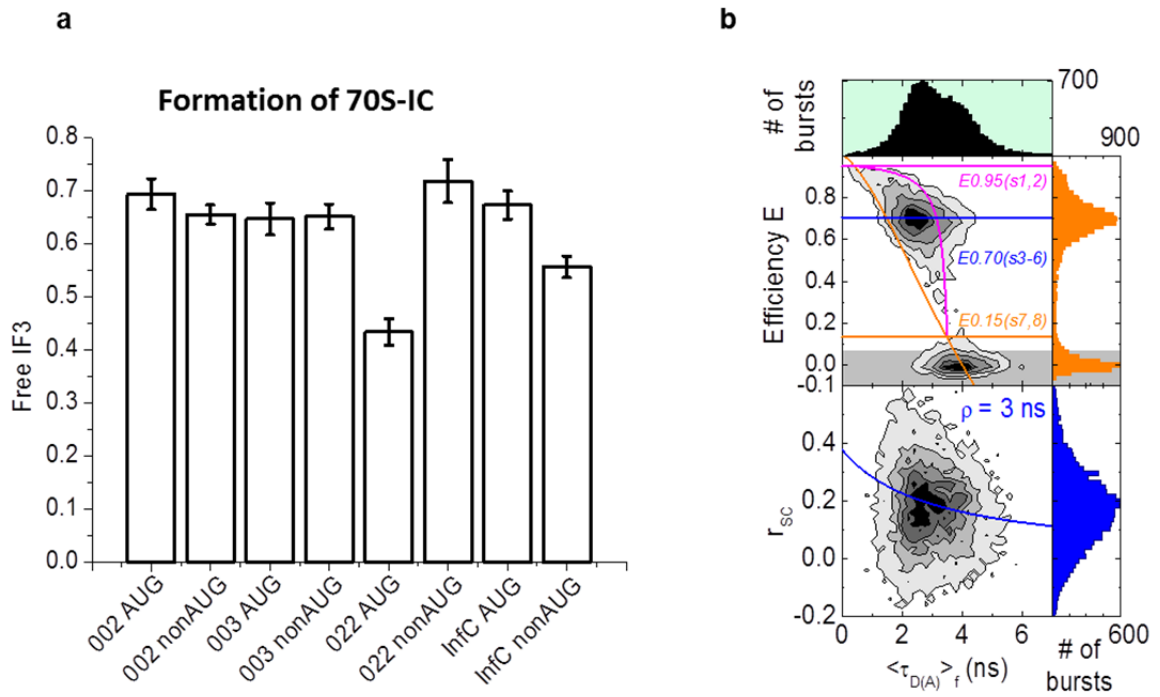
Figure 5.7 **Dissociation of IF3 upon 70S IC formation.** (**a**) mRNA-dependent dissociation of IF3 from 70S after addition of 50S subunits. IC was formed from double labeled IF3 (0.1 nM), 30S subunits (50 nM), IF1 (0.2 µM), IF2 (0.2 µM) with GTP (0.25 mM), fMet-tRNA$^{fMet}$ (0.3 µM), and respective mRNA (0.3 µM). Measurement was started by addition of 0.5 µM 50S and monitored for 60 minutes. The resulting fractions were stable after 10 minutes and analyzed by PDA (Online methods, section 5.2.4). (**b**) Two-dimensional smFRET analysis of FRET efficiency (E) and anisotropy ($r_{SC}$) vs. fluorescence lifetime ($\langle\tau_{D(A)}\rangle_f$) of labeled IF3 chased from the 30S IC with unlabeled IF3 (1 µM unlabeled IF3 added). Horizontal reference lines show E values for *E0.95(s1,2)* (magenta), *E0.70(s3-6)* (blue), and *E0.15(s7,8)* (orange), all from PDA.

141

## 5.2.2 Discussion

The present analysis using an advanced toolbox for hpFRET indicates that free IF3 is a very dynamic molecule that can rapidly sample a large variety of different structural arrangements. At the extremes of the fluctuation states the distance between the FRET dyes in the NTD and CTD changes from 30Å to 69Å (Fig. 5.2e). In free IF3, structural changes can be explained by reversible association between IF3 domains; the ordered-to-disordered transition of the linker region; and the chain dynamics of the unstructured linker (Fig. 5.8). While several conformational states of IF3 have been suggested before based on structural information (Moreau et al, 1997), the functional relevance of the linker dynamics, the richness of the conformational space sampled by IF3, and the kinetics of the transitions were not appreciated. The rate constants of transitions are distributed over four orders of magnitude in time. The conformational landscape is shaped into three major pools by two rate-limiting transitions between: (i) the closed (*AU*) and open (*SU*) IF3 states (characteristic relaxation time of ~ 400 μs) (ii) the unstructured (*SU*) and structured (*SS*) linker (relaxation time of ~ 1.6 μs) and (iii) within the pool (*SU*), multiple isoenergetic IF3 conformations are sampling rapidly (relaxation time of ~ 60 ns). These multiple conformational states generate an almost continuous FRET profile which would not be amenable for analysis by conventional methods. In contrast, the combination of smMFD, seTCSPC, PDA, FCS, and the geometric simulations provide coherent, unique insights into the challenging structural dynamics of the molecule. These tools will pave the way to convert static structural information into dynamic space of molecules in motion.

The intrinsic dynamics of IF3 appears to be faster than its binding to the 30S subunit, estimated to >1 ms based on rapid kinetic studies (Milon et al, 2012). The 30S subunit does not select a particular conformation of free IF3, because binding of IF3 to the 30S subunit does not shift the equilibrium between different IF3 conformations in the free pool (Fig. 5.4a,b). Upon binding, IF3 rearranges into a static open state with FRET efficiency *E0.22(sIC)* corresponding

to 64 Å between the reporters (Fig. 5.8). A similar conformation of IF3 bound to the 30S subunit was observed by single molecule FRET measurements in a TIRF setup, albeit with the fluorescence reporters placed at different positions of NTD and CTD (Elvekrog & Gonzalez, 2013). The static state is enforced by the IF3 recruitment to the 30S subunit is not affected by other initiation components as long as start codon recognition does not take place, i.e. in the 30S PIC (Fig. 5.5). We did not observed IF1 and IF2-dependent conformational changes of IF3 described by Elvekrog and Gonzalez (Elvekrog & Gonzalez, 2013). We assume that the differences in experimental setups, particularly the mRNAs used, labeling positions, buffer conditions and the experimental techniques account for the apparent discrepancy. In structural terms, the 30S-bound open static *E0.22(sIC)* state represents the conformation of IF3 with its both domains bound to the 30S subunit in a stable 30S PIC conformation. In contrast, the fraction of IF3 in the dynamic state *E0.70(sIC)* may represent those molecules that bind to the 30S subunit with one of the domains only, whereas the other domain is free to move. Because the affinity of IF3 CTD is higher than that of the NTD (Fabbretti et al, 2007; Petrelli et al, 2001), it is likely that the molecules which are dynamic are bound with the CTD only, whereas the NTD has cleared the way for the 50S docking (Fabbretti et al, 2007). The dissociation of the NTD may reflect the appearance of an alternative 30S subunit arrangement with the dynamic properties different from those where both domains of IF3 are stably bound.

Recognition of the start codon leads to a remodeling of the 30S PIC towards the formation of the 30S IC (Milon et al, 2012). Although 30SPIC to 30S IC transition destabilizes IF3 binding (Milon et al, 2012), the binding affinity remains sufficiently high to retain most of IF3 in the 30S IC (Fig. 5.6), in agreement with the results of the TIRF studies (Elvekrog & Gonzalez, 2013; MacDougall & Gonzalez, 2014). This strongly disfavors models suggesting that IF3 is released from the ribosome prior to 50S subunit joining (Antoun et al, 2006) and suggests that IF3 dissociation must occur at a later step, e.g. during the 50S subunit joining or during the maturation of the 70S complex.
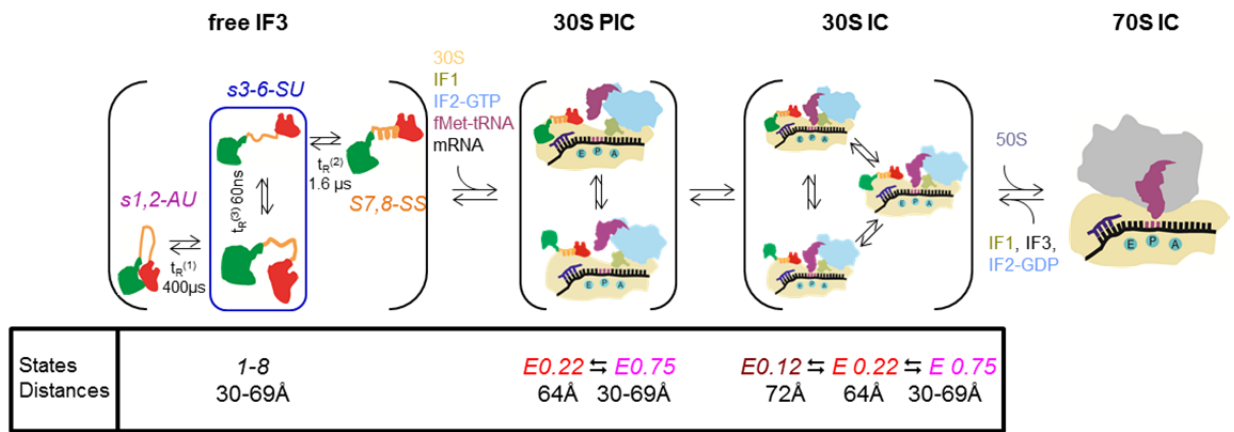
Figure 5.8 **Schematic of conformational transitions of IF3 free in solution and during translation initiation.** States *s1,2-AU*, *s3-6-SU* and *s7,8-SS* represent groups of different IF3 conformations. States *s1,2-AU* are in the associated domain arrangement with unstructured linker (very high and high FRET); states *s3-6-SU* have separated domains and unstructured linker with different FRET efficiencies from very high to low; states *s7,8-SS* have separated domains and structured linker. The equilibrium relaxation times are indicated. Upon 30S PIC formation, IF3 assumes a static open conformation with a 64 Å inter-dye distance (*E0.22(sIC)*) and a dynamic conformation (*E0.70(sIC)*) with distances between 30 and 69 Å. Upon recognition of the start codon, an additional conformation of IF3 is induced with 72 Å between the fluorophores (*E0.12(sIC)*). 50S subunit docking induces rapid dissociation of IF3 from 40-70% of 70S IC.

In 30S IC, a part of IF3 molecules changes the conformation to an extended static *E0.12(sIC)* state (Fig. 5.4d). The portion of the 30S-bound IF3 that undergoes this open (*E0.22*) to extended (*E 0.12)* transition correlates with the nature and strength of the translation initiation region. It is conceivable that the extended conformation is induced by the ribosome as a result of rearrangements initiated by start codon recognition, which is particularly clearly seen in the 30S IC with m002 AUG mRNA (Fig. 5.5). The cryo-EM structure of the same complex suggests that the head of the 30S subunit in the complex is found in a swiveled, rotated conformation (Julian et al, 2011). The extended *E0.12(sIC)* state may represent the IF3 arrangement in a tense 30S IC constrained by both SD-ASD interactions and the codon recognition, whereas the open *E0.22(sIC)* state represent a relaxed IF3 state which is similar in both 30S PIC and 30S IC. The exchange between the two conformations might be too slow compared to the diffusion time through the observation volume of our experiments (~ 2 ms); the observed proportion between the two conformations may reflect the differences in the thermodynamic stabilities of the two

states. We note that with infC mRNA the proportion of IF3 in the *E0.12(sIC)* state was higher on a AUU than on a canonical AUG codon, which may reflect the fact that AUU is a native start codon in infC, and hence the mRNA may be specifically adapted to dampen the effect of the codon-anticodon recognition. The two static states (*E0.22(sIC)* and *E0.12(sIC)*) and an additional dynamic population (*E0.70(sIC)*) of IF3 may reflect different conformations of the 30S subunit in complexes which are likely to be compositionally similar. The dynamic nature of conformational changes of the 30S subunit explains why the sensor of its conformation, i.e. IF3, has to be deformable as well.

Upon addition of the 50S subunit, a portion of IF3 (40-70% depending on the mRNA) dissociates and regains the conformational characteristics of the free IF3. The remaining portion of IF3 may remain bound to the 70S IC, as suggested recently by TIRF experiments (Elvekrog & Gonzalez, 2013). Our data are consistent with the notion that initiation may proceed via heterogeneous pathways (Tsai et al, 2012) and that IF3 may control initiation by stabilizing certain conformations of the initiation complex (Elvekrog & Gonzalez, 2013). Thus, the multiple layers of information that are encoded in the mRNA sequence are deciphered on the ribosome with the help of IF3, thereby contributing to the efficiency of translational control mechanisms in the cell. The fine control of the 30S subunit conformations is made possible by the inherent pliability of the IF3 linker and the adjustment of the IF3 domain orientation to a potentially wide variety of somewhat different 30S IC structures. The present experiments highlight the power of high-resolution confocal single-molecule approaches to delineate structural dynamics in complex, heterogeneous biological processes.

## ACKNOWLEDGMENTS

## AUTHOR CONTRIBUTIONS

D.R. performed all hpFRET measurements and analyzed data, S.F. and D.R. performed FCS analysis, P.M. designed IF3 construct for double labeling and performed biochemical experiments. All authors conceived the research and designed experiments. D.R., M.V.R. and C.A.M.S. wrote the paper with the input of S.F. and P.M.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

## 5.2 Online Methods

### 5.2.1. Biochemical methods.

30S subunits, IF1, IF2, and fMet-tRNA$^{fMet}$ were prepared as described (Milon et al, 2007). IF3 labeling was carried out as following. Wild-type IF3 contains a native Cys at position 65 in the NTD, which is poorly accessible for labeling with maleimides (<5% in 1 hour at non-denaturing conditions). To prepare double-labeled IF3, a second cysteine has been introduced by replacing E166 at a solvent-exposed position of the IF3 CTD by site specific mutagenesis. The double cysteine mutant of IF3 (100 μM) was first labeled for 30 min at RT with a 20-fold molar excess of Alx647-maleimide in labeling buffer (50 mM Tris-HCl pH 7.1, 200 mM NH$_4$Cl, 5% glycerol). The labeling reaction was stopped by addition of 2-mercaptoethanol (6 mM) and purified from the excess of the dye using SP FF column (GE healthcare) by a step gradient elution with 1 M NH$_4$Cl. The eluted protein was preferentially labeled at Cys166 as compared to the labeling of wild-type IF3 under the same conditions (< 5%). In order to overcome the low reactivity of Cys65, Alx647-labeled IF3 was dialyzed against labeling buffer containing 2 M urea and subsequently reacted with Alx488-maleimide for 2 hours at RT. The reaction was stopped by addition of 2-mercaptoethanol (6 mM) and the protein was purified from the excess of the dye as described above. Spectrometric characterization of the double-labeled IF3 indicated the presence of one molecule of each dye covalently attached to IF3. Labeling sites were characterized by a 15% SDS-PAGE. Isolated IF3 domains were prepared by mild trypsin treatment. The fluorescence of the gels was imaged by FL7000 scanner (GE Health Care). All 30S complexes were prepared in buffer A (50 mM Tris-HCl, pH 7.5, 70 mM NH$_4$Cl, 30 mM KCl, 7 mM MgCl$_2$) at room temperature.

### 5.2.2. Multiparameter Fluorescence Detection (MFD)

The custom-made confocal setup for hpFRET utilized a diode laser (LDH-D-C 485, repetition rate at 64 MHz, PicoQuant, Germany) to excite the probe at 485 nm wavelength

through 60X, 1.2 NA collar (0.17) corrected Olympus objective. The fluorescence signal was collected through the same objective and spatially filtered using a 100 μm pinhole. This resulted in a confocal detection volume of several femtoliter (fl), where freely diffusing molecules were detected. The signal was divided into parallel and perpendicular components at two different colors ("green" and "red") through band pass filters, HQ 520/35 and HQ 720/150, for green and red respectively, split further with 50/50 beam splitters and acquired by eight separate detectors: four for green detection (tau-SPAD, PicoQuant, Germany) and four for red detection (APD SPMC-CD-2965, Perkin-Elmer, Germany). For the data registration and time counting a time-correlated single photon counting (TCSPC) module from PicoQuant (HydraHarp 400, Picoquant, Germany) was used. For MFD measurements IF3 was diluted to a picomolar concentration in buffer A. In combination with fl detection volume, this resulted in a final concentration of 0.1 molecules in the observed volume, which allowed us to analyze the bursts of each individual diffusing molecule. The size of the detection volume was controlled daily by FCS using the diffusion time of Rhodamin-110 in water $t_{diff} = 0.2$ ms.

## 5.2.3. FRET Data analysis

To generate the 2D plots we performed bursts analysis of the collected data. Bursts were selected by $2\sigma$ criterion out of the mean background value with a minimum of 60 photons for each burst. Each burst was then fitted using maximum likelihood algorithm (Maus et al, 2001) resulting in a single lifetime for each molecule. The results were plotted as 2D histograms combined with two 1D projections by custom-made software (LabVIEW, National Instruments Co.). For all measurements, we used Alexa 488 as donor dye and Alexa 647 as acceptor dye. This dye pair has a Förster radius $R_0 = 52$ Å assuming the orientation factor $\kappa^2 = 2/3$.

The intensity-based FRET efficiency (E) was obtained by correcting the measured green and red signal $S_G$ and $S_R$, for mean green and red background $\langle B_G \rangle$ and $\langle B_R \rangle$, spectral crosstalk ($\alpha = 1.7\%$), ratio of green and red detection efficiencies ($g_G/g_R = 0.77$) and donor and acceptor

fluorescence quantum yield $\Phi_{FD(0)} = 0.80$ and $\Phi_{FA} = 0.44$. As derived from (Sisamakis et al, 2010) the final equation is:

$$E = \dfrac{\dfrac{S_R - \alpha \cdot \left(S_G - \langle B_G \rangle\right) - \langle B_R \rangle}{g_R}}{\dfrac{\Phi_{FA}}{\Phi_{FD(0)}} \cdot \left(\dfrac{S_G - \langle B_G \rangle}{g_G}\right) + \dfrac{S_R - \alpha \cdot \left(S_G - \langle B_G \rangle\right) - \langle B_R \rangle}{g_R}}$$

(Eq. 5.1)

**Theoretical static FRET line.** Considering the static case (i.e. molecule without dynamics) the FRET efficiency $E_{\text{static}}\left(\tau_{D(A)}\right)$ can be described as a function of donor lifetime in presence of acceptor dye by Eq. 5.2

$$E_{\text{static}} = 1 - \frac{\tau_{D(A)}}{\tau_{D(0)}}$$

(Eq. 5.2)

where $\tau_{D(0)}$ is the donor fluorescence lifetime in the absence of acceptor and $\tau_{D(A)}$ is the mean species lifetime. In the MFD analysis of single-molecule bursts, the maximum likelihood estimator determines only fluorescence weighted average lifetimes $\langle \tau_{D(A)} \rangle_f$ due to the low photon number (~100 photons per burst). To count for dye-linker dynamics, a numerical simulation was performed to create an empirical relation between the species and fluorescence average lifetimes for the selected range of $\langle R_{DA} \rangle$ values using an $4^{\text{th}}$ order polynomial function (Kalinin et al, 2010a; Sisamakis et al, 2010). Using this polynomial function the species lifetimes of the eTCSPC analysis were converted to $\langle \tau_{D(A)} \rangle_f$ in order to compare the results of both methods (Fig 5.2e.) By substituting $\tau_{D(A)}$ in Eq. 5.2 with obtained $4^{\text{th}}$ order polynomial function we can represent the static FRET line as:

$$E_{\text{static,L}} = 1 - \frac{\sum_{i=0}^{4} C_{i,L} \left(\left\langle \tau_{D(A)} \right\rangle_{f,L}\right)^i}{\tau_{D(0)}}$$

(Eq. 5.3)

where $C_{i,L}$ are coefficients of the polynomial function. In our case, the values resulting in the static FRET line such as shown in Fig 5.2 (orange line) were: the width of the linker broadening 12 Å (Case 1 in supplemental of (Kalinin et al, 2010b)), $\tau_{D(0)} = 4.085$ ns and polynomial equation coefficients $C_4 = 0.0078$, $C_3 = -0.1051$, $C_2 = 0.4437$, $C_1 = 0.4137$ and $C_0 = -0.0254$.

**Dynamic FRET line.** Considering the dynamic case (i.e. molecules with conformational dynamics), the same procedure was applied to calculate a dynamic FRET line (Sisamakis et al, 2010) where mixed fluorescent species were the result of an exchange between two fluorescent states with the lifetimes of $\langle \tau_1 \rangle_f$ and $\langle \tau_2 \rangle_f$:

$$E_{dyn,L} = 1 - \frac{\langle \tau_1 \rangle_f \cdot \langle \tau_2 \rangle_f}{\left( \langle \tau_1 \rangle_f + \langle \tau_2 \rangle_f - \sum_{i=0}^{3} C_{i,L} \left( \langle \tau_{D(A)} \rangle_f \right)^i \right)}$$

(Eq. 5.4)

The values used in this paper were:

Purple: $\tau_{D(0)} = 4.1$ ns, $\langle \tau_1 \rangle_f = 0.7$ ns, $\langle \tau_2 \rangle_f = 3.5$ ns and for the polynomial equation coefficients $C_3 = 0.0$, $C_2 = 0.0$, $C_1 = 1.7099$ and $C_0 = -2.4794$.

Magenta: $\tau_{D(0)} = 4.1$ ns, $\langle \tau_1 \rangle_f = 0.4$ ns, $\langle \tau_2 \rangle_f = 3.3$ ns and for the polynomial equation coefficients $C_3 = 0.0$, $C_2 = 0.0$, $C_1 = 2.1199$ and $C_0 = -3.9170$.

The blue dot in Figure 5.2a shows the center of the average of the populations *E0.70(s3-6-SU)* without taking in account the stacking domain effects of populations *AU* and linker effects of the population *SS*. This point was determined by crossing between horizontal for *E0.70(s3-6-SU)* and a theoretical static line which takes in account the broadening due to the dynamics of *E0.70(s3-6-SU)*. The simplest way to calculate this line is from the numerical simulation of the population *SU* (blue histogram in Fig. 5.3a). Fitting this histogram with a Gaussian distribution yielded the width of 17.8 Å. Combining this value with dynamic linker broadening (width = 12 Å) the final broadening was calculated as a simple error propagation case: width = (width$_1^2$

+width$_2^2$)$^{1/2}$ resulting in the final broadening of 21.5 Å. The width is a parameter in the *4th* order polynomial equation as shown above. The final parameters for the static line corrected for overall broadening (half width 10.75 Å) were: $\tau_{D(0)}$ = 4.1 ns and for the polynomial equation coefficients $C_4$ = 0.0074, $C_3$ = -0.1682, $C_2$ = 0.9956, $C_1$ = -0.8533 and $C_0$ = 0.3679. The intercept point between this line and the horizontal *E0.70(s3-6-SU)* missed the measured population, confirming the need for additional populations to describe the experimental results.

**Fluorescence lifetime analysis**. Fluorescence intensity decays were obtained from single molecule data and fitted using the iterative re-convolution approach with various models using single, double or triple exponential decays(Fries et al, 1998). In general, these multi-exponential relaxation models can be described by

$$F_{D(A)}(t) = \left(1 - X_{D(0)}\right) \cdot \sum_i x_{D(A)}^{(i)} \exp(-t / \tau_{D(A)}^{(i)}) + X_{D(0)} \cdot \sum_j x_{D(0)}^{(j)} \exp(-t / \tau_{D(0)}^{(j)}) \qquad \text{(Eq. 5.5)}$$

where $\tau_{D(0)}^{(i)}$ is the donor fluoresce lifetime, $x_{D(0)}^{(i)}$ are the pre-exponential factors. $X_{D(0)}$ is the donor only labeled fraction, $\tau_{D(A)}^{(i)}$ is the performing FRET donor's lifetime with corresponding pre-exponential $x_{D(A)}^{(i)}$ factor. Due to selection procedure, we assumed $X_{D(0)}$ = 0 (Suppl. Fig. 5.S1a). In Suppl. Fig. 5.S1c, we present the data fitting for free IF3 using two and three exponentials – three exponential fit shows $\chi^2$ of 1.15 and flat autocorrelation curve. In comparison, two exponential shows $\chi^2$ of 1.21, while the autocorrelation curve is curved.

**FRET efficiency** *E* values are calculated from mean species lifetimes (seTCSPC) by Eq. 5.2 using $\tau_{D(0)}$=4.15 ns or from the $\langle R_{DA} \rangle_E$ values obtained by PDA by

$$E = \frac{1}{1 + \left(\dfrac{\langle R_{DA} \rangle_E}{R_0}\right)^6}, \qquad \text{(Eq. 5.6)}$$
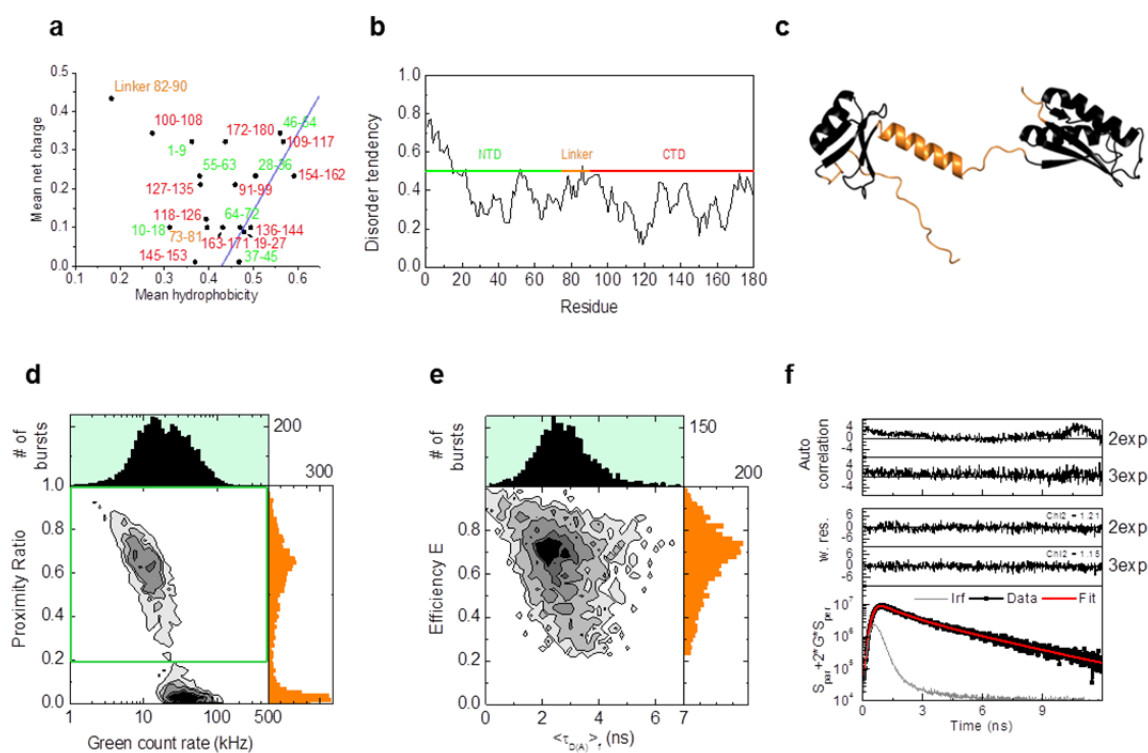
Supplementary Figure 5.S1 **Detailed seTCSPC analysis and the selection of flexible regions for the geometric Markov Chain Monte-Carlo sampling.** (**a**) 2D plot of proximity ratio ($S_R/(S_G + S_R)$) vs. green count rate ($S_G$) for free IF3. The bursts with very high FRET, which were not visible in Fig. 2a due to the restriction of the lifetime determination by the Maximum Likelihood estimator to a minimum photon number (at least 20 green photons), are clearly visible in this plot with low green count rate and proximity ratio close to 1. The green box indicates the bursts that were selected and accumulated for the lifetime analysis in Fig. 5.2b. (**b**) Final FRET efficiency (*E*) vs. fluorescence lifetime ($\langle\tau_{D(A)}\rangle_f$) 2D plot after burst selection. (**c**) Fluorescence lifetime analysis of the main population from the 2D smFRET plot comparing two- and three-exponential fits. Respective autocorrelations are shown on top, weighted residuals and the $\chi 2$ values in the middle, bottom frame shows experimental data of total signal S (black), the instrument response function (grey), and the three exponential fit (red). (**d**) Analysis of IF3 sequence. The sequence was divided in 9 amino acids-long fragments (as indicated) to calculate mean net charge and mean hydrophobicity. The blue line indicates the separation for structured (to the right) and unstructured (to the left) proteins according to ref (Hofmann et al, 2012). (**e**) Alternative analysis of the IF3 sequence. The sequence of the IF3 was analysis using IUPRED software (iupred.enzim.hu). Linker region (orange) shows intermediate tendency for disorder, indicating that it can have either structured or unstructured conformation. (**f**) Structure of IF3 used in the Markov Chain Monte-Carlo simulations. Regions depicted in black were kept fixed in all simulations; regions depicted in orange were flexible. Simulations were done either applying the attractive forces that kept the connecting helical part structured (Fig. 5.3a, orange) or without such forces, which resulted in a much longer linker region and broader distance distribution (Fig. 5.3, blue).

where $\langle R_{DA} \rangle_E$ – is the mean FRET averaged distance between donor and acceptor and $R_0$ is the Förster distance.

**Mean interdye distance** $\langle R_{DA} \rangle$ values were also calculated based on fluorescence decay lifetimes from bulk ensemble measurements (eTCSPC) or selected single molecule bursts (seTCSPC) as

$$\langle R_{DA} \rangle = R_0 \cdot \frac{1}{\sqrt[6]{\dfrac{\tau_{D(0)}}{\tau_{D(A)}} - 1}} \qquad \text{(Eq. 5.7)}$$

**4. Photon distribution analysis (PDA).** To perform PDA the raw photons trace from experimental data were accumulated in 1 ms time windows. FRET efficiency $E$ values were calculated based on donor and acceptor fluorescence intensity $F_D$ and $F_A$ and respective fluorescence quantum yields $\Phi_{FD(0)}$ and $\Phi_{FA}$:

$$E = \frac{F_A / \Phi_{FA}}{F_D / \Phi_{FD(0)} + F_A / \Phi_{FA}} \qquad \text{(Eq. 5.8)}$$

for each window and presented in a 1D histogram (Antonik et al, 2006). The time windows with less than 13 photons or more than 1000 photons were discarded. The FRET efficiency ($E$) and donor to acceptor fluorescence intensity ratio ($F_D/F_A$) histograms were obtained by correcting the measured green and red signal $S_G$ and $S_R$, for mean green and red background $\langle B_G \rangle$ and $\langle B_R \rangle$, spectral crosstalk ($\alpha$ = 1.7%), ratio of green and red detection efficiencies ($g_G/g_R$ = 0.77) and donor and acceptor fluorescence quantum yield $\Phi_{FD(0)}$ = 0.80 and $\Phi_{FA}$ = 0.44. Resulting histograms were analyzed using model which accounts for three different types of FRET distributions (Kalinin et al, 2007): single $E$ (donor-only species), Gaussian distributed static species and dynamic species (Kalinin et al, 2010b; Santoso et al, 2010). Up to two Gaussians, two dynamic and one single E (donor only) were used to fit the $E$ histograms.

For analyzing the results of titration with the 30S subunits (Fig. 5.4d) the following model was used: for free IF3 two distance distributions were fixed to following values: population with fast dynamics approximated by a single Gaussian distribution with $\langle R_{D(A)} \rangle_E$ =45 Å and w = 8 Å and population with slow dynamics with exchange between the end distances of 30 Å and 69 Å; for the static bound IF3 $\langle R_{D(A)} \rangle_E$ =63 Å and w = 9 Å. The amplitudes of all populations and rate constants of the slow exchange dynamics were kept free. The relaxation time of the slow dynamics $t_R \sim 300 \pm 100$ µs did not change significantly with increasing 30S concentration.

For the analysis of the mRNA dissociation (Fig. 5.7a) the last 50 minutes of the corresponding measurement were combined into an $E$ 1D histogram and analysed using the populations and values described above.

## 5.2.5. Fluorescence Correlation Spectroscopy (FCS)

**Conformational dynamics of free IF3**. To study the conformational dynamics of free IF3 with high temporal resolution, FCS curves (green-to-green (GG) and red-to-red (RR) autocorrelations, green-to-red and red-to-green (GR and RG) color cross-correlations) were computed from smMFD data by custom-made software correlator (LabVIEW, National Instruments Co.). The data were analyzed (Felekyan et al, 2013) using the following equations:

$$G_{G,G}(t_c) = 1 + \frac{1}{N_{G,G}} \cdot G_{diff}(t_c) \cdot \left[ 1 - \sum_{i=1}^{4} \left( AC_{G,G}^{(i)} \cdot \left( 1 - e^{-\frac{t_c}{t_R(i)}} \right) \right) \right]$$

$$G_{R,R}(t_c) = 1 + \frac{1}{N_{R,R}} \cdot G_{diff}(t_c) \cdot \left[ 1 - \sum_{i=1}^{4} \left( AC_{R,R}^{(i)} \cdot \left( 1 - e^{-\frac{t_c}{t_R(i)}} \right) \right) \right] \qquad \text{(Eq. 5.9)}$$

$$G_{G,R}(t_c) = 1 + \frac{1}{N_{G,R}} \cdot G_{diff}(t_c) \cdot \left[ 1 - \sum_{i=1}^{4} \left( CC_{G,R}^{(i)} \cdot e^{-\frac{t_c}{t_R(i)}} \right) \right]$$

where $N_x$ is an average number of molecules for corresponding correlation, $G_{diff}(t_c)$ – diffusion term, $AC_{G,G}^{(i)}$ and $AC_{R,R}^{(i)}$ – fractions of respective bunching terms, respectively, $CC_{G,R}^{(i)}$ – fractions of respective anti-correlation terms in cross-correlation curves, and $t_R^{(i)}$ – is the $i$-th relaxation time shared in $i$-th bunching or anti-correlation terms in global fit of FCCS curves.

Diffusion term $G_{diff}(t_c)$ is assumed under assumption of a 3-dimensional (3D) Gaussian shaped detection/illumination volume:

$$G_{diff}(t_c) = \left(1 + \frac{t_c}{t_{diff}}\right)^{-1} \cdot \left(1 + \left(\frac{\omega_0}{z_0}\right)^2 \cdot \frac{t_c}{t_{diff}}\right)^{-\frac{1}{2}}$$
(Eq. 5.10)

where $t_{diff}$ is the diffusion time and $\omega_0$ and $z_0$ are shape parameters of the detection volume which is defined by $w(x,y,z) = \exp(-2(x^2 + y^2)/\omega_0^2)\exp(-2z^2/z_0^2)$.

**Titration of IF3 with the 30S subunit**. To determine the fractions of IF3 free and bound to the 30S subunit via their characteristic diffusion times $t_{d-free}$ and $t_{d-complex}$, respectively, the following equation was used to fit auto-correlation curves $G_{G,G}(t_c)$ of the green signal $S_G$:

$$G_{G,G}(t_c) = 1 + \frac{1}{N} \cdot \left(x_{free} \cdot G_{diff-free}(t_c) + (1 - x_{free}) \cdot G_{diff-complex}(t_c)\right) \cdot \left(1 - T + T \cdot e^{-\frac{t_c}{t_T}}\right)$$
(Eq. 5.11)

where $t_c$ is the correlation time, $N$ is the total number of molecules, $x_{free}$ is the fraction of the unbound IF3 with the diffusion term $G_{diff-free}$, $G_{diff-complex}$ is the diffusion term of IF3 bound to the 30S subunit, and $T$ is the fraction of molecules in triplet (dark) state with the characteristic triplet time $t_T$.

## 5.2.6. Accessible Volume (AV) modeling and geometrical simulations

The AV method was used to predict the position ensemble of a fluorescent label based on its chemical structure and environment (Sindbert et al, 2011a; Woźniak et al, 2008). For the simulation we used the following dye parameters: for Alexa488 – length 20 Å, width 4.5 Å and dye radius 3.5 Å; for Alexa647 – length 22 Å, width 4.5 Å and dye radius 3.5 Å (Fig 5.1b). AV simulations were used to estimate the distances between the two fluorophores in the structural models of IF3 produced by geometrical simulations.

To estimate the possible inter-dye distances in IF3, Markov Chain Monte-Carlo sampling was performed on the connecting linker (residues 72 to 90) between the N- and C-terminal domains using the full length protein model. The initial model was constructed using MODELLER (Rothwell et al, 2013) utilizing the crystal structures of the individual subdomains (PDB: 1tif (Biou et al, 1995) and 2ife (Biou et al, 1995; Moreau et al, 1997)). The initial homology model was mapped to the reduced representation of the protein consisting solely of the C-, $C_\alpha$-, N-, O- and the hydrogen atoms forming the NH-O bonds. The repulsion between the atom pairs (O, N), (C, O) and (C, N) were modeled as repulsive quadratic potential (Kalinin et al, 2012). The sampling was performed on a reduced initial model where only $\varphi$ and $\psi$ torsion angles of the connecting $\alpha$-helical linker were displaced in each iteration step by random value taken from a Gaussian-distribution with a width of 0.002 rad. Two scenarios were used for sampling: (i) the $\alpha$-helical structure of the linker was preserved by modeling the existing hydrogen bonds as simple scaled attractive potential (Lippincott & Schroeder, 1955), (ii) the linker was treated as flexible with no stabilizing internal interactions.

## 5.2.7. Simulation of the FRET data in complex kinetic schemes

To describe all experimental findings, (i.e. the experimental 2D histogram E vs. $\langle\tau_{D(A)}\rangle_f$ (Fig. 5.2a) seTCSPC (Fig. 5.2b) and the cross correlation curve $G_{G,R}(t_c)$ (Fig. 5.2d)), an eight-

state scheme was used. As the first step, we calculated FRET histograms (Gopich & Szabo, 2012a). A kinetic model with discrete conformations was assumed, with the transition between the states described by rate equations. The probability for the system to be in state $i$ at time $t$, $p_i(t)$, satisfies a set of rate equations, which can be written in matrix notation as:

$$\frac{d\boldsymbol{p}}{dt} = \boldsymbol{K} \cdot \boldsymbol{p} \qquad \text{(Eq. 5.12)}$$

where $\boldsymbol{p}$ is a column vector with the components $p_i(t)$ and $\boldsymbol{K}$ is a transition rate matrix representing the rate constants for the transitions between states $i$ and $j$. At long times, $\boldsymbol{p}(t)$ approaches its equilibrium value, $\boldsymbol{p}_{eq}$. The vector of the equilibrium populations $\boldsymbol{p}_{eq}$ is normalized to 1 and satisfies $\boldsymbol{K} \cdot \boldsymbol{p}_{eq} = 0$. For each burst mean averaged efficiency $\langle E \rangle$ and average fluorescence weighted lifetime $\langle \tau_{D(A)} \rangle_f$ can be calculated by:

$$\langle E \rangle = \frac{\sum t_i \cdot E_i}{t_{burst}} \qquad \text{(Eq. 5.13)}$$

and

$$\left\langle \tau_{D(A)} \right\rangle_f = \frac{\sum t_i(\boldsymbol{K}) \cdot \tau_i^2}{\sum t_i(\boldsymbol{K}) \cdot \tau_i} \qquad \text{(Eq. 5.14)}$$

where $t_i(\boldsymbol{K})$ is time spent by a molecule in state $i$ within the duration of the burst, which depends on the transition rate matrix $\boldsymbol{K}$; $E_i$ is the FRET efficiency of the $i$-th state; $t_{burst}$ is the duration of the burst and $\tau_i$ is the fluorescence lifetime of the $i$-th state. Practically, each burst has certain duration and number of photons, which were chosen arbitrary from experimentally measured $t_{burst}$ (duration time) vs. $N$ (number of photons) 2D histogram. The residence times spent by the molecule in different states were calculated using Gillespie algorithm for continuous-time Markov Chain. Then, average fluorescence lifetime $\langle \tau_{D(A)} \rangle_f$ for each burst was calculated by Monte-Carlo simulation of fluorescence emission given FRET efficiencies of each state,

stationary (equilibrium) populations of states were obtained by solving interstate transition dynamics matrix and the residence times obtained on previous step. The following descriptions for the vector $p$ and the rate matrix $K$ (resulting into the equilibrium fractions for the state i, $p_{eq,i}$) and the experimental observables, $E$ and $\tau$, were used in the simulations to recapture the experimental 2D histogram:

| Kinetic state i | State name | Efficiency $E$ | Fluorescence lifetime $\tau$, ns [a] | Equilibrium fraction $p_{eq,i}$ |
|---|---|---|---|---|
| 1 | E0.98(s1-AU) | 0.98 | 0.08 | 0.24 |
| 2 | E0.85(s2-AU) | 0.85 | 0.62 | 0.07 |
| 3 | E0.98(s3-AU) | 0.98 | 0.08 | 0.24 |
| 4 | E0.85(s4-AU) | 0.85 | 0.62 | 0.13 |
| 5 | E0.50(s5-AU) | 0.5 | 2.08 | 0.09 |
| 6 | E0.15(s6-AU) | 0.15 | 3.53 | 0.04 |
| 7 | E0.15(s7-AU) | 0.15 | 3.53 | 0.04 |
| 8 | E0.15(s8-AU) | 0.15 | 3.53 | 0.15 |

[a] calculated by Eq. 5.2 using $\tau_0 = 4.15$ ns

$$K = \begin{pmatrix}
-2 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\
0 & -2 & 0 & 4 & 0 & 0 & 0 & 0 \\
2 & 0 & -30002 & 7000 & 7000 & 0 & 0 & 0 \\
0 & 2 & 15000 & -22004 & 7000 & 0 & 0 & 0 \\
0 & 0 & 15000 & 15000 & -29000 & 7000 & 0 & 0 \\
0 & 0 & 0 & 0 & 15000 & -7580 & 570 & 0 \\
0 & 0 & 0 & 0 & 0 & 580 & -571.1 & 4 \\
0 & 0 & 0 & 0 & 0 & 0 & 1.1 & -4
\end{pmatrix}$$

The simulation procedure was repeated for a high number of bursts to generate E vs $\langle \tau_{D(A)} \rangle_f$ 2D histogram (Fig. 5.3d). Additionally, assuming 3D Gaussian detection volume, the corresponding correlations for green-green ($G_{G,G}$), red-red ($G_{R,R}$) and green-red ($G_{G,R}$) signal can

be calculated with the parameters acquired from the simulations. The resulting values for the simulation and experimental values are shown below and are very similar overall.

| Relaxation times (amplitudes) | $t_R^{(1)}$ | $t_R^{(2)}$ | $t_R^{(3)}$ |
|---|---|---|---|
| Simulation | 762 µs (0.536) | 1.6 µs (0.154) | 73 ns (0.237) |
| | 366 µs (0.014) | | 34 ns (0.014) |
| | 207 µs (0.009) | | 22 ns (0.036) |
| Average weighted relaxation time | 742 µs (0.559) | 1.6 µs (0.154) | 65 ns (0.287) |
| | | | |
| Experiment (FCS) | 474 µs (0.110) | 1.6 µs (0.348) | 63 ns (0.542) |



Supplementary Figure S2 **Schematics of mRNAs used in this work.** (**a**) SD region is highlighted red, AUG start codon in green, non-canonical start codons in orange, and a natural AUU codon in the infC mRNA blue. (**b**) Structural cartoon of mRNA (green) with the SD sequence interacting with the aSD sequence of 16S rRNA on the 30S subunit. Initiator tRNA[fMet] is bound to the start codon AUG separated by a spacer (blue) from the SD sequence.

# 6. Chemical denaturation pathway of T4 Lysozyme and the effect of non-ionic surfactant Tween 20

This chapter is based on collaborative work in a joint project of following persons: Dmitro Rodnin, Hugo Sanabria, Katherina Hemmen, and Claus A.M. Seidel. The contributions to this chapter are as follows: H.S. and K.H. purified and labeled the protein. D.R. H.S., and K.H. and measured and analyzed the FRET experiments. C.A.M.S. supervised the project

DR: 45%

KH: 15%

CAMS: 20%

HS: 20%

## 6.1 Introduction to protein unfolding

As described in the introduction chapter 1.1, correct folding of proteins is the pre-requisite for their correct function; thus, the understanding of folding and unfolding pathways is a key topic in protein science. Protein folding and unfolding in *in vitro* model systems is believed to unravel the fundamental properties of protein dynamics in living cells and provide a detailed insight into the mechanism of folding. There are plenty of denaturants which cause unfolding and many methods which are suitable to study it. The most commonly used chemical denaturants are e.g. chaotropic agents such as urea (Cellitti et al, 2007b), guanidine hydrochloride (Cellitti et al, 2007b; Kiefhaber, 1995; Llinas et al, 1999b; Parker & Marqusee, 1999); acidic (Llinas & Marqusee, 1998; Lopez-Alonso et al, 2010) or basic denaturants (Bieri et al, 1999)); temperature (Bachmann et al, 2005; Cellitti et al, 2007b), pressure (Prigozhin et al, 2013), force unfolding (Jagannathan et al, 2012) or proteolysis (Hilz et al, 1975). Various different methods can be utilized to monitor unfolding, among them most commonly known are circular dichroism spectroscopy, stopped flow (Kiefhaber, 1995), nuclear magnetic resonance spectroscopy (NMR)(Llinas et al, 1999b; Parker & Marqusee, 1999) or optical tweezers (Jagannathan et al, 2012). Methods with high temporal resolution such as fluorescence spectroscopy, are well suited to study protein folding, as high resolution allows for detection of infrequently populated or thermally excited states. Notably, the predicted upper limit for the rate of protein folding is expected to be in the range of 0.5-10 µs (Bowman & Pande, 2010a; Kubelka et al, 2004), which fits well into the temporal resolution of smFRET experiments.

T4L is a protein well studied under different denaturing conditions (Cellitti et al, 2007a; Chang & Li, 2002; Gao et al, 2010; Llinas et al, 1999b). In the simplest scenario T4L is expected to undergo a single transition between two states $C$(folded)↔$U$(unfolded) with $C$ being a single basin or a single native folded state (Lint et al, 1992; Onuchic & Wolynes, 2004). However, there is strong evidence for the existence of an intermediate state on the unfolding pathway (Cellitti et al, 2007a; Chen et al, 2003). This would call for an expansion of the unfolding pathway to at least a three states mechanism $C$↔$I$(intermediate)↔$U$ as proposed by several groups (Bowman & Pande, 2010b; Cellitti et al, 2007a). Furthermore, at least three native states coexist in equilibrium (T4L manuscript). With these characteristics, the question arises of how the three folded functional states, $C_1$, $C_2$ and $C_3$, are connected to the intermediate and the unfolded states? It appears reasonable to assume the existence of an ensemble of intermediate ($I_{1-N}$) states and an ensemble of unfolded ($U_{1-N}$) states (Fig. 6.1). This resembles a kinetic network as proposed by Bowman & Pande (Bowman & Pande, 2010b). However, it is yet unclear, whether any of these intermediate states correspond to one of the folded conformations, e.g.

161

whether the transition from $C_1$ to $U_1$ goes through the state $C_2$. An insight into the potential unfolded state(s) can be gained from the seminal work of Llinas et al. (Llinas & Marqusee, 1998) where a T4L conformation with an unfolded N-terminal domain and folded C-terminal domain was found under acidic conditions.

In general, the following specific questions arise:

- Number and structure of the proposed states $I_{1-N}$ and $U_{1-N}$ – how many states can be detected under given denaturing conditions and what are their structures?

- Connectivity of the states – which states appear or disappear and at which conditions? What is the mechanism and the energetic landscape of unfolding?

- Kinetics under denaturing conditions – what are the rate constants of fluctuations between different protein conformations and how do they change at different concentrations of a denaturant?

- Comparison between different denaturants – how does the type of denaturation influence the unfolding pathway, excited states and kinetics? Due to the setup limitations, smFRET mostly varies buffer conditions to compare different types of denaturation, e.g. urea, guanidine hydrochloride or basic/acidic pH.

To answer these questions the following unfolding experiments of T4L were performed.

- Unfolding of a T4L variant in different urea concentrations ranging from 0 M to 8 M in 16 steps. These conditions were chosen to focus on the midpoint transition of a given variant

- Measurements of a T4L variant at different acidic conditions with pH ranging from 7.5 to 1.5

The different chosen variants show the same behavior in denaturing experiments as determined by Katherina Hemmen (Hemmen, 2015) by CD spectroscopy in urea.

The states which are different from the native states $C_1/C_2/C_3$ are denoted as $X_{(FRET\ level)}^{(condition)}$ with the following FRET levels assigned as subscript: *HF* – high FRET (lifetime $0 - 1.2$ ns); *MF* – middle FRET (lifetime $1.2 - 2.5$ ns); *LF* – low FRET (lifetime $2.5 - 3.8$ ns) and *NF* – no FRET (lifetime $3.8 - 4.3$ ns). The following conditions were used *U* – urea titration, *pH* – pH titration, *TU* – urea titration in the presence of Tween 20 and *TpH* – pH titration in the presence of Tween 20. This notation is used as superscript referring to the specific conditions of each experiment.
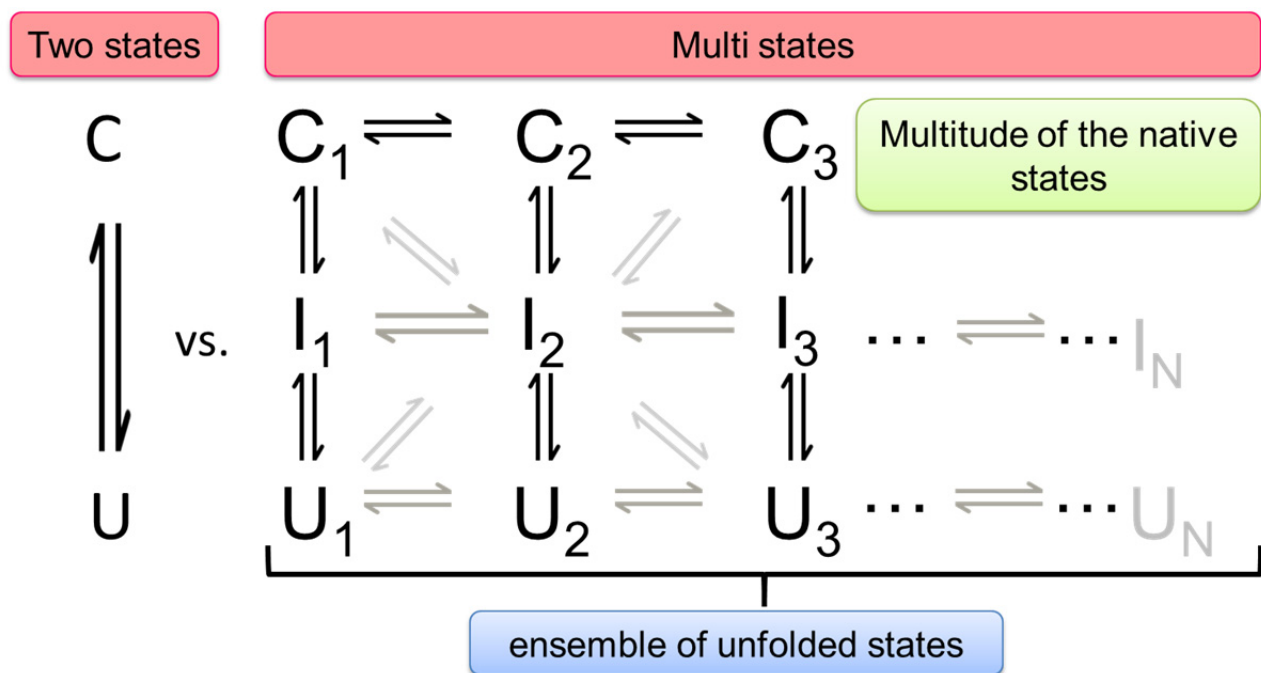
Figure 6.1. **Theoretical unfolding network for a three state protein like T4 lysozyme (Adapted from (Bowman & Pande, 2010b; Hemmen et al, 2014)).** The protein is fluctuating between three different conformations with all possible transitions between multiple intermediate and unfolded states being allowed. Some of the states may show the same FRET values, making the analysis difficult.

## 6.2 Urea and acidic pH dependent unfolding of T4 Lysozyme

### 6.2.1 Urea titration

Unfolding of the K60pAcF/N132C-(DA) variant of T4L was performed by adding different concentrations of urea, a well-established denaturing agent for studies of protein unfolding (Fig. 6.2). As described in Chapter 4, the three conformational states $C_1$, $C_2$ and $C_3$ were detected for this variant with the distances of 49.8 Å, 41.4 Å and 36.5 Å. The states $C_2$ and $C_3$ cannot be distinguished in the analysis of the 2D histogram, as the difference between the states is less than their respective broadness ($\sigma = 6$ Å). Therefore in this chapter for this variant the states $C_2$ and $C_3$ are combined into a single mixed state denoted $C_2$.

Urea (chemical formula $CO(NH_2)_2$) is a powerful protein denaturant which disrupts the noncovalent bonds in the proteins. Upon increase of the urea concentration two transitions are observed: the first transition $T_1^U$ starts at 4 M urea and shows the transition (wine arrow in Figure 6.2) of the populations $C_1$ and $C_2$ (indicated by dashed blue and purple horizontal line in Figure 6.2 respectively) to the intermediate state $X_{NF}^U$ (indicated by grey horizontal line in

163

Figure 6.2) with No-FRET characteristics: fluorescence lifetime $\langle \tau_{D(A)} \rangle_f$ of ~ 4 ns, $F_D/F_A$ values ~ 100 and the interdye distance $R_{DA}^{(T1)} > 80\text{Å}$. The second transition $T_2^U$ (pink arrow in Figure 6.2) becomes visible from 4.75 M urea and occurs from the folded $C_1$ and $C_2$ states that are in
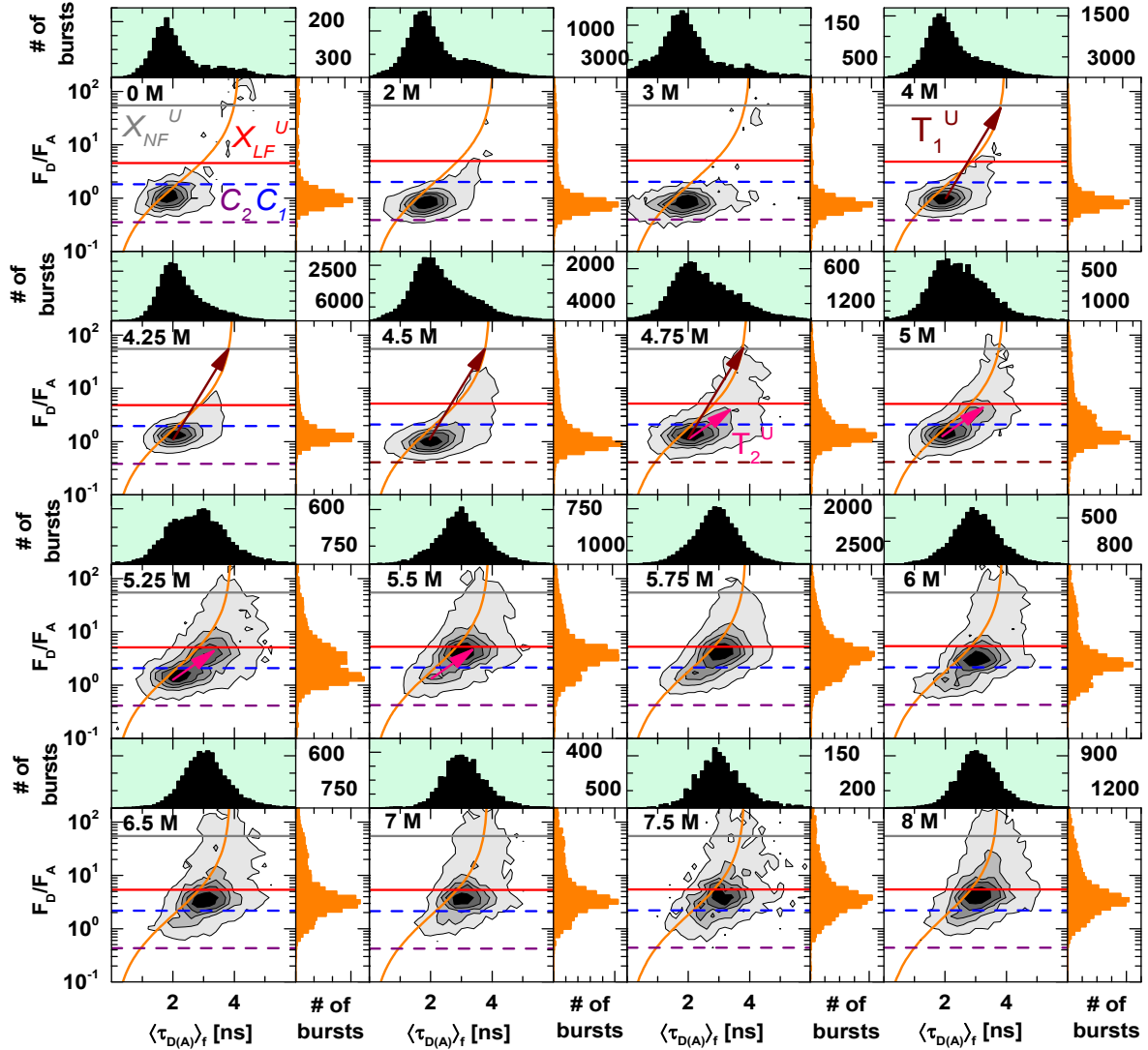


Figure 6.2. **Unfolding of K60pAcF/N132C-(DA) variant of T4L at different urea concentrations**. 2D histograms of $F_D/F_A$ vs $\langle \tau_{D(A)} \rangle_f$ with the respective 1D histograms are shown. The urea concentration is indicated in each panel. Horizontal lines indicate the positions of states $C_1$ (dashed blue), $C_2$ (dashed purple), $X_{NF}^U$ (grey) and $X_{LF}^U$ (red). The wine arrow indicates the transition $T_1^U$ from the $C_1/C_2$ equilibrium to the state $X_{NF}^U$ and the pink arrow indicates the transition $T_2^U$ from the $C_1/C_2$ equilibrium to state $X_{LF}^U$. Orange line represents static linker corrected FRET line as described in section 3.8.4. $F_D/F_A$ values were corrected by the ratio of green to red detection efficiency $g_G/g_R$, green and red background and crosstalk $\alpha$. All parameters are listed in Appendix Table 9.1.

equilibrium to the state $X_{LF}^{U}$ (indicated by red horizontal line in Figure 6.2) with the fluorescence lifetime $\langle\tau_{D(A)}\rangle_f$ of 3.2 ns, $F_D/F_A$ values of ~ 5 and the interdye distance $R_{DA}^{(T2)}$ ~ 64 Å. At 7.5 M and 8 M urea the equilibrium is shifted to the $X_{LF}^{U}$ state with less than 20% of the state $X_{NF}^{U}$ and no folded states $C_1$ and $C_2$. Overall, at least two new FRET populations and two transitions can be clearly distinguished in this titration.

## 6.2.2 pH titration

A detailed titration of the variant S44pAcPhe/I150C-(DA) was performed at different acidic pH values (Fig. 6.3). For all measurements from pH 7.5 up to pH 3, a single population with a fluorescence lifetime $\langle\tau_{D(A)}\rangle_f$ of ~2.7 ns and a broad population toward a short fluorescence lifetime $\langle\tau_{D(A)}\rangle_f$ of ~ 0.5 ns was detected. Both populations are slightly off the static FRET line (orange) indicating a dynamic equilibrium. Analysis by seTCSPC showed the previously identified conformational states $C_1$, $C_2$ and $C_3$ (see chapter 4) with lifetimes of ~3.6 ns (30%, dashed blue horizontal line), ~2.0 ns (46%, dashed purple horizontal line) and ~0.5 ns (24%, dashed dark yellow horizontal line), respectively (Fig. 6.3;Table 6.1). At pH 2.5 a single transition $T_1^{pH}$ (wine arrow in Fig. 6.3) can be identified manifested by the appearance of a very broad population with a lifetime of ~1.5 ns. The broadness and the form of this distribution indicate formation of a highly dynamic mixture of protein species; this notion is further supported by the observation that this population is off the static line (orange). At pH 2 and pH 1.5 the equilibrium between the native states and the unfolded states shifts strongly towards the unfolded state. seTCSPC analysis of pH 1.5 (Table 6.1) detected two populations $X_{NF}^{pH}$ (grey line) and $X_{HF}^{pH}$ (dark yellow) with lifetimes $\tau_{NF}^{pH}$ ~ 4 ns and $\tau_{HF}^{pH}$ ~ 0.4 ns. The strong effect of the acidic pH on the structure and folding kinetics of T4L is clearly demonstrated by the change of populations and a single complex kinetic transition.
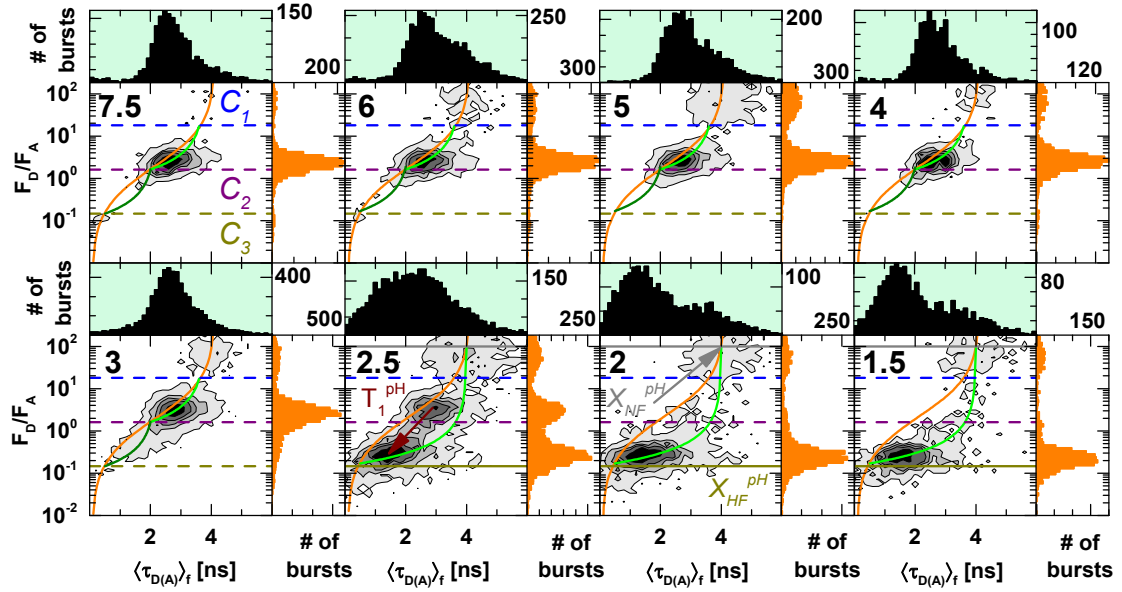
Figure 6.3. **Unfolding of S44pAcPhe/I150C-(DA) variant at acidic pH**. 2D histograms of $F_D/F_A$ vs $\langle\tau_{D(A)}\rangle_f$ with the respective 1D histograms are shown. Horizontal lines indicate the position of states $C_1$ (dashed blue), $C_2$ (dashed purple), $X_{NF}^U$ (grey) and $X_{LF}^U$ or $C_3$ (dark yellow), which show the same lifetime. The wine arrow indicates the transition $T_1^{pH}$ from the $C_1/C_2$ equilibrium population to the mixture of states $X_{NF}^{pH}$ and $X_{HF}^{pH}$. pH is indicated in the corresponding 2D histogram. Orange line represents static linker corrected FRET line, green and olive lines represent dynamic linker corrected FRET lines obtained from the seTCSPC analysis as described in section 3.8.4. $F_D/F_A$ values were corrected by the ratio of green to red detection efficiency $g_G/g_R$, green and red background and crosstalk $\alpha$. All parameters are listed in Appendix Table 9.2

Table 6.1. **Multi exponential fit of donor-only and FRET populations, fluorescence quantum yields of the donor (Alexa488) and acceptor (Alexa647) dyes for S44pAcF/I150C variant at acidic conditions**. The bursts of the donor only or FRET population were selected from the 2D histograms as described in chapter 5 (Fig. 5.S1a,b) and analyzed as described in chapter 3.8.5 of this work.

| Lifetimes of the donor only subpopulation of S44pAcPhe/I150C-(DA) variant at different pH | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Condition | $\tau_1$ [ns] | $x_1$ | $\tau_2$ [ns] | $x_2$ | $\langle\tau_{av}\rangle_x$ [ns] | $\langle\tau_{av}\rangle_f$ [ns] | $\chi_r^2$ | $\Phi_{FD(0)}$ | Method |
| pH 3 | 4.22 | 0.96 | 1.44 | 0.04 | 4.11 | 4.18 | 1.22 | 0.80 | seTCSPC |
| pH 2.5 | 4.21 | 0.95 | 1.61 | 0.05 | 4.08 | 4.16 | 1.19 | 0.80 | seTCSPC |
| pH 2 | 4.24 | 0.95 | 1.44 | 0.05 | 4.11 | 4.19 | 1.13 | 0.80 | seTCSPC |
| pH 1.5 | 4.28 | 0.94 | 1.61 | 0.06 | 4.13 | 4.22 | 1.09 | 0.81 | seTCSPC |

166

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Lifetimes of the FRET subpopulation of S44pAcPhe/I150C-(DA) variant at different pH** | | | | | | | | | | |
| Condition | $\tau_1$ [ns] | $x_1$ | $\tau_2$ [ns] | $x_2$ | $\tau_3$ [ns] | $x_3$ | $\langle\tau_{av}\rangle_x$ [ns] | $\langle\tau_{av}\rangle_f$ [ns] | $\chi_r^2$ | **Method** |
| pH 3 | 3.61 | 0.30 | 1.99 | 0.46 | 0.50 | 0.24 | 2.12 | 2.73 | 1.21 | seTCSPC |
| pH 2.5 | 3.51 | 0.33 | 2.01 | 0.45 | 0.51 | 0.22 | 2.18 | 2.73 | 1.21 | seTCSPC |
| pH 2 | 3.90 | 0.23 | 2.01 | 0.17 | 0.41 | 0.60 | 1.48 | 2.89 | 1.28 | seTCSPC |
| pH 1.5 | 3.91 | 0.34 | — | — | 0.41 | 0.66 | 1.60 | 3.32 | 1.12 | seTCSPC |

## 6.2.3 Discussion of unfolding experiments

Denaturing of the K60pAcF/N132C-(DA) variant by urea shows two excited states: an extended unfolded intermediate state $X_{NF}^U$ which might represent a set of different structures with the interdye distance of >100 Å and an unfolded state $X_{LF}^U$ with the interdye distance of ~ 65 Å. The $X_{NF}^U$ state might partially or fully correspond to a more or less extended polypeptide, because the distance between the two dyes in a 72 amino acid-long fully extended polypeptide is expected to be ~ 260 ± 40 Å depending on the direction of the dye linkers with approximate length of 20 Å each. Mechanistically, this corresponds to dissolving structural elements between the labeling positions 60 and 132 of T4L and forming either a partially or completely extended polypeptide structure. The interdye distance of the state $X_{LF}^U$ is shorter than of state $X_{NF}^U$, which is a clear indication that T4L in the unfolded state $X_{LF}^U$ does not adopt the extended polypeptide structure, but collapses into a molten globule state or into some other unfolded state that is not extended. In contrast, the variant S44pAcPhe/I150C-(DA) at acidic pH introduces a multitude of possible conformations without clear distinction between intermediate and unfolded states. Limiting lifetimes of $\tau_{NF}^{pH} \sim 3.90$ ns and $\tau_{HF}^{pH} \sim 0.4$ ns indicate that the distance distribution ranges from ~25 Å to > 80 Å which includes the whole detectable range for a given dye pair and could be represented structurally by an even broader distance distribution, which most probably corresponds to a slowly interchanging random-coil conformation.

Unfolding of T4L by urea showed two transitions $T_1^U$ and $T_2^U$. The presence of $T_1^U$ supports the theory of an ensemble of intermediate states as it is an equilibrium between the folded states $C_1/C_2$ and intermediate excited state $X_{NF}^U$. Furthermore, this transition occurs earlier than the mid-point of the transitions (5.6 M urea) as was determined by Katherina Hemmen (Hemmen, 2015) for the same T4L variant by CD. In contrast, the transition $T_2^U$ shows the transition between folded states $C_1/C_2$ and unfolded state $X_{LF}^U$ and occurs at the mid-point of transition of the CD measurement. Whether this transition is a three state transition $C\leftrightarrow I\leftrightarrow U$ or a two state $C\leftrightarrow U$ transition cannot be distinguished in this experiment. For the pH titration only

one transition $T_1^{pH}$ is found, from the combination of the folded states $C_1/C_2/C_3$ to the multitude of the unfolded states. Averaging of populations makes it impossible to separate different unfolded states. Nevertheless, these states might include the molten globule state proposed by (Llinas & Marqusee, 1998).

The rate of unfolding induced by urea can be narrowed down to the range of sub-ms to ms by the form and the dynamic averaging behavior of the detected populations in the 2D plots. Kinetics with a broader time range of fluctuations from µs to ms can be attributed to the unfolding by acidic pH as deduced from the form and dynamic averaging of the FRET populations found in the respective 2D plots. This is confirmed by the fFCS analysis performed in collaboration with Prof. Hugo Sanabria (University Clemson, SC, USA) which shows an additional relaxation time in the µs range at acidic pH < 3.

The data show that, denaturation of T4L by protonation or by addition of a chaotropic agent urea follows different unfolding pathways. The intermediate state(s) formed upon denaturation by urea are clearly distinct from the unfolded states, as the unfolded states show shorter fluorescence lifetime than the intermediate states. In the case of pH denaturation, the intermediate steps are indistinguishable from the unfolded states as the unfolded states are represented with a very broad combination of states. In structural terms this suggests that T4L under acidic conditions probably unfolds into a polymer-like extended random-coil structure, while T4L denatured by urea preferably (un)folds into the molten globule state(s) possibly through a fully extended intermediate. This is supported by (Sagle et al, 2009) where two concentration dependent mechanisms of urea action are described: at an intermediate urea concentration, the protein is completely unfolded by urea interaction with the protein backbone (Candotti et al, 2013), whereas for the high urea concentration the unfolding behavior is dominated by the interactions between urea molecules crosslinking the protein into a molten globule state(s).

The kinetic behavior is also different, as the protonation dynamics spans a much broader range of times from µs to ms while the urea denaturation only shows slower kinetics in the range of sub-ms to ms. Moreover, both experiments do not resolve the proposed $I \leftrightarrow U$ transition. During the urea-induced denaturation this transition is probably hidden within the transition $T_2^U$, which is supported by further data gathered by Katherina Hemmen (Hemmen, 2015). In case of acidic denaturation, this transition as well as $C \leftrightarrow I$ transition might be very fast, represented by the µs relaxation time detected by fFCS. These results lead to the conclusion that the unfolding pathway and the mechanism of denaturation by urea and by protonation are significantly

different. To deepen the understanding of the unfolding mechanism and pathway of T4L in urea, this is analyzed in detail by Katherina Hemmen in her upcoming thesis (Hemmen, 2015).

## 6.3 Effect of non-ionic surfactants on unfolding

### 6.3.1 Introduction

smFRET spectroscopy requires only a miniscule amount of ~100 pM of the labeled protein for the measurement which is a significant advantage over other methods which require up to mM concentration of the protein. Unfortunately, there are also two main drawbacks – measuring intermolecular FRET requires high affinity partners and the reporter molecules are easily lost due to adsorption on the surface of the measuring chamber or coverslip. As shown in Figure 6.4a, adsorption has a profound effect on the signal of the FRET species as its concentration decreases with time. To prevent unwanted adsorption of the biomolecules to glass surfaces, non-ionic surfactants are commonly used in single-molecule FRET spectroscopy (Borgia et al, 2012; Wunderlich et al, 2013). Non-ionic surfactants are also used in conjunction with denaturing conditions to study folding/unfolding of proteins or peptides. It is assumed that they do not interact with biomolecules (Eckert et al, 2006; Wahlgren et al, 1997) or change their folding landscape. However, this assumption has not been experimentally tested in combination with denaturing conditions.

**A**        **B**



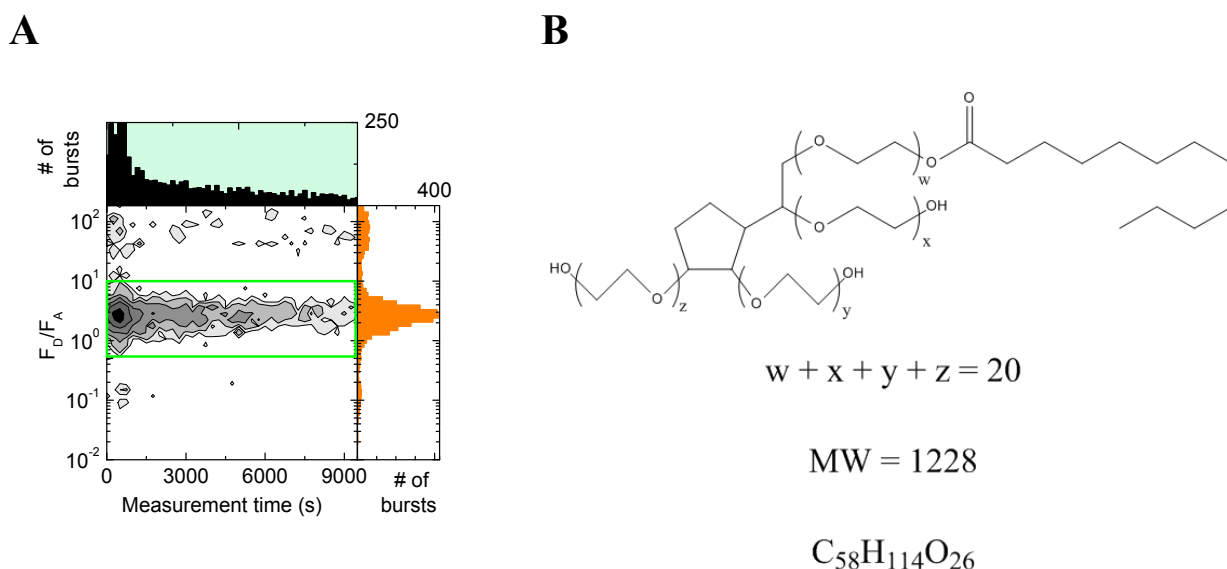Figure 6.4. **Illustration of adsorption of T4L and the structure of Tween 20**. (**A**) Exemplary measurement of S44pAcF/I150C-(DA) without Tween 20. 2D histogram of $F_D/F_A$ vs measurement time with the respective 1D histograms is shown. The green box indicates the FRET species, whose signal strongly decreases with time due to adsorption artefacts. (**B**) Structure, molecular weight and chemical formula of Tween 20.

Tween 20 is a non-ionic surfactant (Fig. 6.4b) often used in screenings (Mcpherson et al, 1986), folding studies (Borgia et al, 2012) and microfluidics assisted experiments (Choi & Chae, 2010; Hofmann et al, 2010; Pfeil et al, 2009). Tween 20 is used at $10^{-2}$ % w/v (the critical micelle concentration (CMC) of Tween 20 is $0.74 \cdot 10^{-2}$ % w/v) by the group of Prof. Ben Schuler for their microfluidic measurements (Borgia et al, 2012).

The following questions were not addressed previously and will be studied in this section:

- How does Tween 20 influence conformational states detected upon unfolding by urea or by acidic pH?

- How does Tween 20 influence the unfolding kinetics?

- What is the influence of Tween 20 at different concentrations on the mechanism of unfolding in the presence of different denaturants?

To address these points, the following unfolding experiments were performed with T4L in presence of Tween 20:

- Unfolding of a T4L variant at different urea concentrations in the absence and presence of $10^{-2}$ % w/v Tween 20.

- Unfolding of a T4L variant at different acidic pH in the absence and presence of $10^{-2}$ % w/v Tween 20

- Measurements at 5.25 M urea and pH 2 for the corresponding variant at different Tween 20 concentrations ranging from $10^{-5}$ to $10^{-2}$ % w/v Tween 20

As described in chapter 6.1, the states which are different from the native states $C_1/C_2/C_3$ will be denoted as $X_{(FRET\ level)}^{(condition)}$ with the following FRET levels assigned as subscript: $HF$ – high FRET (lifetime 0 – 1.2 ns); $MF$ – middle FRET (lifetime 1.2 – 2.5 ns); $LF$ – low FRET (lifetime 2.5 – 3.8 ns) and $NF$ – no FRET (lifetime 3.8 – 4.3 ns). The following conditions will be used $U$ – urea titration, $pH$ – pH titration, $TU$ – urea titration with added Tween 20 and $TpH$ – pH titration with added Tween 20. This notation will be used as superscript to address the specific experiment that was performed.


## 6.3.2 Tween 20 influence on urea unfolding

To test the influence of Tween 20 on the denaturation by urea, unfolding of the K60pAcF/N132C-(DA) variant was carried out in the absence and presence of $10^{-2}$ % w/v Tween

20 at 0, 4.5, 5.25 and 7.5 M urea (Fig. 6.5). In the absence of urea the position and the shape of the main populations are nearly identical. eTCSPC analysis detects two populations $C_1$ (dashed blue line) and $C_2$ (dashed purple line) with two lifetimes of ~2 ns and ~0.4 ns (Table 6.2). At 4.5 M urea there is a clear difference between the samples with and without Tween 20, as the main population in the sample with Tween 20 is shifted toward longer fluorescence lifetimes from ~2 ns to ~3 ns, whereas the sample without Tween 20 shows only slight tailing toward the longer fluorescence lifetime. For the sample without Tween 20, seTCSPC analysis indicates an equilibrium between three states $X_{LF}^U$, $X_{MF}^U$ and $X_{HF}^U$ with lifetimes of ~3.3 ns, ~1.8 ns and ~0.5 ns, respectively (Table 6.2). For the sample in the presence of Tween 20 we detect also at least three populations $X_{LF}^{TU}$, $X_{MF}^{TU}$ and $X_{HF}^{TU}$ with nearly same lifetimes, but the equilibrium is shifted toward the state $X_{LF}^{TU}$. A similar pattern applies for 5.25 M urea – here the equilibrium between the states $X_{LF}^U$ and $X_{MF}^U$ in the absence of Tween 20 is at the mid-point (50/50%), while for the sample with Tween 20 the equilibrium is much closer to the state $X_{LF}^{TU}$. For 7.5 M urea the equilibrium between $X_{LF}^U$ and $X_{MF}^U$ (or $X_{LF}^{TU}$ and $X_{MF}^{TU}$) is strongly shifted towards the $X_{LF}^U$ ($X_{LF}^{TU}$) and is similar for both in presence and absence of Tween 20. Another important
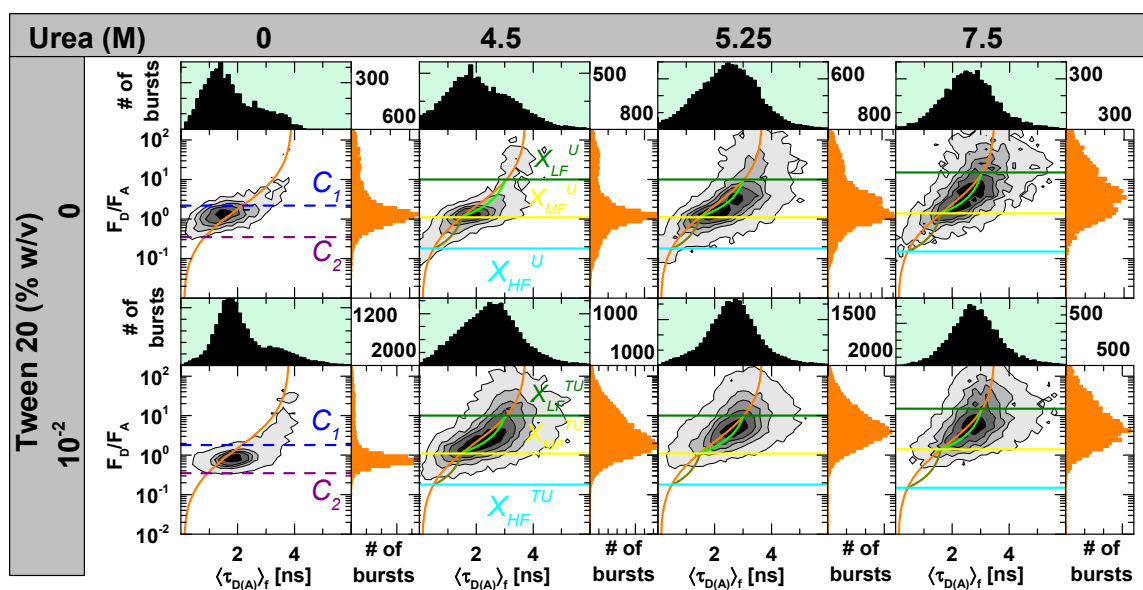


Figure 6.5. **Urea titration series of K60pAcF/N132C-(DA) variant without (upper row) and with $10^{-2}$ % Tween 20 (lower row).** 2D histograms of $F_D/F_A$ vs $\langle\tau_{D(A)}\rangle_f$ with the respective 1D histograms are shown. Horizontal lines indicate the position of states $C_1$ (dashed blue), $C_2$ (dashed purple), $X_{LF}^U$ and $X_{LF}^{TU}$ (olive), $X_{MF}^U$ and $X_{MF}^{TU}$ (yellow) and $X_{HF}^U$ and $X_{HF}^{TU}$ (cyan). Urea concentration is indicated on top of the respective plot, Tween 20 concentration to the left of the graphs. The orange line represents static linker corrected FRET line, green and dark yellow lines represent dynamic linker corrected FRET lines obtained from the seTCSPC analysis as described in section 3.8.4. $F_D/F_A$ values were corrected by the ratio of green to red detection efficiency $g_G/g_R$, green and red background and crosstalk $\alpha$. All parameters are listed in Appendix Table 9.3.

171

Table 6.2. **Multi exponential fit of donor-only and FRET populations, fluorescence quantum yields of the donor (Alexa488) and acceptor (Alexa647) dyes for K60pAcF/N132C-(DA) variant in urea**. The bursts of the donor only or FRET population were selected from the 2D histograms as described in chapter 5 (Fig. 5.S1a,b) and analyzed as described in chapter 3.8.5 of this work. Ensemble TCSPC measurements were performed and analyzed by Katherina Hemmen (Hemmen, 2015).

| Lifetimes of donor only sample of K60pAcF/N132C-(D) variant without Tween 20 at different urea concentrations | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Condition | $\tau_1$ [ns] | $x_1$ | $\tau_2$ [ns] | $x_2$ | $\langle\tau_{av}\rangle_x$ [ns] | $\langle\tau_{av}\rangle_f$ [ns] | $\chi_r^2$ | $\Phi_{FD(0)}$ | Method |
| 0 M urea | 4.08 | 0.89 | 1.78 | 0.11 | 3.83 | 3.96 | 1.11 | 0.74 | eTCSPC |
| 4.5 M urea | 4.02 | 0.90 | 1.35 | 0.10 | 3.74 | 3.92 | 1.05 | 0.73 | eTCSPC |
| 5.25 M urea | 4.06 | 0.85 | 1.64 | 0.15 | 3.70 | 3.90 | 1.17 | 0.72 | eTCSPC |
| 7.5 M urea | 3.98 | 0.82 | 1.47 | 0.18 | 3.53 | 3.79 | 1.14 | 0.69 | eTCSPC |
| Lifetimes of donor only (sub)population of K60pAcF/N132C-(DA) variant $10^{-2}$ % w/v Tween 20 at different urea concentrations | | | | | | | | | |
| Condition | $\tau_1$ [ns] | $x_1$ | $\tau_2$ [ns] | $x_2$ | $\langle\tau_{av}\rangle_x$ [ns] | $\langle\tau_{av}\rangle_f$ [ns] | $\chi_r^2$ | $\Phi_{FD(0)}$ | Method |
| 0 M urea | 4.09 | 0.88 | 1.52 | 0.12 | 3.79 | 3.97 | 1.08 | 0.74 | seTCSPC |
| 4.5 M urea | 4.14 | 0.83 | 1.82 | 0.17 | 3.75 | 3.95 | 1.09 | 0.73 | seTCSPC |
| 5.25 M urea | 4.04 | 0.88 | 1.11 | 0.12 | 3.69 | 3.93 | 1.16 | 0.72 | seTCSPC |
| 7.5 M urea | 3.98 | 0.82 | 1.47 | 0.18 | 3.53 | 3.79 | 1.08 | 0.69 | seTCSPC |

| Lifetimes of FRET (sub)population of K60pAcF/N132C-(DA) variant without Tween 20 at different urea concentrations | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Condition | $\tau_1$ [ns] | $x_1$ | $\tau_2$ [ns] | $x_2$ | $\tau_3$ [ns] | $x_3$ | $\langle\tau_{av}\rangle_x$ [ns] | $\langle\tau_{av}\rangle_f$ [ns] | $\chi_r^2$ | Method |
| 0 M urea | 2.05 | 0.52 | — | — | 0.44 | 0.48 | 1.28 | 1.78 | 1.23 | eTCSPC |
| 4.5 M urea | 3.37 | 0.16 | 1.85 | 0.49 | 0.58 | 0.25 | 1.59 | 2.25 | 1.26 | eTCSPC |
| 5.25 M urea | 3.33 | 0.28 | 1.32 | 0.51 | 0.41 | 0.21 | 1.69 | 2.38 | 1.31 | eTCSPC |
| 7.5 M urea | 3.23 | 0.50 | 1.60 | 0.21 | 0.45 | 0.19 | 2.04 | 2.84 | 1.18 | eTCSPC |
| Lifetimes of FRET (sub)population of K60pAcF/N132C-(DA) variant with $10^{-2}$ % w/v Tween 20 at different urea concentrations | | | | | | | | | |
| Condition | $\tau_1$ [ns] | $x_1$ | $\tau_2$ [ns] | $x_2$ | $\tau_3$ [ns] | $x_3$ | $\langle\tau_{av}\rangle_x$ [ns] | $\langle\tau_{av}\rangle_f$ [ns] | $\chi_r^2$ | Method |
| 0 M urea | 2.08 | 0.47 | — | — | 0.31 | 0.53 | 1.14 | 1.83 | 1.31 | eTCSPC |
| 4.5 M urea | 3.32 | 0.24 | 1.75 | 0.43 | 0.51 | 0.23 | 1.67 | 2.41 | 1.26 | seTCSPC |
| 5.25 M urea | 2.98 | 0.61 | 1.61 | 0.36 | 0.47 | 0.03 | 2.41 | 2.64 | 1.27 | seTCSPC |
| 7.5 M urea | 2.95 | 0.63 | 1.49 | 0.30 | 0.40 | 0.07 | 2.33 | 2.64 | 1.42 | seTCSPC |

difference is the amount of the high FRET states $X_{HF}^U$ and $X_{HF}^{TU}$ – for the sample in absence of Tween 20, broad tailing toward the short lifetime is observed with the corresponding amplitude of $X_{HF}^U$ of ~20%. With Tween 20 this state is strongly disfavored, as the tailing is nearly absent and the amplitude of the state $X_{HF}^{TU}$ is < 10%. In summary, three very similar populations with different slow sub-ms equilibria are detected for all denaturing conditions.

## 6.3.3 Tween 20 influence on unfolding by acidic pH

To test the influence of Tween 20 on the denaturation by acidic pH, the S44pAcF/I150C-(DA) variant was unfolded in the absence or presence of $10^{-2}$ % w/v Tween 20 at pH 3, 2.5, 2 and 1.5 (Fig. 6.6). For the pH titration series in the absence of Tween 20 (Fig. 6.6, upper row) we observed the shift of the main population from the fluorescence lifetime $\langle \tau_{D(A)} \rangle_f$ of 2.8 ns at pH 3 to a shorter fluorescence lifetime $\langle \tau_{D(A)} \rangle_f$ of 1.3 ns at pH 1.5. In all cases the population is located off the static FRET line, indicating a dynamic equilibrium between three native states $C_1$, $C_2$ and $C_3$ for pH 3, whereas for pH 1.5 the equilibrium is between two states $X_{NF}^{pH}$ and $X_{HF}^{pH}$ with lifetimes $\tau_{NF}^{pH} \sim 4$ ns and $\tau_{HF}^{pH} \sim 0.4$ ns, as determined by seTCSPC (Table 6.3). For the pH titration series in the presence of $10^{-2}$ % w/v Tween 20 (Fig. 6.6, lower row), no shift toward shorter lifetime is observed; in contrast, a slight shift of the FRET population to a longer lifetime is visible. At pH 2.5, 2.0 and 1.5 seTCSPC analysis detects three populations: the native-like populations $X_{HF}^{TpH}$ and $X_{MF}^{TpH}$ and a population $X_{NF}^{TpH}$ with a lifetime of ~ 4.0 ns. As the pH becomes more acidic, the amplitude of the population $X_{HF}^{TpH}$ decreases, while that of population $X_{MF}^{TpH}$ increases. In summary, a strong effect of Tween 20 on the structure and dynamics of T4L is observed as we detect different conformations and different averaging behavior of the protein depending on the presence of Tween 20.
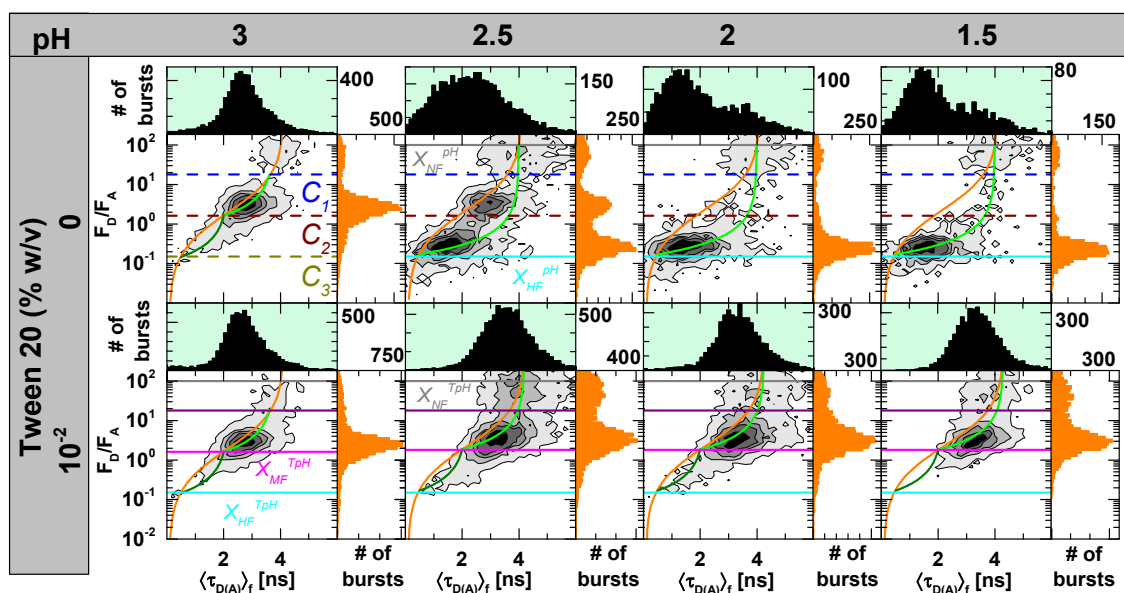
Figure 6.6. **pH titration series of S44pAcPhe/I150C-(DA) variant without (upper row) and with $10^{-2}$ % Tween 20 (lower row)**. 2D histograms of $F_D/F_A$ vs $\langle\tau_{D(A)}\rangle_f$ with the respective 1D histograms are shown. Horizontal lines indicate the positions of the states $C_1$ (dashed blue), $C_2$ (dashed purple), $C_3$ (dashed dark yellow), $X_{NF}^{pH}$ (grey), $X_{MF}^{TpH}$ (magenta) and $X_{HF}^{pH}$ and $X_{HF}^{TpH}$ (cyan), which show a similar lifetime. pH is indicated on top of the corresponding plot, Tween 20 concentration to the left of the graphs. Orange line represents static linker corrected FRET line, green and olive lines represent dynamic linker corrected FRET lines obtained from the seTCSPC analysis as described in section 3.8.4. $F_D/F_A$ values were corrected by the ratio of green to red detection efficiency $g_G/g_R$, green and red background and crosstalk $\alpha$. All parameters are listed in Appendix Table 9.3.

Table 6.3. **Multi exponential fit of donor-only and FRET populations, fluorescence quantum yields of the donor (Alexa488) and acceptor (Alexa647) dyes for S44pAcF/I150C variant at acidic conditions**. The bursts of the donor only or FRET population were selected from the 2D histograms as described in chapter 5 (Fig. 5.S1a,b) and analyzed as described in chapter 3.8.5 of this work.

| Lifetimes of donor only (sub)population of S44pAcPhe/I150C-(DA) variant without Tween 20 at different pH | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Condition | $\tau_1$ [ns] | $x_1$ | $\tau_2$ [ns] | $x_2$ | $\langle\tau_{av}\rangle_x$ [ns] | $\langle\tau_{av}\rangle_f$ [ns] | $\chi_r^2$ | $\Phi_{FD(0)}$ | Method |
| pH 3 | 4.22 | 0.96 | 1.44 | 0.04 | 4.10 | 4.18 | 1.22 | 0.80 | seTCSPC |
| pH 2.5 | 4.21 | 0.95 | 1.61 | 0.05 | 4.08 | 4.16 | 1.19 | 0.80 | seTCSPC |
| pH 2 | 4.24 | 0.95 | 1.44 | 0.05 | 4.11 | 4.19 | 1.13 | 0.8 | seTCSPC |
| pH 1.5 | 4.28 | 0.94 | 1.61 | 0.06 | 4.13 | 4.22 | 1.09 | 0.81 | seTCSPC |

| Lifetimes of donor only (sub)population of S44pAcPhe/I150C-(DA) variant with $10^{-2}$ % w/v Tween 20 at different pH | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Condition | $\tau_1$ [ns] | $x_1$ | $\tau_2$ [ns] | $x_2$ | $\langle\tau_{av}\rangle_x$ [ns] | $\langle\tau_{av}\rangle_f$ [ns] | $\chi_r^2$ | $\Phi_{FD(0)}$ | Method |
| pH 3 | 4.18 | 0.95 | 1.49 | 0.05 | 4.05 | 4.13 | 1.07 | 0.79 | seTCSPC |
| pH 2.5 | 4.26 | 0.96 | 1.50 | 0.04 | 4.14 | 4.22 | 1.11 | 0.81 | seTCSPC |
| pH 2 | 4.31 | 0.96 | 1.55 | 0.04 | 4.20 | 4.27 | 1.17 | 0.82 | seTCSPC |
| pH 1.5 | 4.38 | 0.95 | 1.62 | 0.05 | 4.25 | 4.33 | 1.15 | 0.83 | seTCSPC |

| Lifetimes of FRET (sub)population of S44pAcPhe/I150C-(DA) variant without Tween 20 at different pH | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Condition | $\tau_1$ [ns] | $x_1$ | $\tau_2$ [ns] | $x_2$ | $\tau_3$ [ns] | $x_3$ | $\langle\tau_{av}\rangle_x$ [ns] | $\langle\tau_{av}\rangle_f$ [ns] | $\chi_r^2$ | Method |
| pH 3 | 3.61 | 0.30 | 1.99 | 0.46 | 0.50 | 0.24 | 2.12 | 2.73 | 1.21 | seTCSPC |
| pH 2.5 | 3.51 | 0.13 | 2.01 | 0.28 | 0.51 | 0.59 | 1.32 | 2.19 | 1.21 | seTCSPC |
| pH 2 | 3.90 | 0.09 | 2.01 | 0.09 | 0.41 | 0.82 | 0.87 | 2.15 | 1.28 | seTCSPC |
| pH 1.5 | 4.02 | 0.08 | — | — | 0.51 | 0.92 | 0.79 | 1.94 | 1.12 | seTCSPC |

| Lifetimes of FRET (sub)population of S44pAcPhe/I150C-(DA) variant with $10^{-2}$ % w/v Tween 20 at different pH | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Condition | $\tau_1$ [ns] | $x_1$ | $\tau_2$ [ns] | $x_2$ | $\tau_3$ [ns] | $x_3$ | $\langle\tau_{av}\rangle_x$ [ns] | $\langle\tau_{av}\rangle_f$ [ns] | $\chi_r^2$ | Method |
| pH 3 | 3.48 | 0.34 | 2.04 | 0.43 | 0.47 | 0.23 | 2.17 | 2.75 | 1.23 | seTCSPC |
| pH 2.5 | 3.97 | 0.32 | 2.01 | 0.47 | 0.51 | 0.21 | 2.32 | 3.01 | 1.16 | seTCSPC |
| pH 2 | 3.94 | 0.35 | 2.01 | 0.49 | 0.51 | 0.16 | 2.45 | 3.05 | 1.19 | seTCSPC |
| pH 1.5 | 3.98 | 0.29 | 2.01 | 0.51 | 0.51 | 0.10 | 2.23 | 3.00 | 1.17 | seTCSPC |

## 6.3.4 Unfolding of T4L at different Tween 20 concentrations in the presence of 5.25 M urea

For the urea experiments, the K60pAcF/N132C-(DA) variant in buffer containing 5.25 M urea at different Tween 20 concentration ranging from 0 to $10^{-2}$ % w/v (above CMC) was used (Fig. 6.7, upper panel). As described in section 6.3.2 in the absence of Tween 20, this T4L variant shows two distinct FRET populations at $\langle\tau_{D(A)}\rangle_f$ of ~2 ns and ~2.8 ns and a broad population towards the very short lifetime (Fig. 6.7). seTCSPC analysis showed that these populations are in a dynamic exchange between ~3.3 ns ($X_{LF}^{TU}$) and ~1.5 ns ($X_{MF}^{TU}$) and between ~0.5 ns ($X_{HF}^{TU}$) and ~1.5 ns ($X_{MF}^{TU}$), respectively (Table 6.4). Addition of $10^{-5}$ % w/v of Tween 20 shifts the equilibrium away from $X_{HF}^{TU}$; its fraction decreases to <10% in seTCSPC analysis and no broad population is detectable (Table 6.4). Furthermore, the equilibrium between $X_{LF}^{TU}$ and $X_{MF}^{TU}$ shifts strongly towards the population $X_{LF}^{TU}$ (Table 6.4). This equilibrium shift
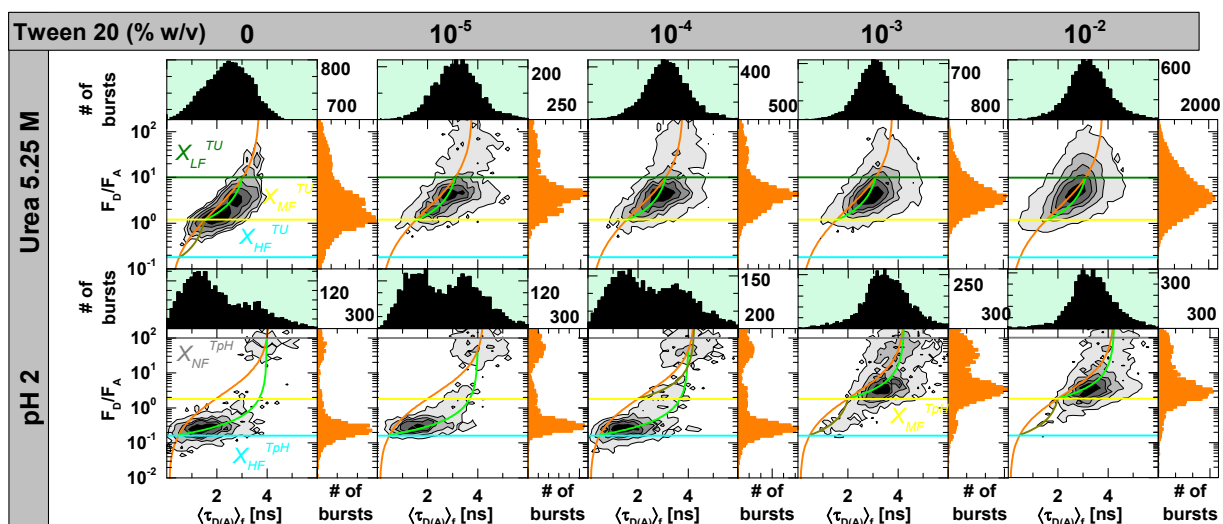
Figure 6.7. **Urea titration series of K60pAcPhe/N132C-(DA) mutant (upper row) and pH titration of S44pAcPhe/I150C-(DA) mutant (lower row) at different Tween20 concentrations.** 2D histograms of $F_D/F_A$ vs $\langle\tau_{D(A)}\rangle_f$ with the respective 1D histograms are shown. Horizontal lines indicate the position of the states $X_{LF}^{TU}$ (green), $X_{MF}^{TU}$ and $X_{MF}^{TpH}$ (yellow), $X_{HF}^{TU}$ and $X_{HF}^{TpH}$ (cyan) and $X_{NF}^{TpH}$ (grey). Tween 20 concentration is indicated on top, type of denaturant indicated to the left of the graph. The orange line represents static linker corrected FRET line, green and olive lines represent dynamic linker corrected FRET lines obtained from the seTCSPC analysis as described in section 3.8.4. $F_D/F_A$ values were corrected by the ratio of green to red detection efficiency $g_G/g_R$, green and red background and crosstalk $\alpha$. All parameters are listed in Appendix Table 9.4.

Table 6.4. **Multi exponential fit of donor-only and FRET populations, fluorescence quantum yields of the donor (Alexa488) and acceptor (Alexa647) dyes for K60pAcF/N132C-(DA) variant in urea and S44pAcPhe/I150C-(DA) variant in pH2 at different Tween 20 concentrations**. The bursts of the donor only or FRET population were selected from the 2D histograms as described chapter 5 (Fig. 5.S1a,b) and analyzed as described in chapter 3.8.5 of this work. Ensemble TCSPC measurements were performed and analyzed by Katherina Hemmen (Hemmen, 2015).

| Lifetimes of donor only (sub)population of K60pAcF/N132C-(DA) variant at 5.25 M urea at different Tween 20 concentrations | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Condition | $\tau_1$ [ns] | $x_1$ | $\tau_2$ [ns] | $x_2$ | $\langle\tau_{av}\rangle_x$ [ns] | $\langle\tau_{av}\rangle_f$ [ns] | $\chi_r^2$ | $\Phi_{FD(0)}$ | Method |
| 0 % w/v | 4.06 | 0.85 | 1.64 | 0.15 | 3.70 | 3.90 | 1.17 | 0.72 | eTCSPC |
| $10^{-5}$ % w/v | 4.03 | 0.93 | 1.31 | 0.07 | 3.85 | 3.97 | 1.08 | 0.74 | seTCSPC |
| $10^{-4}$ % w/v | 4.13 | 0.88 | 1.41 | 0.12 | 3.80 | 4.01 | 1.10 | 0.77 | seTCSPC |
| $10^{-3}$ % w/v | 4.01 | 0.89 | 1.75 | 0.11 | 3.76 | 3.89 | 1.10 | 0.74 | seTCSPC |
| $10^{-2}$ % w/v | 4.04 | 0.88 | 1.11 | 0.12 | 3.69 | 3.93 | 1.16 | 0.72 | seTCSPC |

176

**Lifetimes of donor only (sub)population S44pAcF/I150C-(DA) variant at pH 2 at different Tween 20 concentrations**

| Condition | $\tau_1$ [ns] | $x_1$ | $\tau_2$ [ns] | $x_2$ | $\langle\tau_{av}\rangle_x$ [ns] | $\langle\tau_{av}\rangle_f$ [ns] | $\chi_r^2$ | $\Phi_{FD(0)}$ | Method |
|---|---|---|---|---|---|---|---|---|---|
| 0 % w/v | 4.24 | 0.95 | 1.44 | 0.05 | 4.11 | 4.19 | 1.13 | 0.8 | seTCSPC |
| $10^{-5}$ % w/v | 4.31 | 0.96 | 1.67 | 0.04 | 4.20 | 4.27 | 1.10 | 0.82 | seTCSPC |
| $10^{-4}$ % w/v | 4.35 | 0.95 | 1.52 | 0.05 | 4.22 | 4.30 | 1.08 | 0.82 | seTCSPC |
| $10^{-3}$ % w/v | 4.41 | 0.93 | 1.51 | 0.07 | 4.21 | 4.34 | 1.15 | 0.82 | seTCSPC |
| $10^{-2}$ % w/v | 4.31 | 0.96 | 1.55 | 0.04 | 4.20 | 4.27 | 1.17 | 0.82 | seTCSPC |

**Lifetimes of FRET (sub)population of K60pAcF/N132C-(DA) variant at 5.25 M urea at different Tween 20 concentrations**

| Condition | $\tau_1$ [ns] | $x_1$ | $\tau_2$ [ns] | $x_2$ | $\tau_3$ [ns] | $x_3$ | $\langle\tau_{av}\rangle_x$ [ns] | $\langle\tau_{av}\rangle_f$ [ns] | $\chi_r^2$ | Method |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 % w/v | 3.33 | 0.28 | 1.32 | 0.51 | 0.41 | 0.21 | 1.69 | 2.38 | 1.31 | eTCSPC |
| $10^{-5}$ % w/v | 3.08 | 0.47 | 1.42 | 0.44 | 0.44 | 0.09 | 2.11 | 2.54 | 1.35 | seTCSPC |
| $10^{-4}$ % w/v | 3.16 | 0.51 | 1.58 | 0.42 | 0.48 | 0.07 | 2.31 | 2.67 | 1.37 | seTCSPC |
| $10^{-3}$ % w/v | 3.12 | 0.55 | 1.61 | 0.39 | 0.51 | 0.06 | 2.37 | 2.69 | 1.29 | seTCSPC |
| $10^{-2}$ % w/v | 2.98 | 0.61 | 1.61 | 0.36 | 0.47 | 0.03 | 2.41 | 2.64 | 1.27 | seTCSPC |

**Lifetimes of FRET (sub)population of S44pAcF/I150C-(DA) variant at pH 2 at different Tween 20 concentrations**

| Condition | $\tau_1$ [ns] | $x_1$ | $\tau_2$ [ns] | $x_2$ | $\tau_3$ [ns] | $x_3$ | $\langle\tau_{av}\rangle_x$ [ns] | $\langle\tau_{av}\rangle_f$ [ns] | $\chi_r^2$ | Method |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 % w/v | 3.90 | 0.09 | 2.01 | 0.09 | 0.41 | 0.82 | 0.87 | 2.15 | 1.28 | seTCSPC |
| $10^{-5}$ % w/v | 3.95 | 0.13 | 2.08 | 0.12 | 0.40 | 0.73 | 1.06 | 2.53 | 1.27 | seTCSPC |
| $10^{-4}$ % w/v | 3.92 | 0.15 | 2.01 | 0.16 | 0.32 | 0.69 | 1.13 | 2.67 | 1.31 | seTCSPC |
| $10^{-3}$ % w/v | 3.95 | 0.27 | 2.12 | 0.48 | 0.61 | 0.25 | 2.24 | 2.89 | 1.22 | seTCSPC |
| $10^{-2}$ % w/v | 3.94 | 0.35 | 2.01 | 0.49 | 0.51 | 0.16 | 2.45 | 3.05 | 1.19 | seTCSPC |

is amplified by increasing the Tween 20 concentrations with the fraction of $X_{LF}^{TU}$ reaching a maximum of 60 % at $10^{-2}$ % w/v Tween 20. In summary, even very low concentrations of $10^{-5}$ % w/v of Tween 20 induce a shift of equilibrium from $X_{HF}^{TU}$ to $X_{LF}^{TU}$.

## 6.3.5 Unfolding of T4L at different Tween 20 concentrations at pH 2

For the pH experiments, the S44pAcPhe/I150C-(DA) variant in pH 2 buffer at different Tween 20 concentrations ranging from 0 to $10^{-2}$ % w/v was studied (Fig. 6.7, lower panel). In the absence of Tween 20 a single FRET population at $\langle\tau_{D(A)}\rangle_f$ of 1.4 ns is visible. The form of this

177

population clearly indicates dynamics, as it is nearly horizontal and does not follow the expected line for the static populations (orange line). seTCSPC analysis is dominated by two populations $X_{NF}^{TpH}$ and $X_{HF}^{TpH}$ with lifetimes $\tau_{NF}^{TpH} \sim 3.9$ ns and $\tau_{HF}^{TpH} \sim 0.4$ ns with a third detectable population $X_{MF}^{TpH}$ with the lifetime $\tau_{MF}^{TpH} \sim 2.0$ ns. Addition of $10^{-5}$ % w/v of Tween 20 shifts the equilibrium toward the populations $X_{NF}^{TpH}$ and $X_{MF}^{TpH}$. Upon increasing the concentration of Tween 20 the population fraction of $X_{MF}^{TpH}$ increases up to nearly 50%, while the fraction of $X_{HF}^{TpH}$ is strongly reduced with the $x_{HF}^{TpH} \sim 15\%$.


## 6.3.6 Discussion of the effects of the non-ionic surfactant Tween 20 on urea- and pH-induced unfolding

As shown in this study, addition of the non-ionic detergent Tween 20 has a significant influence on the structure of T4L at denaturing conditions. The exact effect is dependent on the type of induced unfolding (urea/protonation), making it difficult to predict the consequences for an unknown sample and denaturation method.

Denaturation by urea in the presence of Tween 20 shifts the equilibrium toward the low FRET state of $X_{LF}^{TU}$ decreasing the probability of the state $X_{HF}^{TU}$ and therefore lowering its detectability. Tween 20 promotes the shift of the mid-point of unfolding toward lower urea concentrations by preferentially stabilizing the protein in the state $X_{LF}^{TU}$ (Fig. 6.5). Thus, surfactant and denaturant show cooperative behavior for unfolding of T4L: while urea shifts the equilibrium toward the intermediate/unfolded states by dissolving the structural elements of the protein and binding to the exposed protein backbone (Candotti et al, 2013), Tween 20 stabilizes the protein in the unfolded state by binding to it and shifting the equilibrium further toward the state $X_{LF}^{TU}$.

At acidic pH in the absence of Tween 20, T4L is protonated and unfolds into a molten globule non-native state (Llinas & Marqusee, 1998). In the presence of $10^{-2}$ % w/v Tween 20 T4L shows a different behavior whereby the protein is either stabilized by non-covalent binding of Tween 20 in the native-like states $X_{MF}^{TpH}$ and $X_{HF}^{TpH}$ or unfolds into the $X_{NF}^{TpH}$ state. While there are clearly similarities in the lifetimes of the states $X_{HF}^{pH}$ and $X_{HF}^{TpH}$ as well as between the states $X_{NF}^{pH}$ and $X_{NF}^{TpH}$, the corresponding amplitudes indicate that their origin is different, as in the absence of Tween 20 the $X_{HF}^{pH}$ state is strongly favored while in the presence of Tween 20 the $X_{HF}^{TpH}$ state is strongly disfavored. Furthermore, the states $X_{NF}^{pH}$ and $X_{NF}^{TpH}$ themselves might correspond to an ensemble of diverse states, making the comparison ambiguous.

In contrast to the equilibrium between the populations where the surfactant induces new conformations, the dynamics of unfolding does not show new kinetic terms as there is no significant change in the averaging behavior or the distribution of FRET populations in the 2D histograms. Their form indicates that the unfolding kinetics of this variant are in the sub-ms to ms range. Nevertheless, the shift in equilibrium indicates that the kinetic terms are changed by the addition of Tween 20. In contrast, the averaging behavior is significantly different for the acidic pH denaturation, with a very broad horizontal FRET population in absence of Tween 20 and more concentrated diagonal FRET population in presence of Tween 20. Moreover, fFCS analysis performed in cooperation with Prof. Hugo Sanabria (University Clemson, SC, USA) indicates an additional kinetic term in the μs range for the titration in absence of Tween 20, while this term disappears in the comparable experiments in the presence of Tween 20.

Titration of Tween 20 clearly demonstrates the variability of the surfactant effect on the unfolding by different denaturants. The effective concentrations of Tween 20 is between $10^{-3}$ and $10^{-4}$ % w/v in acidic conditions, whereas for the urea denaturation the effect on unfolding is detectable at concentrations as low as $10^{-5}$ % w/v. Furthermore, both effective concentrations are lower than CMC, indicating that the influence of Tween 20 is not due to micelle/protein interaction but is probably attributed to single Tween 20 molecules interacting with protein.

The effect of Tween 20 probably involves non-covalent and non-specific binding to the hydrophobic core of the T4L areas exposed during unfolding. Otherwise, an effect on the T4L would be already visible either in our measurement or previous studies (Eckert et al, 2006; Wahlgren et al, 1997) who did not detect any effect on protein structure for other non-ionic surfactants at native conditions. This non-covalent binding may lead to the stabilization of the protein in the preferred excited state, which is probably selected by the properties of the exposed residues. The overall connectivity of the states is strongly influenced by Tween 20 in the presence of either denaturant, especially for the high FRET states $X_{HF}^{TU}$ and $X_{HF}^{TpH}$. Both of these states are strongly disfavored in presence of Tween 20. It is possible, that the non-covalent binding of Tween 20 also sterically hinders the formation of very compact molten globule states, while allowing the protein to adopt other less compact conformations (urea) or to transition into a multitude of polymer-like random coil structures (pH). To further elucidate the mechanism of Tween 20 influencing the unfolding of T4L, additional experiments in cooperation with Hugo Sanabria will be performed and are not a part of this dissertation. Furthermore, the influence of other detergents that could point to specific interactions is not the objective of this study.

The strong overall effects on kinetics and structure of the examined T4L variants make Tween 20 a very undesirable addition to the unfolding experiments. To prevent the adsorption of labeled T4L to the walls of the measuring chamber we proceeded to use an excess of unlabeled T4L ($\sim$1 µM). Nevertheless, for other proteins this method could still lead to the change of the behavior of the labeled molecules in the measurement (Gibb et al, 2014; Graham et al, 2011) and has to be tested for each new case.

# 7. Conclusions and Outlook

The present work provides new insights into the dynamics of proteins, with the two proteins studied in this work, T4L and IF3 serving as examples for different dynamic regimes. Comparing the behavior of these two proteins, I was able to show that the structure of the linker has a significant influence on the conformational properties of the protein. While IF3 can be considered as an example of a protein that fluctuates freely within a quasi-continuum of different domain arrangements and adopts many different conformations with at least eight distinguishable states, T4L is restricted to only three fairly rigid structures. Moreover, kinetic behavior of these proteins is very different: while the dynamics of IF3 can be very rapid (in the nanoseconds range), T4L shows only slower motions in the microsecond range which are associated with transitions between certain distinct structures. These results show the power of combined smFRET approach, as the combination of (s)eTCSPC, (f)FCS and MFD provide simultaneously time-resolved information on protein structure and dynamics.

## 7.1 T4 Lysozyme

For T4 lysozyme, three distinct ground-state conformation, $C_1$, $C_2$, and $C_3$ were identified. Two distinct structures from the PDB database were assigned as states that likely correspond to the $C_1$ and $C_2$ structures. In contrast, state $C_3$ is a new state that has not been previously identified. Additionally, the complete kinetic scheme of the protein was described for the functional cycle of the protein. The novel compact state $C_3$ turned out to be involved in the product release as expected from the Michaelis-Menten model for non-ATP/GTP driven proteins, which appears to be the first time that such intermediate is experimentally shown for a protein. Further work is now performed using computational methods to determine the structure of $C_3$. Here, the combination of molecular dynamic simulations and structural modeling combined with the distance information derived from 24 T4L mutants will allow the structural identification of the previously hidden state and will validate the notion of the $C_3$ acting as the product release state.

## 7.2 Initiation factor 3

For the initiation factor 3 structural dynamics of the protein free in solution and on the ribosome at different stages of translation initiation were studied. Eight distinct interconverting states of the IF3 in solution were identified, answering some long-standing question concerning

the structure of the IF3 linker. Using newest developed simulation tools for the first time, it was shown that three relaxation rates were required to describe the protein dynamics in solution. This dynamics changes when IF3 binds to the ribosome dependent on the functional state of the complex and the translation initiation region of the mRNAs. Additionally, IF3 on the ribosome may adopt different arrangements, with the C-terminal domain docking or undocking from the 30S subunit.

Despite the progress in measuring IF3 dynamics, a number of questions remain. The discrepancy between our studies and the results of the Gonzalez group are clearly striking and were a subject of significant discussion by the referees of the manuscript. One potential way to understand the source of differences would be to use the exact functional complexes and buffer conditions used by that group. Furthermore, it remains to be shown how the dynamics of IF3 contribute to the fidelity of the start site selection and subunit joining. The position of IF3 on the 30S subunit and the potential link between the movements of 30S subunit and of IF3 could not be elucidated by the current experimental setup. Here, the labeling of the ribosomal proteins of the 30S subunit and the reconstitution of the ribosome as described by Noller (Culver & Noller, 1999; Green & Noller, 1999) would be required to answer these outstanding questions in the ribosome field. The labeling of other parts of the initiation machinery (50S subunit, tRNA, initiation factors) will allow for comprehensive mapping of the interactions and description of the kinetic scheme for the ribosome initiation. This future work cannot be performed by a single group and requires collaborative efforts of specialists in single-molecule detection, biochemistry and genetics. However, the expected overall impact on the field would be significant and could resolve many long-standing questions in translation.

## 7.3 Unfolding of T4 Lysozyme and surfactant effects on the denaturation

Different pathways of the protein unfolding under acidic pH and urea were described in this work. This qualitative description of the unfolding of the T4 lysozyme served as a starting point for a more quantitative work, as already performed by Katherina Hemmen who is exploring the unfolding landscape and the multitude of unfolded states of T4L in different urea conditions. This work in progress is especially challenging owing to limitations of FRET in the range of the detected distances. Additionally, due to three existing structures for the folded T4L, the identification of the intermediate or unfolded states is challenging and requires a significant amount of biochemical and structural work. For the unfolding by acidic pH, additional

experiments are required to understand the mechanism of the unfolding and the underlying structural changes. Single molecule and FCS experiments on other selected variants of T4L with different labeling patterns would provide additional information on the intermediate and unfolded states and their dynamics.

Effect of the non-ionic surfactant Tween 20 on the T4L is another important part of this work. Tween 20 strongly influences T4L conformations and dynamics and should not be used without controls for biochemical procedures including reduction of the adsorption of proteins on the walls of measuring equipment (well chambers or microfluidic devices). For the denaturing conditions of acidic pH or high urea, the detergent impacts the protein unfolding behavior significantly and should be considered by anyone using Tween 20 or any other non-ionic surfactant in their experiments in denaturing conditions. It is important to note that even such a low concentrations as $10^{-5}$% still may impact the system significantly, especially under urea denaturation. The mechanism by which Tween 20 influences the energetic landscape of T4L unfolding is not yet clear and will be further examined in collaboration with Prof. Hugo Sanabria. Here, additional structural simulations of the atomistic interaction between Tween 20 and T4L are needed to identify potential binding sites of Tween 20 on T4L and might allow to fully deduce the mechanism of the interaction.

## 7.4 Developing comprehensive framework for generating and analysis of the smFRET experiments

Simulation of multiple conformation states (e.g. eight for IF3) using previous kinetic and structural knowledge in combination with simultaneous analysis comprises a new step in the analysis of fluorescence data. Here, additional work has to be done towards improving the work-flow. In the future, data on distances ((s)eTCSPC), kinetics ((f)FCS), and MFD histograms should be validated simultaneously. At least a basic routine should be implemented to vary parameters and reduce the amount of manual "try and error" for the user. Providing such a tools would be of great use for many fields where FRET information is obtained and might be also helpful to other experimental methods which are able to obtain structural and kinetic data from a single measurement. Improving the analysis routines would make the multiparameter single molecule FRET approach an outstanding tool to explore structural and dynamical landscape of proteins and macromolecular complexes, which is one of the most important questions of modern biochemistry.

184

My second supervisor **Prof. Dr. Holger Gohlke**.

All my **colleagues** for the great working atmosphere and fun time together.

I want to thank even more my family and especially my mother **Prof. Dr. Marina Rodnina-Wintermeyer**, for teaching me, supporting me and helping me through all the easy and hard times; for the seeding and finishing the IF3 project.

Finally, I want to thank my wife **Daria Rodnina** who always motivated me and helped me with this work. My words are not enough, I love you!

# 9. Appendix

Table 9.1. **Static FRET lines of K60pAcF/N132C-(DA) variant of T4L as shown in the Figure 6.2.** The measurements were performed as indicated in the first column by Dmytro Rodnin (DR), Katherina Hemmen (KH) or Master Student Soheila Rezaei Adariani (SRA) under supervision of Dmytro Rodnin.

| Static FRET lines for the given urea concentrations | | |
|---|---|---|
| $g_G/g_R = 0.7 – 0.8$ | $B_G$ and $B_R = 0.4 – 2.0$ kHz | $\alpha = 1.7\text{-}2.0\%$ |
| DR | 0 M urea | $F_D/F_A=(0.8067/0.3900)/((4.1617/((-0.0363*x\textasciicircum3)+(0.2647*x\textasciicircum2)+0.5377*x-0.0499))-1)$ |
| DR | 2 M urea | $F_D/F_A=(0.7877/0.3700)/((4.0545/((-0.0381*x\textasciicircum3)+(0.2728*x\textasciicircum2)+0.5318*x-0.0486))-1)$ |
| SRA | 3 M urea | $F_D/F_A=(0.7806/0.3700)/((4.0123/((-0.0391*x\textasciicircum3)+(0.2770*x\textasciicircum2)+0.5284*x-0.0480))-1)$ |
| DR | 4 M urea | $F_D/F_A=(0.7698/0.3900)/((3.9612/((-0.0401*x\textasciicircum3)+(0.2805*x\textasciicircum2)+0.5288*x-0.0478))-1)$ |
| DR | 4.25 M urea | $F_D/F_A=(0.7698/0.3900)/((3.9612/((-0.0401*x\textasciicircum3)+(0.2805*x\textasciicircum2)+0.5288*x-0.0478))-1)$ |
| SRA | 4.5 M urea | $F_D/F_A=(0.7630/0.3700)/((3.9221/((-0.0410*x\textasciicircum3)+(0.2845*x\textasciicircum2)+0.5260*x-0.0474))-1)$ |
| KHM | 4.75 M urea | $F_D/F_A=(0.7630/0.3700)/((3.9221/((-0.0410*x\textasciicircum3)+(0.2845*x\textasciicircum2)+0.5260*x-0.0474))-1)$ |
| SRA | 5 M urea | $F_D/F_A=(0.7578/0.3700)/((3.9074/((-0.0412*x\textasciicircum3)+(0.2840*x\textasciicircum2)+0.5299*x-0.0477))-1)$ |
| DR | 5.25 M urea | $F_D/F_A=(0.7578/0.3700)/((3.9074/((-0.0412*x\textasciicircum3)+(0.2840*x\textasciicircum2)+0.5299*x-0.0477))-1)$ |
| SRA | 5.5 M urea | $F_D/F_A=(0.7596/0.3700)/((3.8988/((-0.0412*x\textasciicircum3)+(0.2857*x\textasciicircum2)+0.5235*x-0.0468))-1)$ |
| KHM | 5.75 M urea | $F_D/F_A=(0.7596/0.3700)/((3.8988/((-0.0412*x\textasciicircum3)+(0.2857*x\textasciicircum2)+0.5235*x-0.0468))-1)$ |
| SRA | 6 M urea | $F_D/F_A=(0.7614/0.3700)/((3.8957/((-0.0411*x\textasciicircum3)+(0.2864*x\textasciicircum2)+0.5194*x-0.0463))-1)$ |
| SRA | 6.5 M urea | $F_D/F_A=(0.7575/0.3700)/((3.8760/((-0.0415*x\textasciicircum3)+(0.2878*x\textasciicircum2)+0.5187*x-0.0461))-1)$ |
| SRA | 7 M urea | $F_D/F_A=(0.7553/0.3800)/((3.8645/((-0.0417*x\textasciicircum3)+(0.2888*x\textasciicircum2)+0.5183*x-0.0460))-1)$ |
| KHM | 7.5 M urea | $F_D/F_A=(0.7512/0.3700)/((3.8493/((-0.0420*x\textasciicircum3)+(0.2894*x\textasciicircum2)+0.5197*x-0.0460))-1)$ |
| SRA | 8 M urea | $F_D/F_A=(0.7480/0.4000)/((3.7400/((-0.0436*x\textasciicircum3)+(0.3035*x\textasciicircum2)+0.4850*x-0.0412))-1)$ |

Table 9.2. **Static and dynamic FRET lines for S44pAcF/I150C variant at acidic conditions as shown in the Figure 6.3 and 6.6**.

| FRET lines of S44pAcPhe/I150C-(DA) variant at acidic pH titration | | |
|---|---|---|
| $g_G/g_R = 0.7 – 0.8$ | $B_G$ and $B_R = 0.2 – 1.1$ kHz | $\alpha = 1.7\%$ |
| pH 1.5 0 % w/v Tween 20 | Color | Equation |
| Static | Orange | $F_D/F_A=(0.8200/0.3200)/((4.1000/((-0.0356*x\textasciicircum3)+(0.2711*x\textasciicircum2)+0.4964*x-0.0440))-1)$ |
| Dynamic $\tau_1 = 4.0$ ns, $\tau_2 = 0.5$ ns | Green | $F_D/F_A=0.2000/(0.3200*((1/4.0000+1/0.5000-(2.0951*x-4.3808)/(0.5000*4.0000))-1/4.1000))$ |
| pH 2.0 0 % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A=(0.8200/0.3200)/((4.1000/((-0.0356*x\textasciicircum3)+(0.2711*x\textasciicircum2)+0.4964*x-0.0440))-1)$ |
| Dynamic | Green | $F_D/F_A=0.2000/(0.3200*((1/4.0000+1/0.5000-(2.0951*x-$ |

| | | |
|---|---|---|
| $\tau_1$ = 4.0 ns, $\tau_2$ = 0.5 ns | | 4.3808)/(0.5000*4.0000))-1/4.1000)) |
| pH 2.5 0 % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.8200/0.3200)/((4.1000/((-0.0356*x^3)+(0.2711*x^2)+0.4964*x-0.0440))-1) |
| Dynamic<br>$\tau_1$ = 4.0 ns, $\tau_2$ = 0.5 ns | Green | $F_D/F_A$=0.2000/(0.3200*((1/4.0000+1/0.5000-(2.0951*x-4.3808)/(0.5000*4.0000))-1/4.1000)) |
| pH 3.0 0 % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.8200/0.3700)/((4.1000/((-0.0356*x^3)+(0.2711*x^2)+0.4964*x-0.0440))-1) |
| Dynamic<br>$\tau_1$ = 3.5 ns, $\tau_2$ = 2.0 ns | Green | $F_D/F_A$=0.2000/(0.3700*((1/3.5000+1/2.0000-(1.3216*x-1.1429)/(2.0000*3.5000))-1/4.2000)) |
| Dynamic<br>$\tau_1$ = 2.0 ns, $\tau_2$ = 0.5 ns | Olive | $F_D/F_A$=0.2000/(0.3700*((1/2.0000+1/0.50002.2432*x-2.5578)/(0.5000*2.0000))-1/4.2000)) |
| pH 1.5 $10^{-2}$ % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.8600/0.3200)/((4.3000/((-0.0320*x^3)+(0.2557*x^2)+0.5023*x-0.0454))-1) |
| Dynamic<br>$\tau_1$ = 4.3 ns, $\tau_2$ = 2.0 ns | Green | $F_D/F_A$=0.2000/(0.3200*((1/4.3000+1/2.0000-((0.0000*x^3)+(-0.0000*x^2)+1.2806*x-1.2063)/(2.0000*4.3000))-1/4.3000)) |
| Dynamic<br>$\tau_1$ = 2.0 ns, $\tau_2$ = 0.5 ns | Olive | $F_D/F_A$=0.2000/(0.3200*((1/2.0000+1/0.5000-(2.2536*x-2.5822)/(0.5000*2.0000))-1/4.3000)) |
| pH 2.0 $10^{-2}$ % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.8500/0.3200)/((4.2500/((-0.0328*x^3)+(0.2594*x^2)+0.5009*x-0.0451))-1) |
| Dynamic<br>$\tau_1$ = 4.25 ns, $\tau_2$ = 2.0ns | Green | $F_D/F_A$=0.2000/(0.3200*((1/4.2500+1/2.0000-(1.2766*x-1.1755)/(2.0000*4.2500))-1/4.2500)) |
| Dynamic<br>$\tau_1$ = 2.0 ns, $\tau_2$ = 0.5 ns | Olive | $F_D/F_A$=0.2000/(0.3200*((1/2.0000+1/0.5000-(2.2485*x-2.5702)/(0.5000*2.0000))-1/4.2500)) |
| pH 2.5 $10^{-2}$ % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.8400/0.3200)/((4.2000/((-0.0337*x^3)+(0.2632*x^2)+0.4994*x-0.0447))-1) |
| Dynamic<br>$\tau_1$ = 4.2 ns, $\tau_2$ = 2.0 ns | Green | $F_D/F_A$=0.2000/(0.3200*((1/4.2000+1/2.0000-(+1.2726*x-1.1448)/(2.0000*4.2000))-1/4.2000)) |
| Dynamic<br>$\tau_1$ = 2.0 ns, $\tau_2$ = 0.5 ns | Olive | $F_D/F_A$=0.2000/(0.3200*((1/2.0000+1/0.5000-(2.2432*x-2.5578)/(0.5000*2.0000))-1/4.2000)) |
| pH 3.0 $10^{-2}$ % w/v Tween 20 | | |

| | | |
|---|---|---|
| Static | Orange | $F_D/F_A=(0.8200/0.3700)/((4.1000/((-0.0356*x^3)+(0.2711*x^2)+0.4964*x-0.0440))-1)$ |
| Dynamic $\tau_1 = 3.5$ ns, $\tau_2 = 2.0$ ns | Green | $F_D/F_A=0.2000/(0.3700*((1/3.5000+1/2.0000-(1.3216*x-1.1429)/(2.0000*3.5000))-1/4.2000))$ |
| Dynamic $\tau_1 = 2.0$ ns, $\tau_2 = 0.5$ ns | Olive | $F_D/F_A=0.2000/(0.3700*((1/2.0000+1/0.5000-(2.2432*x-2.5578)/(0.5000*2.0000))-1/4.2000))$ |

Table 9.3. **Static and dynamic FRET lines for K60pAcF/N132C-(DA) variant at different urea concentrations as shown in the Figure 6.5**.

| FRET lines for K60pAcF/N132C-(DA) variant at different urea concentrations | | |
|---|---|---|
| $g_G/g_R = 0.77$ | $B_G$ and $B_R = 0.4 - 1.9$ kHz | $\alpha = 1.7\%$ |
| 0 M urea 0 % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A=(0.7630/0.3700)/((3.9221/((-0.0410*x^3)+(0.2845*x^2)+0.5260*x-0.0474))-1)$ |
| 4.5 M urea 0 % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A=(0.7500/0.3200)/((3.7500/((-0.0433*x^3)+(0.3025*x^2)+0.4853*x-0.0413))-1)$ |
| Dynamic $\tau_1 = 3.0$ ns, $\tau_2 = 1.5$ ns | Green | $F_D/F_A=0.2000/(0.3200*((1/1.5000+1/3.0000-(1.4180*x-1.2719)/(3.0000*1.5000))-1/3.7000))$ |
| Dynamic $\tau_1 = 1.5$ ns, $\tau_2 = 0.5$ ns | Olive | $F_D/F_A=0.2000/(0.3200*((1/1.5000+1/0.5000-(2.2626*x-2.0012)/(0.5000*1.5000))-1/3.7000))$ |
| 5.25 M urea 0 % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A=(0.7400/0.3200)/((3.7000/((-0.0446*x^3)+(0.3075*x^2)+0.4836*x-0.0409))-1)$ |
| Dynamic $\tau_1 = 3.0$ ns, $\tau_2 = 1.5$ ns | Green | $F_D/F_A=0.2000/(0.3200*((1/1.5000+1/3.0000-(1.4180*x-1.2719)/(3.0000*1.5000))-1/3.7000))$ |
| Dynamic $\tau_1 = 1.5$ ns, $\tau_2 = 0.5$ ns | | $F_D/F_A=0.2000/(0.3200*((1/1.5000+1/0.5000-((0.0000*x^3)+(-0.0000*x^2)+2.2626*x+-2.0012)/(0.5000*1.5000))-1/3.7000))$ |
| 7.5 M urea 0 % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A=(0.7000/0.3200)/((3.5000/((-0.0505*x^3)+(0.3293*x^2)+0.4765*x-0.0393))-1)$ |
| Dynamic $\tau_1 = 3.0$ ns, $\tau_2 = 1.6$ ns | Green | $F_D/F_A=0.2000/(0.3200*((1/1.6000+1/3.0000-(1.3538*x-1.0715)/(3.0000*1.6000))-1/3.5000))$ |
| Dynamic $\tau_1 = 1.6$ ns, $\tau_2 = 0.45$ ns | Olive | $F_D/F_A=0.2000/(0.3200*((1/1.6000+1/0.4500-(2.2977*x-2.1521)/(0.4500*1.6000))-1/3.5000))$ |
| 0 M urea $10^{-2}$ % w/v Tween 20 | | |

| | | |
|---|---|---|
| Static | Orange | $F_D/F_A$=(0.7091/0.4700)/((3.8152/((-0.0464*x^3)+(0.2887*x^2)+0.5864*x-0.0552))-1) |
| 4.5 M urea $10^{-2}$ % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.7500/0.3200)/((3.7500/((-0.0433*x^3)+(0.3025*x^2)+0.4853*x-0.0413))-1) |
| Dynamic $\tau_1$ = 3.0 ns, $\tau_2$ = 1.5ns | Green | $F_D/F_A$=0.2000/(0.3200*((1/1.5000+1/3.0000-(1.4180*x-1.2719)/(3.0000*1.5000))-1/3.7000)) |
| Dynamic $\tau_1$ = 1.5 ns, $\tau_2$ = 0.5 ns | Olive | $F_D/F_A$=0.2000/(0.3200*((1/1.5000+1/0.5000-(2.2626*x-2.0012)/(0.5000*1.5000))-1/3.7000)) |
| 5.25 M urea $10^{-2}$ % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.7400/0.3200)/((3.7000/((-0.0446*x^3)+(0.3075*x^2)+0.4836*x-0.0409))-1) |
| Dynamic $\tau_1$ = 3.0 ns, $\tau_2$ = 1.5 ns | Green | $F_D/F_A$=0.2000/(0.3200*((1/1.5000+1/3.0000-(1.4180*x-1.2719)/(3.0000*1.5000))-1/3.7000)) |
| Dynamic $\tau_1$ = 1.5 ns, $\tau_2$ = 0.5 ns | Olive | $F_D/F_A$=0.2000/(0.3200*((1/1.5000+1/0.5000-(2.2626*x-2.0012)/(0.5000*1.5000))-1/3.7000)) |
| 7.5 M Urea $10^{-2}$ % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.7000/0.3200)/((3.5000/((-0.0505*x^3)+(0.3293*x^2)+0.4765*x-0.0393))-1) |
| Dynamic $\tau_1$ = 3.0 ns, $\tau_2$ = 1.6 ns | Green | $F_D/F_A$=0.2000/(0.3200*((1/1.6000+1/3.0000-(1.3538*x-1.0715)/(3.0000*1.6000))-1/3.5000)) |
| Dynamic $\tau_1$ = 1.5 ns, $\tau_2$ = 0.5 ns | Olive | $F_D/F_A$=0.2000/(0.3200*((1/1.6000+1/0.4500-(2.2977*x-2.1521)/(0.4500*1.6000))-1/3.5000)) |

Table 9.4. **Static and dynamic FRET lines for K60pAcF/N132C-(DA) and S44pAcF/I150C variant at different Tween 20 concentrations as shown in the Figure 6.7.**

| FRET lines for K60pAcF/N132C-(DA) variant at 5.25 M | | |
|---|---|---|
| $g_G/g_R$ = 0.77 | $B_G$ and $B_R$ = 0.4 – 2.5 kHz | $\alpha$ = 1.7% |
| 0 % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.7400/0.3200)/((3.7000/((-0.0446*x^3)+(0.3075*x^2)+0.4836*x-0.0409))-1) |
| Dynamic $\tau_1$ = 3.0 ns, $\tau_2$ = 1.5 ns | Green | $F_D/F_A$=0.2000/(0.3200*((1/1.5000+1/3.0000-(1.4180*x-1.2719)/(3.0000*1.5000))-1/3.7000)) |
| Dynamic $\tau_1$ = 1.5 ns, $\tau_2$ = 0.5 ns | Olive | $F_D/F_A$=0.2000/(0.3200*((1/1.5000+1/0.5000-(2.2626*x-2.0012)/(0.5000*1.5000))-1/3.7000)) |
| $10^{-5}$ % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.7600/0.3200)/((3.8000/((-0.0421*x^3)+(0.2976*x^2)+0.4870*x- |

| | | |
|---|---|---|
| | | 0.0417))-1) |
| Dynamic $\tau_1$ = 3.1 ns, $\tau_2$ = 1.5 ns | Green | $F_D/F_A$=0.2000/(0.3200*((1/3.1000+1/1.5000-(1.4241*x-1.3324)/(1.5000*3.1000))-1/3.8000)) |
| $10^{-4}$ % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.7600/0.3200)/((3.8000/((-0.0421*x^3)+(0.2976*x^2)+0.4870*x+-0.0417))-1) |
| Dynamic $\tau_1$ = 3.1 ns, $\tau_2$ = 1.5 ns | Green | $F_D/F_A$=0.2000/(0.3200*((1/3.1000+1/1.5000-(1.4241*x-1.3324)/(1.5000*3.1000))-1/3.8000)) |
| $10^{-3}$ % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.7600/0.3200)/((3.8000/((-0.0421*x^3)+(0.2976*x^2)+0.4870*x-0.0417))-1) |
| Dynamic $\tau_1$ = 3.1 ns, $\tau_2$ = 1.5 ns | Green | $F_D/F_A$=0.2000/(0.3200*((1/3.1000+1/1.5000-(1.4241*x-1.3324)/(1.5000*3.1000))-1/3.8000)) |
| $10^{-2}$ % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.7600/0.3200)/((3.8000/((-0.0421*x^3)+(0.2976*x^2)+0.4870*x-0.0417))-1) |
| Dynamic $\tau_1$ = 3.1 ns, $\tau_2$ = 1.5 ns | Green | $F_D/F_A$=0.2000/(0.3200*((1/3.1000+1/1.5000-(1.4241*x-1.3324)/(1.5000*3.1000))-1/3.8000)) |
| **FRET lines for S44pAcF/I150C variant at pH 2** | | |
| 0 % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.8200/0.3200)/((4.1000/((-0.0356*x^3)+(0.2711*x^2)+0.4964*x-0.0440))-1) |
| Dynamic $\tau_1$ = 4.0 ns, $\tau_2$ = 0.5ns | Green | $F_D/F_A$=0.2000/(0.3200*((1/4.0000+1/0.5000-(2.0951*x-4.3808)/(0.5000*4.0000))-1/4.1000)) |
| $10^{-5}$ % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.8500/0.3200)/((4.2500/((-0.0328*x^3)+(0.2594*x^2)+0.5009*x-0.0451))-1) |
| Dynamic $\tau_1$ = 4.0 ns, $\tau_2$ = 0.5 ns | Green | $F_D/F_A$=0.2000/(0.3200*((1/4.0000+1/0.5000-(2.0951*x-4.3808)/(0.5000*4.0000))-1/4.1000)) |
| $10^{-4}$ % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.8500/0.3200)/((4.2500/((-0.0328*x^3)+(0.2594*x^2)+0.5009*x+-0.0451))-1) |
| Dynamic $\tau_1$ = 4.0 ns, $\tau_2$ = 0.5 ns | Green | $F_D/F_A$=0.2000/(0.3200*((1/4.0000+1/0.5000-(2.0951*x-4.3808)/(0.5000*4.0000))-1/4.1000)) |
| Dynamic $\tau_1$ = 4.25 ns, $\tau_2$ = 2.0 ns | Olive | $F_D/F_A$=0.2000/(0.3200*((1/4.2500+1/2.0000-(1.2766*x-1.1755)/(2.0000*4.2500))-1/4.2500)) |
| $10^{-3}$ % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.8500/0.3200)/((4.2500/((-0.0328*x^3)+(0.2594*x^2)+0.5009*x-0.0451))-1) |
| Dynamic | Green | $F_D/F_A$=0.2000/(0.3200*((1/4.2500+1/2.0000-(1.2766*x- |

| | | |
|---|---|---|
| $\tau_1$ = 4.25 ns, $\tau_2$ = 2.0 ns | | 1.1755)/(2.0000*4.2500))-1/4.2500)) |
| Dynamic<br>$\tau_1$ = 2.0 ns, $\tau_2$ = 0.5 ns | Olive | $F_D/F_A$=0.2000/(0.3200*((1/2.0000+1/0.5000-(2.2485*x-2.5702)/(0.5000*2.0000))-1/4.2500)) |
| $10^{-2}$ % w/v Tween 20 | | |
| Static | Orange | $F_D/F_A$=(0.8500/0.3200)/((4.2500/((-0.0328*x^3)+(0.2594*x^2)+0.5009*x-0.0451))-1) |
| Dynamic<br>$\tau_1$ = 4.25 ns, $\tau_2$ = 2.0 ns | Green | $F_D/F_A$=0.2000/(0.3200*((1/4.2500+1/2.0000-(1.2766*x-1.1755)/(2.0000*4.2500))-1/4.2500)) |
| Dynamic<br>$\tau_1$ = 2.0 ns, $\tau_2$ = 0.5 ns | Olive | $F_D/F_A$=0.2000/(0.3200*((1/2.0000+1/0.5000-(2.2485*x-2.5702)/(0.5000*2.0000))-1/4.2500)) |

# 10. References

Abrahams JP, Leslie AG, Lutter R, Walker JE (1994) Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria. *Nature* **370:** 621-628

Al-Soufi W, Reija B, Novo M, Felekyan S, Kühnemuth R, Seidel CAM (2005) Fluorescence correlation spectroscopy, a tool to investigate supramolecular dynamics: Inclusion complexes of pyronines with cyclodextrin. *Journal of the American Chemical Society* **127:** 8775-8784

Allen GS, Frank J (2007) Structural insights on the translation initiation complex: ghosts of a universal initiation complex. *Molecular microbiology* **63:** 941-950

Antonik M, Felekyan S, Gaiduk A, Seidel CAM (2006) Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *Journal of Physical Chemistry B* **110:** 6970-6978

Antoun A, Pavlov MY, Lovmar M, Ehrenberg M (2006) How initiation factors maximize the accuracy of tRNA selection in initiation of bacterial protein synthesis. *Molecular cell* **23:** 183-193

Arnold GE, Ornstein RL (1992) A molecular dynamics simulation of bacteriophage T4 lysozyme. *Protein Eng* **5:** 703-714

Baase WA, Liu L, Tronrud DE, Matthews BW (2010) Lessons from the lysozyme of phage T4. *Protein Sci* **19:** 631-641

Bachmann A, Kiefhaber T, Boudko S, Engel J, Bachinger HP (2005) Collagen triple-helix formation in all-trans chains proceeds by a nucleation/growth mechanism with a purely entropic barrier. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 13897-13902

Becker W (2005) Advanced Time-Correlated Single Photon Counting Techniques. *Springer Series in Chemical Physics* **81:** 331

Bieri O, Wildegger G, Bachmann A, Wagner C, Kiefhaber T (1999) A salt-induced kinetic intermediate is on a new parallel pathway of lysozyme folding. *Biochemistry* **38:** 12460-12470

Biou V, Shu F, Ramakrishnan V (1995) X-ray crystallography shows that translational initiation factor IF3 consists of two compact alpha/beta domains linked by an alpha-helix. *EMBO J* **14:** 4056-4064

Böhmer M, Wahl M, Rahn HJ, Erdmann R, Enderlein J (2002) Time-resolved fluorescence correlation spectroscopy. *Chem Phys Lett* **353:** 439-445

Borgia A, Wensley BG, Soranno A, Nettels D, Borgia MB, Hoffmann A, Pfeil SH, Lipman EA, Clarke J, Schuler B (2012) Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. *Nat Commun* **3**

Bouvignies G, Vallurupalli P, Hansen DF, Correia BE, Lange O, Bah A, Vernon RM, Dahlquist FW, Baker D, Kay LE (2011) Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **477:** 111-114

Bowman GR, Pande VS (2010a) Millisecond Timescale, Atomistic Protein Folding Simulations Yield a Network Theory for Protein Folding. *Biophysical Journal* **98:** 614a-614a

Bowman GR, Pande VS (2010b) Protein folded states are kinetic hubs. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 10890-10895

Brustad EM, Lemke EA, Schultz PG, Deniz AA (2008) A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer. *J Am Chem Soc* **130:** 17664-17665

Cai Q, Kusnetzow AK, Hideg K, Price EA, Haworth IS, Qin PZ (2007) Nanometer distance measurements in RNA using site-directed spin labeling. *Biophys J* **93:** 2110-2117

Calogero RA, Pon CL, Canonaco MA, Gualerzi CO (1988) Selection of the mRNA translation initiation region by Escherichia coli ribosomes. *Proceedings of the National Academy of Sciences of the United States of America* **85:** 6427-6431

Candotti M, Esteban-Martin S, Salvatella X, Orozco M (2013) Toward an atomistic description of the urea-denatured state of proteins. *Proceedings of the National Academy of Sciences of the United States of America* **110:** 5933-5938

Cellitti J, Bernstein R, Marqusee S (2007a) Exploring subdomain cooperativity in T4 lysozyme II: Uncovering the C-terminal subdomain as a hidden intermediate in the kinetic folding pathway. *Protein Science* **16:** 852-862

Cellitti J, Llinas M, Echols N, Shank EA, Gillespie B, Kwon E, Crowder SM, Dahlquist FW, Alber T, Marqusee S (2007b) Exploring subdomain cooperativity in T4 lysozyme I: Structural and energetic studies of a circular permutant and protein fragment. *Protein Science* **16:** 842-851

Chang JY, Li L (2002) The unfolding mechanism and the disulfide structures of denatured lysozyme. *FEBS letters* **511:** 73-78

Chen Y, Hu DH, Vorpagel ER, Lu HP (2003) Probing single-molecule T4 lysozyme conformational dynamics by intramolecular fluorescence energy transfer. *Journal of Physical Chemistry B* **107:** 7947-7956

Chiang YW, Borbat PP, Freed JH (2005a) The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. *Journal of magnetic resonance* **172:** 279-295

Chiang YW, Borbat PP, Freed JH (2005b) Maximum entropy: a complement to Tikhonov regularization for determination of pair distance distributions by pulsed ESR. *Journal of magnetic resonance* **177:** 184-196

Choi S, Chae J (2010) Methods of reducing non-specific adsorption in microfluidic biosensors. *J Micromech Microeng* **20**

Choi Y, Moody IS, Sims PC, Hunt SR, Corso BL, Perez I, Weiss GA, Collins PG (2012) Single-molecule lysozyme dynamics monitored by an electronic circuit. *Science* **335:** 319-324

Chung HS, Louis JM, Eaton WA (2009) Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 11837-11844

Culver GM, Noller HF (1999) Efficient reconstitution of functional Escherichia coli 30S ribosomal subunits from a complete set of recombinant small subunit ribosomal proteins. *Rna* **5:** 832-843

Dale RE, Eisinger J, Blumberg WE (1979) Orientational freedom of molecular probes - Orientation factor in intra-molecular energy transfer. *Biophysical Journal* **26:** 161-193

Dallas A, Noller HF (2001) Interaction of translation initiation factor 3 with the 30S ribosomal subunit. *Molecular cell* **8:** 855-864

de Groot BL, Hayward S, van Aalten DM, Amadei A, Berendsen HJ (1998) Domain motions in bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data. *Proteins* **31:** 116-127

Dill KA, MacCallum JL (2012) The Protein-Folding Problem, 50 Years On. *Science* **338:** 1042-1046

Dix JA, Hom EF, Verkman AS (2006) Fluorescence correlation spectroscopy simulations of photophysical phenomena and molecular interactions: a molecular dynamics/monte carlo approach. *The journal of physical chemistry B* **110:** 1896-1906

Dixon MM, Nicholson H, Shewchuk L, Baase WA, Matthews BW (1992) Structure of a hinge-bending bacteriophage T4 lysozyme mutant, Ile3-->Pro. *J Mol Biol* **227:** 917-933

Durant JA, Chen C, Laue TM, Moody TP, Allison SA (2002) Use of T4 lysozyme charge mutants to examine electrophoretic models. *Biophysical chemistry* **101-102:** 593-609

Eckert C, Durchschlag H, Tiefenback K (2006) Thermodynamic Analysis of Lysozyme Denaturation by Surfactants. *Progress in Colloid and Polymer Science* **133:** 123-130

Elson EL (2013) 40 Years of FCS: How It All Began. In *Fluorescence Fluctuation Spectroscopy*, Tetin SY (ed), Vol. 518, pp 1-10. San Diego: Elsevier Academic Press Inc

Elson EL, Magde D (1974) Fluorescence Correlation Spectroscopy. I. Conceptual Basis and Theory. *Biopolymers* **13:** 1-27

Elvekrog MM, Gonzalez RL, Jr. (2013) Conformational selection of translation initiation factor 3 signals proper substrate selection. *Nat Struct Mol Biol* **20:** 628-633

Enderlein J, Robbins DL, Ambrose WP, Goodwin PM, Keller RA (1997) Statistics of single-molecule detection. *Journal of Physical Chemistry B* **101:** 3626-3632

Fabbretti A, Pon CL, Hennelly SP, Hill WE, Lodmell JS, Gualerzi CO (2007) The real-time path of translation factor IF3 onto and off the ribosome. *Molecular cell* **25:** 285-296

Felekyan S, Kalinin S, Sanabria H, Valeri A, Seidel CAM (2012) Filtered FCS: species auto- and cross-correlation functions highlight binding and dynamics in biomolecules. *Chemphyschem* **13:** 1036-1053

Felekyan S, Kühnemuth R, Kudryavtsev V, Sandhagen C, Becker W, Seidel CAM (2005) Full correlation from picoseconds to seconds by time-resolved and time-correlated single photon detection. *Rev Sci Instrum* **76:** 083104

Felekyan S, Sanabria H, Kalinin S, Kühnemuth R, Seidel CAM (2013) Analyzing Förster resonance energy transfer (FRET) with fluctuation algorithms. *Methods in Enzymology* **519:** 39-85

Fierz B, Reiner A, Kiefhaber T (2009) Local conformational dynamics in alpha-helices measured by fast triplet transfer. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 1057-1062

Fleissner MR, Brustad EM, Kalai T, Altenbach C, Cascio D, Peters FB, Hideg K, Peuker S, Schultz PG, Hubbell WL (2009) Site-directed spin labeling of a genetically encoded unnatural

amino acid. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 21637-21642

Fries JR, Brand L, Eggeling C, Köllner M, Seidel CAM (1998) Quantitative identification of different single molecules by selective time-resolved confocal fluorescence spectroscopy. *J Phys Chem A* **102:** 6601-6613

Frydman J (2001) Folding of newly translated proteins in vivo: The role of molecular chaperones. *Annu Rev Biochem* **70:** 603-647

Gao M, She ZS, Zhou RH (2010) Key Residues that Play a Critical Role in Urea-Induced Lysozyme Unfolding. *Journal of Physical Chemistry B* **114:** 15687-15693

Garcia C, Fortier PL, Blanquet S, Lallemand JY, Dardel F (1995a) 1H and 15N resonance assignments and structure of the N-terminal domain of Escherichia coli initiation factor 3. *European journal of biochemistry / FEBS* **228:** 395-402

Garcia C, Fortier PL, Blanquet S, Lallemand JY, Dardel F (1995b) Solution structure of the ribosome-binding domain of E. coli translation initiation factor IF3. Homology with the U1A protein of the eukaryotic spliceosome. *J Mol Biol* **254:** 247-259

Gardino AK, Villali J, Kivenson A, Lei M, Liu CF, Steindel P, Eisenmesser EZ, Labeikovsky W, Wolf-Watz M, Clarkson MW, Kern D (2009) Transient non-native hydrogen bonds promote activation of a signaling protein. *Cell* **139:** 1109-1118

Gibb B, Ye LF, Gergoudis SC, Kwon YH, Niu HY, Sung P, Greene EC (2014) Concentration-Dependent Exchange of Replication Protein A on Single-Stranded DNA Revealed by Single-Molecule Imaging. *Plos One* **9**

Gohlke H, Hergert U, Meyer T, Mulnaes D, Grieshaber MK, Smits SH, Schmitt L (2013) Binding region of alanopine dehydrogenase predicted by unbiased molecular dynamics simulations of ligand diffusion. *Journal of Chemical Information and Modeling* **53:** 2493–2498

Gopich I, Szabo A (2005) Theory of photon statistics in single-molecule Forster resonance energy transfer. *Journal of Chemical Physics* **122:** 14707

Gopich IV, Szabo A (2003) Single-macromolecule fluorescence resonance energy transfer and free-energy profiles. *Journal of Physical Chemistry B* **107:** 5058-5063

Gopich IV, Szabo A (2007) Single-molecule FRET with diffusion and conformational dynamics. *Journal of Physical Chemistry B* **111:** 12925-12932

Gopich IV, Szabo A (2012a) Theory of Single-Molecule Fret Efficiency Histograms. *Adv Chem Phys* **146:** 245-297

Gopich IV, Szabo A (2012b) Theory of the energy transfer efficiency and fluorescence lifetime distribution in single-molecule FRET. *Proceedings of the National Academy of Sciences of the United States of America* **109:** 7747-7752

Goto NK, Skrynnikov NR, Dahlquist FW, Kay LE (2001) What is the average conformation of bacteriophage T4 lysozyme in solution? A domain orientation study using dipolar couplings measured by solution NMR. *J Mol Biol* **308:** 745-764

Graham JS, Johnson RC, Marko JF (2011) Concentration-dependent exchange accelerates turnover of proteins bound to double-stranded DNA. *Nucleic Acids Res* **39:** 2249-2259

Green R, Noller HF (1999) Reconstitution of functional 50S ribosomes from in vitro transcripts of Bacillus stearothermophilus 23S rRNA. *Biochemistry* **38:** 1772-1779

Grunberg-Manago M, Dessen P, Pantaloni D, Godefroy-Colburn T, Wolfe AD, Dondon J (1975) Light-scattering studies showing the effect of initiation factors on the reversible dissociation of Escherichia coli ribosomes. *Journal of Molecular Biology* **94:** 461-478

Gualerzi CO, Brandi L, Caserta E, Garofalo C, Lammi M, La Teana A, Petrelli D, Spurio R, Tomsic J, Pon CL (2001) Initiation factors in the early events of mRNA translation in bacteria. *Cold Spring Harb Symp Quant Biol* **66:** 363-376

Gurunathan K, Levitus M (2010) FRET Fluctuation Spectroscopy of Diffusing Biopolymers: Contributions of Conformational Dynamics and Translational Diffusion. *Journal of Physical Chemistry B* **114:** 980-986

Hammes GG (1964) Mechanism of Enzyme Catalysis. *Nature* **204:** 342-343

Hemmen K (2015) Dissertation. *Heinrich-Heine-University, Düsseldorf*

Hemmen K, Rodnin D, Rohrbeck D, Adariani SR, Sanabria H, Seidel CAM (2014) Deciphering Folding Pathways of Phage T4 Lysozyme: Influence of Multiple Conformations. *Biophysical Journal* **106:** 259a-259a

Henzler-Wildman K, Kern D (2007) Dynamic personalities of proteins. *Nature* **450:** 964-972

Hilz H, Wiegers U, Adamietz P (1975) Stimulation of Proteinase-K Action by Denaturing Agents - Application to Isolation of Nucleic-Acids and Degradation of Masked Proteins. *European Journal of Biochemistry* **56:** 103-108

Hofmann H, Hillger F, Pfeil SH, Hoffmann A, Streich D, Haenni D, Nettels D, Lipman EA, Schuler B (2010) Single-molecule spectroscopy of protein folding in a chaperonin cage. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 11793-11798

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

Hu D, Lu HP (2004) Placing single-molecule T4 lysozyme enzymes on a bacterial cell surface: toward probing single-molecule enzymatic reaction in living cells. *Biophys J* **87:** 656-661

Hua Y, Raleigh DP (1998) Conformational analysis of the interdomain linker of the central homology region of chloroplast initiation factor IF3 supports a structural model of two compact domains connected by a flexible tether. *FEBS Lett* **433:** 153-156

Jagannathan B, Elms PJ, Bustamante C, Marqusee S (2012) Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein. *Proceedings of the National Academy of Sciences of the United States of America* **109:** 17820-17825

Johnson CK (2006) Calmodulin, conformational states, and calcium signaling. A single-molecule perspective. *Biochemistry* **45:** 14233-14246

Johnson KA, Goody RS (2011) The original Michaelis constant: Translation of the 1913 Michaelis–Menten paper. *Biochemistry* **50:** 8264-8269

Julian P, Milon P, Agirrezabala X, Lasso G, Gil D, Rodnina MV, Valle M (2011) The cryo-EM Structure of a complete 30S translation initiation complex from Escherichia coli. *PLoS Biol* **9:** e1001095

Kalinin S, Felekyan S, Antonik M, Seidel CAM (2007) Probability distribution analysis of single-molecule fluorescence anisotropy and resonance energy transfer. *Journal of Physical Chemistry B* **111:** 10253-10262

Kalinin S, Peulen T, Sindbert S, Rothwell PJ, Berger S, Restle T, Goody RS, Gohlke H, Seidel CAM (2012) A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nature methods* **9:** 1218-1225

Kalinin S, Sisamakis E, Magennis SW, Felekyan S, Seidel CAM (2010a) On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits. *Journal of Physical Chemistry B* **114:** 6197–6206

Kalinin S, Valeri A, Antonik M, Felekyan S, Seidel CAM (2010b) Detection of structural dynamics by FRET: a photon distribution and fluorescence lifetime analysis of systems with multiple states. *J Phys Chem B* **114:** 7983-7995

Kask P, Palo K, Ullmann D, Gall K (1999) Fluorescence-intensity distribution analysis and its application in biomolecular detection technology. *Proceedings of the National Academy of Sciences of the United States of America* **96:** 13756-13761

Kiefhaber T (1995) Kinetic Traps in Lysozyme Folding. *Proceedings of the National Academy of Sciences of the United States of America* **92:** 9029-9033

Kim SA, Heinze KG, Bacia K, Waxham MN, Schwille P (2005) Two-photon cross-correlation analysis of intracellular reactions with variable stoichiometry. *Biophysical Journal* **88:** 4319-4336

Kleckner IR, Foster MP (2011) An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta* **1814:** 942-968

Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H (2002) Long-range interactions within a nonnative protein. *Science* **295:** 1719-1722

Koshioka M, Saski K, Masuhara H (1995) Time-Dependent Fluorescene Depolarization Analysis in Three Dimensional Microscopy. *Applied Spectroscopy* **49:** 224-228

Kou SC, Cherayil BJ, Min W, English BP, Xie XS (2005) Single-molecule Michaelis-Menten equations. *Journal of Physical Chemistry B* **109:** 19068-19081

Krieger F, Fierz B, Axthelm F, Joder K, Meyer D, Kiefhaber T (2004) Intrachain diffusion in a protein loop fragment from carp parvalbumin. *Chem Phys* **307:** 209-215

Krieger F, Fierz B, Bieri O, Drewello M, Kiefhaber T (2003) Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *Journal of Molecular Biology* **332:** 265-274

Kubelka J, Hofrichter J, Eaton WA (2004) The protein folding 'speed limit'. *Current opinion in structural biology* **14:** 76-88

Kühnemuth R, Seidel CAM (2001) Principles of single molecule multiparameter fluorescence spectroscopy. *Single Molecules* **2:** 251-254

Kuroki R, Weaver LH, Matthews BW (1993) A covalent enzyme-substrate intermediate with saccharide distortion in a mutant T4 lysozyme. *Science* **262:** 2030-2033

Kuroki R, Weaver LH, Matthews BW (1995) Structure-based design of a lysozyme with altered catalytic activity. *Nat Struct Biol* **2:** 1007-1011

Kycia JH, Biou V, Shu F, Gerchman SE, Graziano V, Ramakrishnan V (1995) Prokaryotic translation initiation factor IF3 is an elongated protein consisting of two crystallizable domains. *Biochemistry* **34:** 6183-6187

La Teana A, Pon CL, Gualerzi CO (1993) Translation of mRNAs with degenerate initiation triplet AUU displays high initiation factor 2 dependence and is subject to initiation factor 3 repression. *Proceedings of the National Academy of Sciences of the United States of America* **90:** 4161-4165

Lakowicz JR (2006) *Principles of Fluorescence Spectroscopy*, Third edn. New York: Springer.

Lange OF, Grubmüller H (2008) Full correlation analysis of conformational protein dynamics. *Proteins* **70:** 1294-1312

Laurence TA, Kapanidis AN, Kong XX, Chemla DS, Weiss S (2004) Photon arrival-time interval distribution (PAID): A novel tool for analyzing molecular interactions. *Journal of Physical Chemistry B* **108:** 3051-3067

Laursen BS, Sorensen HP, Mortensen KK, Sperling-Petersen HU (2005) Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev* **69:** 101-123

Lemke EA (2011) Site-specific labeling of proteins for single-molecule FRET measurements using genetically encoded ketone functionalities. *Methods in Molecular Biology* **751:** 3-15

Levitus M (2010) Relaxation Kinetics by Fluorescence Correlation Spectroscopy: Determination of Kinetic Parameters in the Presence of Fluorescent Impurities. *Journal of Physical Chemistry Letters* **1:** 1346-1350

Lint JR, Leopold BD, Hurst GA, Hamrick WJ (1992) Determining Effective Study Area Size from Marked and Harvested Wild Turkey Gobblers. *J Wildlife Manage* **56:** 556-562

Lippincott ER, Schroeder R (1955) One-Dimensional Model of the Hydrogen Bond. *Journal of Chemical Physics* **23:** 1099-1106

Llinas M, Gillespie B, Dahlquist FW, Marqusee S (1999a) The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nat Struct Biol* **6:** 1072-1078

Llinas M, Gillespie B, Dahlquist FW, Marqusee S (1999b) The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nature Structural Biology* **6:** 1072-1078

Llinas M, Marqusee S (1998) Subdomain interactions as a determinant in the folding and stability of T4 lysozyme. *Protein Science* **7:** 96-104

Lopez-Alonso JP, Bruix M, Font J, Ribo M, Vilanova M, Jimenez MA, Santoro J, Gonzalez C, Laurents DV (2010) NMR Spectroscopy Reveals that RNase A is Chiefly Denatured in 40% Acetic Acid: Implications for Oligomer Formation by 3D Domain Swapping. *Journal of the American Chemical Society* **132:** 1621-1630

Lu HP (2011) Revealing time bunching effect in single-molecule enzyme conformational dynamics. *Physical Chemistry Chemical Physics* **13:** 6734-6749

Lu M, Lu HP (2014) Probing protein multidimensional conformational fluctuations by single-molecule multiparameter photon stamping spectroscopy. *The journal of physical chemistry B* **118:** 11943-11955

MacDougall DD, Gonzalez RL, Jr. (2014) Translation initiation factor 3 regulates switching between different modes of ribosomal subunit joining. *Journal of Molecular Biology*

Magde D, Elson EL, Webb WW (1972) Thermodynamic fluctuations in a reacting system - measurement by fluorescence correlation spectroscopy. *Physical Review Letters* **29:** 705-708

Magde D, Elson EL, Webb WW (1974) Fluorescence correlation spectroscopy. II. An experimental realization. *Biopolymers* **13:** 29-61

Matthews BW, Nicholson H, Becktel WJ (1987) Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proceedings of the National Academy of Sciences of the United States of America* **84:** 6663-6667

Matthews BW, Remington SJ (1974) The three dimensional structure of the lysozyme from bacteriophage T4. *Proceedings of the National Academy of Sciences of the United States of America* **71:** 4178-4182

Maus M, Cotlet M, Hofkens J, Gensch T, De Schryver FC, Schäffer J, Seidel CAM (2001) An experimental comparison of the maximum likelihood estimation and nonlinear least-squares fluorescence lifetime analysis of single molecules. *Analytical chemistry* **73:** 2078-2086

McCutcheon JP, Agrawal RK, Philips SM, Grassucci RA, Gerchman SE, Clemons WM, Jr., Ramakrishnan V, Frank J (1999) Location of translational initiation factor IF3 on the small ribosomal subunit. *Proceedings of the National Academy of Sciences of the United States of America* **96:** 4301-4306

McHaourab HS, Lietzow MA, Hideg K, Hubbell WL (1996) Motion of spin-labeled side chains in T4 lysozyme. Correlation with protein structure and dynamics. *Biochemistry* **35:** 7692-7704

McHaourab HS, Oh KJ, Fang CJ, Hubbell WL (1997) Conformation of T4 lysozyme in solution. Hinge-bending motion and the substrate-induced conformational transition studied by site-directed spin labeling. *Biochemistry* **36:** 307-316

Mcpherson A, Koszelak S, Axelrod H, Day J, Robinson L, Mcgrath M, Williams R, Cascio D (1986) The Effects of Neutral Detergents on the Crystallization of Soluble-Proteins. *J Cryst Growth* **76:** 547-553

Milon P, Carotti M, Konevega AL, Wintermeyer W, Rodnina MV, Gualerzi CO (2010) The ribosome-bound initiation factor 2 recruits initiator tRNA to the 30S initiation complex. *EMBO Reports* **11:** 312-316

Milon P, Konevega AL, Gualerzi CO, Rodnina MV (2008) Kinetic checkpoint at a late step in translation initiation. *Molecular cell* **30:** 712-720

Milon P, Konevega AL, Peske F, Fabbretti A, Gualerzi CO, Rodnina MV (2007) Transient kinetics, fluorescence, and FRET in studies of initiation of translation in bacteria. *Methods Enzymol* **430:** 1-30

Milon P, Maracci C, Filonava L, Gualerzi CO, Rodnina MV (2012) Real-time assembly landscape of bacterial 30S translation initiation complex. *Nature structural & molecular biology* **19:** 609-615

Milon P, Rodnina MV (2012) Kinetic control of translation initiation in bacteria. *Crit Rev Biochem Mol Biol* **47:** 334-348

Mittermaier A, Kay LE (2006) New tools provide new insights in NMR studies of protein dynamics. *Science* **312:** 224-228

Mooers BH, Baase WA, Wray JW, Matthews BW (2009) Contributions of all 20 amino acids at site 96 to the stability and structure of T4 lysozyme. *Protein Sci* **18:** 871-880

Moreau M, de Cock E, Fortier PL, Garcia C, Albaret C, Blanquet S, Lallemand JY, Dardel F (1997) Heteronuclear NMR studies of E. coli translation initiation factor IF3. Evidence that the inter-domain region is disordered in solution. *J Mol Biol* **266:** 15-22

Mulder FA, Mittermaier A, Hon B, Dahlquist FW, Kay LE (2001) Studying excited states of proteins by NMR spectroscopy. *Nature Structural Biology* **8:** 932-935

Muller-Spath S, Soranno A, Hirschfeld V, Hofmann H, Ruegger S, Reymond L, Nettels D, Schuler B (2010) From the Cover: Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 14609-14614

Muschielok A, Andrecka J, Jawhari A, Bruckner F, Cramer P, Michaelis J (2008) A nano-positioning system for macromolecular structural analysis. *Nat Methods* **5:** 965-971

Myasnikov AG, Simonetti A, Marzi S, Klaholz BP (2009) Structure-function insights into prokaryotic and eukaryotic translation initiation. *Current opinion in structural biology* **19:** 300-309

Nojiri M, Saito T (1997) Structure and function of poly(3-hydroxybutyrate) depolymerase from Alcaligenes faecalis T1. *J Bacteriol* **179:** 6965-6970

Onuchic JN, Wolynes PG (2004) Theory of protein folding. *Current opinion in structural biology* **14:** 70-75

Palo K, Mets U, Loorits V, Kask P (2006) Calculation of photon-count number distributions via master equations. *Biophysical Journal* **90:** 2179-2191

Parker MJ, Marqusee S (1999) The cooperativity of burst phase reactions explored. *Journal of Molecular Biology* **293:** 1195-1210

Petrelli D, LaTeana A, Garofalo C, Spurio R, Pon CL, Gualerzi CO (2001) Translation initiation factor IF3: two domains, five functions, one mechanism? *The EMBO journal* **20:** 4560-4569

Pfeil SH, Wickersham CE, Hoffmann A, Lipman EA (2009) A microfluidic mixing system for single-molecule measurements. *Rev Sci Instrum* **80**

Price ES, Aleksiejew M, Johnson CK (2011) FRET-FCS detection of intralobe dynamics in calmodulin. *Journal of Physical Chemistry B* **115:** 9320-9326

Price ES, DeVore MS, Johnson CK (2010) Detecting Intramolecular Dynamics and Multiple Förster Resonance Energy Transfer States by Fluorescence Correlation Spectroscopy. *Journal of Physical Chemistry B* **114:** 5895-5902

Prigozhin MB, Liu YX, Wirth AJ, Kapoor S, Winter R, Schulten K, Gruebele M (2013) Misplaced helix slows down ultrafast pressure-jump protein folding. *Proceedings of the National Academy of Sciences of the United States of America* **110:** 8087-8092

Quillin ML, Breyer WA, Griswold IJ, Matthews BW (2000) Size versus polarizability in protein-ligand interactions: binding of noble gases within engineered cavities in phage T4 lysozyme. *J Mol Biol* **302:** 955-977

Rothwell PJ, Allen WJ, Sisamakis E, Kalinin S, Felekyan S, Widengren J, Waksman G, Seidel CAM (2013) dNTP-dependent conformational transitions in the fingers subdomain of Klentaq1

DNA polymerase: insights into the role of the "nucleotide-binding" state. *Journal of Biological Chemistry* **288:** 13575-13591


Rothwell PJ, Berger S, Kensch O, Felekyan S, Antonik M, Wohrl BM, Restle T, Goody RS, Seidel CAM (2003) Multiparameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 reverse transcriptase: primer/template complexes. *Proceedings of the National Academy of Sciences of the United States of America* **100:** 1655-1660


Sagle LB, Zhang YJ, Litosh VA, Chen X, Cho Y, Cremer PS (2009) Investigating the Hydrogen-Bonding Model of Urea Denaturation. *Journal of the American Chemical Society* **131:** 9304-9310


Santoso Y, Joyce CM, Potapova O, Le Reste L, Hohlbein J, Torella JP, Grindley ND, Kapanidis AN (2010) Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 715-720


Schaffer J, Volkmer A, Eggeling C, Subramaniam V, Striker G, Seidel CAM (1999) Identification of single molecules in aqueous solution by time-resolved fluorescence anisotropy. *J Phys Chem A* **103:** 331-336


Schwille P. (2001) Fluorescence Correlation Spectroscopy. Theory and Applications. In E.L. Elson RR (ed.). Springer, Berlin, Vol. 65, pp. 360-378.


Selkoe DJ (2003) Folding proteins in fatal ways. *Nature* **426:** 900-904


Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y, Wriggers W (2010) Atomic-level characterization of the structural dynamics of proteins. *Science* **330:** 341-346


Shoichet BK, Baase WA, Kuroki R, Matthews BW (1995) A relationship between protein stability and protein function. *Proceedings of the National Academy of Sciences of the United States of America* **92:** 452-456


Sindbert S, Kalinin S, Nguyen H, Kienzler A, Clima L, Bannwarth W, Appel B, Muller S, Seidel CA (2011a) Accurate distance determination of nucleic acids via Forster resonance energy transfer: implications of dye linker length and rigidity. *J Am Chem Soc* **133:** 2463-2480


Sindbert S, Kalinin S, Nguyen H, Kienzler A, Clima L, Bannwarth W, Appel B, Muller S, Seidel CAM (2011b) Accurate distance determination of nucleic acids via Forster resonance energy transfer: implications of dye linker length and rigidity. *Journal of the American Chemical Society* **133:** 2463-2480


Sisamakis E, Valeri A, Kalinin S, Rothwell PJ, Seidel CAM (2010) Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods Enzymol* **475:** 455-514

Skrynnikov NR, Dahlquist FW, Kay LE (2002) Reconstructing NMR spectra of "invisible" excited protein states using HSQC and HMQC experiments. *J Am Chem Soc* **124:** 12352-12360

Slaughter BD, Allen MW, Unruh JR, Bieber Urbauer RJ, Johnson CK (2004) Single-Molecule Resonance Energy Transfer and Fluorescence Correlation Spectroscopy of Calmodulin in Solution. *Journal of Physical Chemistry B* **108:** 10388-10397

Slaughter BD, Bieber-Urbauer RJ, Johnson CK (2005a) Single-molecule tracking of sub-millisecond domain motion in calmodulin. *Journal of Physical Chemistry B* **109:** 12658-12662

Slaughter BD, Unruh JR, Allen MW, Bieber Urbauer RJ, Johnson CK (2005b) Conformational substates of calmodulin revealed by single-pair fluorescence resonance energy transfer: influence of solution conditions and oxidative modification. *Biochemistry* **44:** 3694-3707

Slaughter BD, Urbauer RJ, Urbauer JL, Johnson CK (2007) Mechanism of calmodulin recognition of the binding domain of isoform 1b of the plasma membrane Ca(2+)-ATPase: kinetic pathway and effects of methionine oxidation. *Biochemistry* **46:** 4045-4054

Smock RG, Gierasch LM (2009) Sending signals dynamically. *Science* **324:** 198-203

Sussman JK, Simons EL, Simons RW (1996) Escherichia coli translation initiation factor 3 discriminates the initiation codon in vivo. *Molecular microbiology* **21:** 347-360

Tompa P (2012) On the supertertiary structure of proteins. *Nat Chem Biol* **8:** 597-600

Torres T, Levitus M (2007) Measuring Conformational Dynamics: A New FCS-FRET Approach. *Journal of Physical Chemistry B* **111:** 7392-7400

Tsai A, Petrov A, Marshall RA, Korlach J, Uemura S, Puglisi JD (2012) Heterogeneous pathways and timing of factor departure during translation initiation. *Nature* **487:** 390-393

Tsugita A, Inouye M (1968) Purification of bacteriophage T4 lysozyme. *J Biol Chem* **243:** 391-397

Uversky VN (2009) Intrinsic disorder in proteins associated with neurodegenerative diseases. *Front Biosci-Landmrk* **14:** 5188-5238

Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **41:** 415-427

van der Meer BW, Cooker G, Chen SY (1994) *Resonance Energy Transfer: Theory and Data*, New York: VCH Publishers.

Wahlgren M, Kedström J, Arnebrant T (1997) The interactions in solution between nonionic surfactant and globular proteins: effects on cloud point. *Journal of Dispersion Science and Technology* **18:** 449-458

Weiel J, Hershey JW (1978) Fluorescence polarization studies of the binding of fluorescein-labeled initiation factor IF3 to 30 S ribosomal subunits from Escherichia coli. *FEBS letters* **87:** 103-106

Weiss S (2000) Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nature Structural Biology* **7:** 724-729

Widengren J, Kudryavtsev V, Antonik M, Berger S, Gerken M, Seidel CA (2006) Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. *Analytical chemistry* **78:** 2039-2050

Wintermeyer W, Gualerzi C (1983) Effect of Escherichia coli initiation factors on the kinetics of N-AcPhe-tRNAPhe binding to 30S ribosomal subunits. A fluorescence stopped-flow study. *Biochemistry* **22:** 690-694

Woźniak AK, Schröder G, Grubmüller H, Seidel CAM, Oesterhelt F (2008) Single molecule FRET measures bends and kinks in DNA. *ProcNatlAcadSciUSA* **105:** 18337-18342

Wunderlich B, Nettels D, Benke S, Clark J, Weidner S, Hofmann H, Pfeil SH, Schuler B (2013) Microfluidic mixer designed for performing single-molecule kinetics with confocal detection on timescales from milliseconds to minutes. *Nat Protoc* **8:** 1459-1474

Yirdaw RB, McHaourab HS (2012) Direct observation of T4 Lysozyme hinge-bending motion by fluorescence correlation spectroscopy. *Biophysical Journal* **103:** 1525-1536

Yu NJ, Spremulli LL (1997) Structural and mechanistic studies on chloroplast translational initiation factor 3 from Euglena gracilis. *Biochemistry* **36:** 14827-14835

Yusupova G, Jenner L, Rees B, Moras D, Yusupov M (2006) Structural basis for messenger RNA movement on the ribosome. *Nature* **444:** 391-394

Zacharias M (2008) Combining elastic network analysis and molecular dynamics simulations by hamiltonian replica exchange. *Journal of Chemical Theory and Computation* **4:** 477-487

Zhang K, Yang H (2005) Photon-by-photon determination of emission bursts from diffusing single chromophores. *Journal of Physical Chemistry B* **109:** 21930-21937

Zhang XJ, Wozniak JA, Matthews BW (1995) Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. *J Mol Biol* **250:** 527-552

Zheng D, Lu HP (2014) Single-molecule enzymatic conformational dynamics: spilling out the product molecules. *The journal of physical chemistry B* **118:** 9128-9140

Hiermit erkläre ich, dass ich die Promotion mit dem Thema

„Protein dynamics as studied by single molecule Förster resonance energy transfer"

Am Institute für Physikalische Chemie II der Heinrich-Heine-Universität Düsseldorf unter der Leitung von Prof. Dr. Claus A.M. Seidel eigenständig und ohne unerlaubte Hilfe angefertigt und in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht habe.

Es existieren keine vorherigen Promotionsversuche.

Düsseldorf, den 13.07.2015

Dmitro Rodnin