

Structure and conformational fluctuations of Phage T4 Lysozyme under native and denaturing conditions

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Katherina Hemmen
aus Wildeshausen

Düsseldorf, Juni 2016

aus dem Institut für Molekulare Physikalische Chemie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

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

Korreferent: Prof. Dr. Dieter Willbold

Tag der mündlichen Prüfung: 5. Juli 2016

ZUSAMMENFASSUNG

In der vorliegenden Arbeit wurden Einzelmolekül- und ensemble Fluoreszenzmethoden angewandt, um das Verhalten des Modellenzym Lysozym aus dem T4 Phagen (T4L) unter nativen und denaturierenden Bedingungen zu charakterisieren.

Bei der Substratumsetzung bewegen die beiden Subdomänen von T4L sich in einer Scharnierbewegung zueinander. Wir untersuchten die Dynamik dieser Bewegung durch ein Netz von FRET-Paaren und ordneten den beobachteten Übergängen und konformativen Zuständen einzelne Schritte im enzymatischen Zyklus zu. Durch eine globale Analyse von 24 FRET-Varianten wurden drei verschiedene Konformationen identifiziert. Zwei Konformationen ähneln den bekannten "offen" und "geschlossen" Zuständen von T4L, während die dritte (ein thermisch angeregten Zustand) bisher nicht bekannt ist. Durch Varianten, die die enzymatische Aktivität von T4L beeinflussen, wurde die Energielandschaft der katalytischen Spaltung im enzymatischen Zyklus abgeleitet. Mit diesem Wissen wurde der Entfaltungspfad von T4L durch einen kombinierten Ansatz aus CD-Spektroskopie und Fluoreszenzmessungen untersucht. Dies ermöglichte die Identifizierung eines Entfaltungspfades über mindestens zwei Intermediatzustände. Wir finden einen sequentiellen Verlust von Strukturelementen und schlagen vor, dass das Konformationsgleichgewicht nativer Zustände eine entscheidende Rolle bei der Entfaltung des Proteins, das in Subdomänen organisiert ist, spielt. Fluktuationsanalyse deckt relevante Zeitskalen von Nano- bis Millisekunden ab. Dies ermöglicht die Herleitung eines Entfaltungsschemas, das schnelle (sub Mikrosekunden) und langsame (über Millisekunden) Kinetiken berücksichtigt. Der denaturierte Zustand von T4L wurde in 7.5 M Harnstoff durch ein FRET-Abstandsnetzwerk untersucht. Harnstoff denaturiertes T4L verhält sich wie ein nicht-ideales Polymer, und zeigte große Heterogenität auf der Einzelmolekülebene. Regionen in T4L, die scheinbare residuelle Ordnung zeigten, wurden mittels einer Steifigkeitsmatrix identifiziert und die lokale Seitenkettenmobilität mit Fluoreszenzanisotropiemessungen ausgewertet. Während die einzelnen Subdomänen gut voneinander getrennt sind, fanden wir, dass beide, insbesondere die C-terminale Subdomäne, nativ-artig angeordnet sind und eine unebene Energielandschaft zeigen, die mittels Fluktuationsanalyse beschrieben wurde.

Insgesamt scheint die N-terminale Domäne von T4L weniger stabil zu sein. Dies führt zu einer sequentiellen Entfaltung von T4L. Dies könnte für die Funktion von Bedeutung sein, da die Katalyse mit Konformationsänderungen innerhalb der Proteinstruktur in Beziehung steht. Daher sind sowohl die Proteinstruktur als auch die Dynamik hierarchisch organisiert. Insgesamt lässt sich sagen, dass die Kombination verschiedenster Fluoreszenzmethoden sehr gut geeignet ist, um Proteindynamiken zeitlich über sieben Größenordnungen zu beschreiben und wenig populierte, angeregte Zustände zu entdecken. Es konnten noch existierende Lücken in dem schon sehr gut untersuchten System T4L geschlossen werden, was das herausragende Potenzial von Fluoreszenz-basierten Methoden in der Strukturbiologie aufzeigt.

SUMMARY

In the presented work single-molecule and ensemble fluorescence methods were applied to the model enzyme Phage T4 Lysozyme (T4L) to characterize its behavior under native and denaturing conditions.

For T4L it is known that its two subdomains undergo a hinge-bending motion while processing its substrate. We studied the dynamics of this motion by a network of FRET-pairs and related the observed transitions and conformational states to distinct steps in the enzymatic cycle. By a global analysis of 24 FRET variants, three distinct conformations were identified. Two conformations resemble known “open” and “closed” states of T4L, while the third (thermally excited state) is so far unknown. By variants, which alter the enzymatic activity of T4L, the energy landscape of the catalytic cleavage cycle was derived.

With this knowledge the unfolding pathway of T4L was studied by a combined approach of CD spectroscopy and fluorescence measurements. This allowed disentangling of the unfolding steps with at least two intermediates. We find a sequential loss of structural elements and propose that the conformational equilibrium of native states plays a critical role in the unfolding of the protein, which is organized in subdomains. Fluctuation analysis covers relevant timescales from nano- to milliseconds. This allows deriving an unfolding reaction scheme, which considers fast (sub microsecond) and slow (above milliseconds) kinetics.

The denatured state of T4L was studied in 7.5 M urea by a distance network. It was found that urea denatured T4L behaves like a non-ideal polymer, which showed large heterogeneity on the single-molecule level. Regions in T4L, which showed apparent residual order, were identified by building a stiffness matrix and evaluating the local side chain mobility with fluorescence anisotropy. While being well separated, we found that the individual subdomains, specifically the C-terminal subdomain, are arranged in a native-like fashion and show a rough energy landscape, as probed by fluctuation analysis.

Overall, the N-terminal subdomain seems to be less stable. This results in a sequential unfolding of T4L. This might be relevant for the function, as the catalysis is related to conformational changes within the protein structure. Therefore, both protein structure and dynamics are hierarchically organized. In this work, it could be shown that the combination of several different fluorescence methods is well suited to firstly detect so far “hidden” excited, i.e. minuscule populated, states and describe fast dynamics (e.g. enzymatic cleavage under native conditions, unfolding dynamics) over more than seven orders of magnitude in time. Here, still existing gaps in the knowledge of the very well-studied model enzyme T4L could be filled – proving the outstanding potential of fluorescence in the field of structure biology where classical methods reach their limits.

CONTENTS

ZUSAMMENFASSUNG	i
CONTENTS	iii
ABBREVIATIONS & SYMBOLS	1
1. Introduction	3
1.1. Protein Structure and function.....	3
1.2. Protein unfolding.....	5
1.2.1. The unfolded state of proteins (DSE)	6
1.2.2. Methods to study the DSE	7
1.3. Phage T4 Lysozyme – a model enzyme.....	8
1.3.1. Native structure and function.....	8
1.3.2. Unfolding of T4L.....	9
1.4. Aims	10
2. Theoretical Background	14
2.1. Fluorescence.....	14
2.1.1. Fluorescence Lifetime.....	15
2.1.2. Fluorescence Anisotropy	16
2.2. Förster Resonance Energy Transfer (FRET).....	17
2.3. Fluorescence spectroscopic methods	18
2.4. Single Molecule Multiparameter Fluorescence Detection (MFD).....	19
2.5. Photon Distribution Analysis (PDA).....	21
2.6. Fluorescence Correlation Spectroscopy (FCS)	22
2.7. Accessible Volume Simulation (AVSim)	24
3. Structure and function of T4L under native conditions.....	26
3.1. Dynamics & Function of transient states in single enzyme molecules	26
3.2. Manuscript.....	27
3.2.1. Introduction.....	28

CONTENTS

3.2.2.	Results.....	30
3.2.3.	Discussion	42
3.2.4.	Materials and Methods.....	47
3.3.	Supplementary Information.....	50
3.3.1.	Materials and Methods.....	50
3.3.2.	Supporting results	65
3.4.	Outlook.....	92
4.	Unfolding pathway of T4L	93
4.1.	Multiparameter Fluorescence spectroscopy reveals sub-millisecond intermediate transitions in protein unfolding	93
4.2.	Manuscript.....	94
4.2.1.	Introduction.....	95
4.2.2.	Results.....	97
4.2.3.	Discussion	104
4.2.4.	Materials & Methods	111
4.3.	Supplementary Information.....	114
4.3.1.	Methods.....	114
4.3.2.	Supporting results	128
5.	The denatured state of T4L.....	168
5.1.	Fluorescence spectroscopic toolkit resolves the heterogeneity of denatured states.....	168
5.2.	Manuscript.....	169
5.2.1.	Introduction.....	170
5.2.2.	Results.....	172
5.2.3.	Discussion	183
5.2.4.	Materials & Methods	191
5.3.	Supplementary Information.....	194
5.3.1.	Methods.....	194

CONTENTS

5.3.2. Supporting results	210
6. Summary & Outlook	241
6.1. Structure and function of T4L under native conditions	241
6.2. Unfolding pathway of T4L.....	242
6.3. Characterization of the denatured state of T4L.....	243
6.4. Fluorescence as a complementary tool in structural biology	244
7. References	245
8. Appendix	253
8.1. Vector Map and Nucleotide Sequence pET11a-T4L wt*	253
8.2. Additional Materials & Methods.....	254
8.2.1. Materials	254
8.2.2. Molecular biological methods.....	257
8.2.3. Protein production & purification.....	261
8.2.4. Protein analytical methods.....	262
8.2.5. Steady-state fluorescence measurements.....	266
8.2.6. Bioinformatics & Software	268
8.3. Exemplary results for Protein Production, Purification & Labeling	270
8.3.1. Site-directed mutagenesis	270
8.3.2. Protein production & purification.....	270
8.3.3. Site-specific labeling.....	271
8.4. Influence of urea on the Förster Radius	272
8.5. Steady-state Fluorescence	273
<i>PRESENTATIONS & PUBLICATIONS</i>	275
<i>DANKSAGUNG</i>	277

CONTENTS

ABBREVIATIONS & SYMBOLS

	Meaning		Meaning
S_1, S_n	1 st /n-excited electronic singlet state	k_{12}, k_{21}	exchange rate constant
(f)FCS	(filtered) fluorescence correlation spectroscopy	kDa	kilo Dalton
(s)eTCPSC	(sub)ensemble time-correlated single photon counting	K_{eq}	equilibrium constant, k_{12}/k_{21}
(s/d)PDA	(static/dynamic) photon distribution analysis	k_F	rate constant of fluorescence
*g	g force	k_{FRET}	FRET rate constant
$\langle B_g \rangle, \langle B_r \rangle$	green and red background count rates	k_{IC}	rate constant of internal conversion
$\langle R_{DA} \rangle$	mean donor-acceptor distance	k_{ISC}	rate constant of intersystem crossing
$\langle R_{DA} \rangle_E$	FRET-efficiency weighted average distance	LB	lysogeny broth
$\langle \tau \rangle_f$	fluorescence-weighted averaged fluorescence lifetime	LF	low FRET
$\langle \tau \rangle_x$	species-weighted averaged fluorescence lifetime	L_{link}	linker length in AVSim
$\langle \tau_{D(A)} \rangle_f$	fluorescence-weighted donor lifetime	m	slope of the unfolding rise term
Å	Angström (10^{-10} m)	MFD	Multiparameter fluorescence detection
aa	amino acid	m_N, m_D	slope of the native and denatured state baseline
Abs	absorption	MWCO	molecular weight cut off
ACF	autocorrelation	N	native state, native protein
Adir	direct acceptor excitation of double labeled sample	N	number of particles in detection volume
Amp	Ampicillin	n_D	refractive index
AVSim	accessible volume simulation	NMR	nuclear magnetic resonance
b	anisotropy fraction	OD ₆₀₀	optical density at wavelength of 600nm
BIFL	burst-integrated fluorescence lifetime	pAcF	<i>para</i> -Acetylphenylalanine
BSA	bovine serum albumin	PBS	phosphate buffered saline
C_1, C_2, C_3	native conformations of T4L	PCR	polymerase chain reaction
Cam	Chloramphenicol	PDB	Protein database
CatIEX	cation ion exchange chromatography	R	universal gas constant, 8.3145 J/K*mol
CCF	cross-correlation	$\langle R_{DA} \rangle_E$	mean distance from PDA
CD	Circular dichroism	$r, r(t)$	(time-resolved) anisotropy
cf	correction factor	r_0	fundamental anisotropy
c_p, c_{A488}, c_{A647}	concentration of protein, donor dye (Alexa 488), acceptor dye (Alexa647)	R_0	Förster radius
CV	column volume	r_{ss}	steady-state anisotropy
Cys	Cystein	r_∞	residual anisotropy
d	path length of cuvette	R_D, R_A	position of donor/ acceptor molecule
DA	double labeled sample	$\langle R_{DA} \rangle$	mean distance between the donor and acceptor dye of Gaussian distribution
ddH ₂ O	double-distilled water	R_{dye}	dye radius in AVSim
dF	dilution factor	$r_{ee,L}$	maximum, linker-corrected end-to-end distance
DNA	Deoxyribonucleic acid	R_{mp}	distance between mean donor and acceptor positions
dNTP	desoxy-nucleotide triphosphate	Rpm	rounds per minute
DOnly, D0	signal or sample labeled with Donor only	S	uncorrected signal
DSE	ensemble of denatured states	S_0	electronic ground state
$E, \langle E \rangle$	(time-averaged) FRET efficiency	$sCCF, sACF$	species-selective cross- or auto correlation function
<i>E. coli</i>	<i>Escherichia coli</i>	SD200	superdex 200 (GPC column)
EPR	electron paramagnetic resonance	SDS	sodium dodecyl sulfate
F	corrected fluorescence signal	SDS	SDS polyacrylamide gel electrophoresis

ABBREVIATIONS & SYMBOLS

		PAGE	
f	form factor	T	temperature in Kelvin
F_{\parallel}, F_{\perp}	fluorescence signal detected either vertically or horizontally after vertical excitation	T_1	1 st excited triplet state
$F_{D(A)}(t)$	fluorescence intensity decay of double labeled sample	J	spectral overlap integral
F_D, F_A	fluorescence signal of donor/ acceptor dye	T4L	lysozyme from the T4 phage
$F_{DOnly}(t)$	fluorescence intensity decay of donor only labeled sample	TAE	Tris-Acetate-EDTA buffer
FRET	Förster resonance energy transfer	t_d, t_{diff}	diffusion time
G	g factor, Gerätefaktor	Tris	Tris(hydroxymethyl)aminomethane
$G(t_c)$	correlation curve	Trp	tryptophan
$G_{ab}(t_c)$	antibunching term in fFCS	TW	time window in PDA
$G_{abun}(t_c)$	antibunching term in FCS	U	unfolded state, unfolded protein
$G_b(t_c)$	bunching term in fFCS	UV	ultra violett
$G_{bun}(t_c)$	bunching term in FCS	V	volt
GC	Gaussian chain	V	vertical position of polarizer (0°)
$G_{cab}(t_c)$	antibunching term in FCS in CCF	w/v	Mass concentration (weight/ volume)
$G_{diff}(t_c)$	diffusion term in FCS, fFCS	WLC	Worm like chain
GdmCl	guanidine hydrochloride	w_{Link}	linker width in AVSim
g_g, g_r	detection efficiencies in green and red	x	fraction
goi	gene of interest	y_N, y_D	native and denatured state baseline
GPC	gel permeation chromatography	α	crosstalk
$G_{rc}(t_c)$	rotational correlation term in FCS	β	angle between excitation and emission dipole
$G_{rc,c}(t_c)$	rotational correlation term in FCS for CCF	ΔG^0	free energy of unfolding
$G_{slow}(t_c)$	slow dynamics term in fFCS	$\varepsilon_P, \varepsilon_D, \varepsilon_A$	extinction coefficient of protein, donor/ acceptor dye
H	horizontal position of polarizer (90°)	η	viscosity
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	κ^2	orientation factor
HF	high FRET	λ	wavelength [nm]
I, J	(un) folding intermediate	ρ	rotational correlation times
IEX	Ion exchange chromatography	σ_{DA}	half-width of Gaussian distribution
IPTG	isopropyl- β -D-thiogalactopyranosid	τ	fluorescence lifetime
J	spectral overlap integral	Φ_F	fluorescence quantum yield
k	inverse of the fluorescence lifetime τ	ω_0/z_0	geometry factor

1. Introduction

Proteins have many important functions in the entirety of organisms – starting from simple cytoskeleton forming proteins up to complex (bio-molecular) machineries built up out of dozens of subunits. They are produced from ribosomes as unstructured polypeptide chains and then either fold by themselves or assisted with “helper-proteins” (chaperones) into their functional, three dimensional structures. How does accurate protein folding occur exactly? This is still one of the unsolved questions in modern biophysics. It is especially intriguing as many diseases have been shown to be associated with an accumulation of misfolded and/or aggregated proteins e.g. Alzheimer’s (1). So far the main focus in understanding proteins and their behavior has been lying on the structures captured by X-ray crystallography, which represent frequently the thermodynamically most stable states (this ignores potential crystal packing effects). However, only recently it is an emerging point of view that the dynamics, which link the different crystallized forms is also important, if not even more important, than those “static” structures. Multidimensional NMR methods can determine residue-specific the high-resolution solution structure of small proteins and the dynamic motions within the protein (2, 3). Only the dynamic exchange and the equilibrium between different native protein conformations involving thermally excited states can give insight into how those bio-molecular machines really work (4).

The focus of this thesis will lie on the dynamics of an enzyme under native conditions, the involved (excited) states and conformational transitions and how they relate to the enzymatic function (Chapter 3). Next, the energy landscape of the unfolding pathway, which leads the protein from its folded to the unfolded state, is studied with a particular focus on intermediate transitions and structures (Chapter 4). Finally, the residual structure and dynamics within the denatured state will be characterized (Chapter 5). To be able to capture the dynamic processes on a wide range of time scales and to disentangle the heterogeneity of the observed states, a broad combination of mostly (single-molecule) fluorescence methods will be used, supported by other spectroscopic methods (e.g. CD spectroscopy) and computer simulations.

1.1. Protein Structure and function

Proteins are produced as a linear chain usually out of twenty different amino acids at the ribosome in the cell. Their so-called primary structure (sequence of amino acids in the peptide chain) ranges from only a few tens (peptides) up to several hundreds of amino acid residues and the order of these determine the proteins’ 3-dimensional structure and function.

This linear chain firstly folds into short strands of secondary structure elements, α -helices and β -sheets connected by turns, loops and unstructured regions, which are then rearranged to the protein's unique 3-dimensional structure and proteins composed of several chains are arranged in a quaternary structure. The protein fold is frequently organized in (sub-) domains, which either have each their own function to fulfill or the dynamic interaction with the other (sub-) domains is required for function. In contrast to domains, subdomains are not stable on their own, but need other (sub-) domains of the protein for stabilization. The dynamic interactions of these (sub-) domains lead to the fact, that a single structure is no longer a sufficient description for the respective protein. These conformationally different states are often summarized as “supertertiary” structure or structural ensembles (5).

This leads to the following views. Fully folded proteins are inherently dynamic on a wide range of time scales – starting from bond vibrations in the fs range (Figure 1.1.1) over side chain rotamers in the ps - ns range up to larger domain motions in the ms range (4). The dynamics is needed to fulfill a broad variety of important functions inside and outside the cells (6). The cells are homeostats, which react to their changing environment by the specific production of distinct proteins. Enzymes e.g. catalyze nearly all the multitude of chemical reactions taking place in a cell, whereas hormones transmit signals to coordinate the biological processes inside a cell and between different cells, tissues and organs. Additional proteins are needed to provide the cell's structure (cytoskeleton), to transport ions and small molecules, and to protect against foreign particles, bacteria and viruses. These mediated functions generally occur on time scales larger than ms and the mechanism of action depends highly on their structure and structural dynamics.

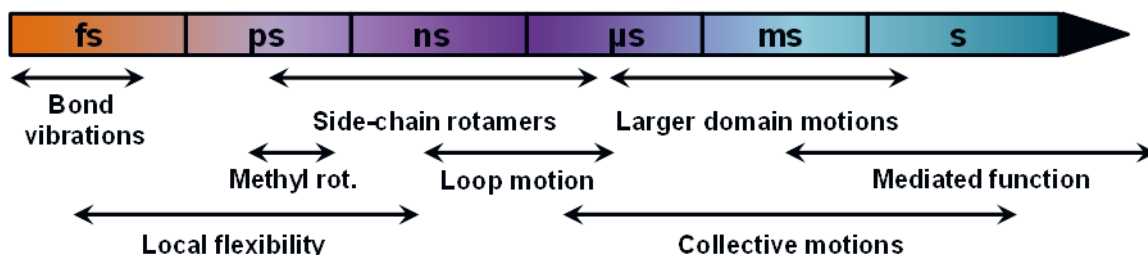


Figure 1.1.1 Time scales of dynamics in proteins. In general, a protein displays dynamic behavior in a wide range of time scales – starting from fast bond vibrations and methyl rotations of its monomers' amino acids in the femto seconds (fs, orange) over the local backbone mobility up to the correlated functional motions of larger protein domains in the seconds (s, turquoise) range. Modified from (4).

Problems might arise, when due to so far unknown changes in the cells, misfolded protein moieties become the most stable conformations – a situation known from several neurodegenerative diseases like Alzheimer's. Here, abnormally folded proteins form plaques

in the brain and finally block signal transduction in the nerve cells (1). Diseases like this strengthen the need in understanding how protein folding – unfolding occurs and how to influence the delicate equilibrium.

1.2. Protein unfolding

In general, the native, folded structure of proteins is stable only under physiological conditions (pH, temperature or solvent). Thus, their native contacts can be easily dissolved e.g. by an increase (or for some decrease) in temperature or by the addition of denaturants, where the two most commonly used are urea and guanidine hydrochloride (GdmCl). The progress of unfolding can be followed by a number of experimental measures. Circular dichroism spectroscopy e.g. reports on the secondary structure of proteins, Tryptophane fluorescence probes on changes in their immediate environment, whereas calorimetry assesses the energetics (stability) of the unfolding structures. Overall, unfolding is observed by probing a (spectroscopic) signal which shows a difference in amplitude for the folded and unfolded state. As depicted in Figure 1.2.1A, its contribution is determined by the equilibrium of the enthalpy and entropy (of protein, solvent and denaturants), and thus the resulting free energy, of both native and unfolded states. Under native conditions, the folded structure is by design the most stable conformation, unfolded moieties are high in energy and – if populated – then only in a minuscule amount. Under denaturing conditions it is *vice versa*, the unfolded state is more stable than the native conformation.

Mostly, the recorded change in the observed signal is seen to follow a sigmoidal shape (Figure 1.2.1A) which can be described with a two state model developed by (7) (Eq.1.2.1):

$$y = \frac{y_N + m_N \cdot c + (y_D + m_D \cdot c) \cdot \exp(-\Delta G^0 - m \cdot c / R \cdot T)}{1 + \exp(-\Delta G^0 - m \cdot c / R \cdot T)} \quad 1.2.1$$

With y_N and y_D being the native and denatured state baseline, m_N and m_D the native and denatured state slope, m is the slope of the rise part and ΔG^0 the free energy of unfolding. However, this model assumes that the folded protein unfolds in a single step reaction, folded \rightarrow denatured conformation, which has been shown not to reflect the real unfolding behavior of the great majority of proteins (8) beyond the two state equilibrium of N and U (Figure 1.2.1B). A more detailed model on protein unfolding is given by the so-called *folding funnel* shown in Figure 1.2.1B (9, 10). Here, the free energy landscape is believed to narrow down, starting from a nearly flat area under denaturing conditions, where the unfolded

conformations exchange quickly between each other, over the rough funnel sides, depicting possibly intermediate and/or trapped states, up to the narrow funnel tip representing the folded protein conformation(s), while minimizing the free energy of the protein.

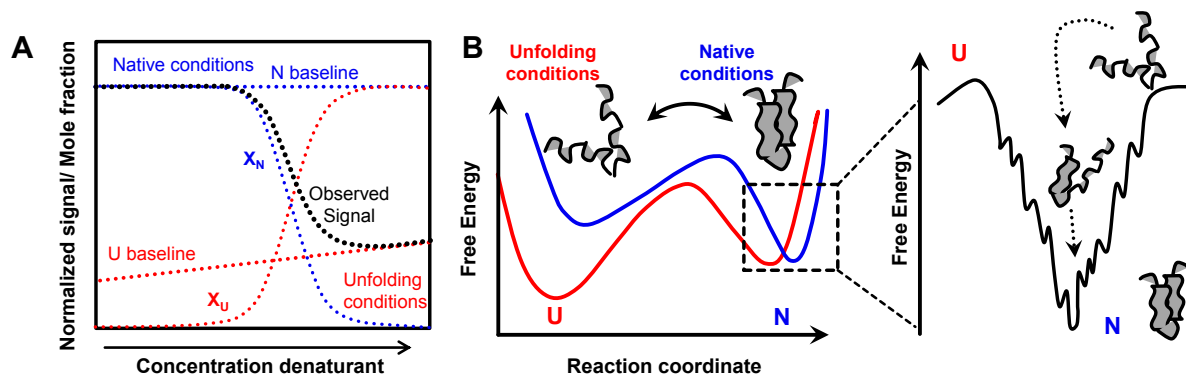


Figure 1.2.1 Protein unfolding occurs from equilibrium of native states to an ensemble of unfolded states (DSE). (A) Under native conditions, the folded state(s) (N, blue) is energetically preferred, whereas under denaturing conditions the DSE (U, red) is preferred. (B) For a typical two state protein, the observed signal changes in a sigmoidal manner (black) where the slope of the native (blue dotted) and unfolded baseline (red dotted) is sample depending. Fraction of native (N, blue) and unfolded (U, red) molecules adds up to 1. Modified and compiled from: (7, 9, 11). The folding funnel depicts a more realistic model of protein folding; the native protein state resembles the free energy minimum under native conditions, a large ensemble of unfolded conformations folds with towards it by decreasing in free energy while possibly encountering several local minima (trapped states) and/ or intermediate states.

1.2.1. The unfolded state of proteins (DSE)

As sketched in Figure 1.2.1B, the denatured form of proteins should not be assigned to a single, stable structure but rather is believed to consist of many, interchanging conformations (denoted as denatured state ensemble, DSE) without noteworthy energy barriers between them (12). Nowadays it is usual to interpret the observed experimental results in terms of various polymer models, which have originally been defined in polymer physics to describe the polymer's behavior using principles from statistical mechanics (13-15). The most commonly used simple models here are the Ideal or Gaussian chain model (GC) and the Worm-like chain model (WLC) (16).

However, the long-standing belief of the fact that the properties of the DSE can be described by polymer models implicates the complete lack of tertiary or secondary structure is currently weakening. Although the average ensemble might behave like a chemical polymer, for the individual conformers it is indeed likely to contain (transient) secondary structure – native or non-native – or even show tertiary structure interactions (15, 17). This issue has been the focus of Fitzkee and Rose in 2004 who took the crystal structures of more than 30 proteins and simulated their behavior if only ~ 8 % of their residues would be as flexible as expected in an unfolded protein. Surprisingly, this degree of flexibility is enough to reproduce the

polymer measures determined experimentally and in simulations under highly denaturing conditions (18).

1.2.2. Methods to study the DSE

The DSE and the way from the native protein to its unfolded state can be studied with many different methods summarized in the lower part of Figure 1.2.2, with the time scales for processes generally occurring while protein folds are shown on top (19). The extraction of folding and unfolding rates is mostly done with stopped-flow methods, where the sample is rapidly diluted into an unfolding or refolding condition and the signal evolution on time scales longer than a few milliseconds is recorded. Typically, this method is combined with site-directed mutagenesis to selectively (de-)stabilize parts of the protein (20, 21).

At the moment, the DSE itself is mostly studied by several NMR methods which cover a broad range of time scales from picoseconds to seconds and, by a combination of several methods e.g. NOE effects and residual dipolar couplings (22), the number of restraints obtained for the DSE can be enhanced – the lack of sufficient restraints and the large dynamic heterogeneity are being the major problems. However, of course, for the DSE no structure alike the native state can be obtained. The experimentally determined parameters are primarily used to locate possible residual structure (or intermediate structures in slightly destabilizing conditions). The time- and ensemble-averaged experimental measures have to be analyzed carefully and validated by independent methods (e. g. fluorescence) and/ or by re-calculating the experimental measures from the gained simulated structures (22). The use of restraints gained from fluorescence measurements, which can probe the protein's behavior over a comparable time scale as NMR, is now an emerging field.

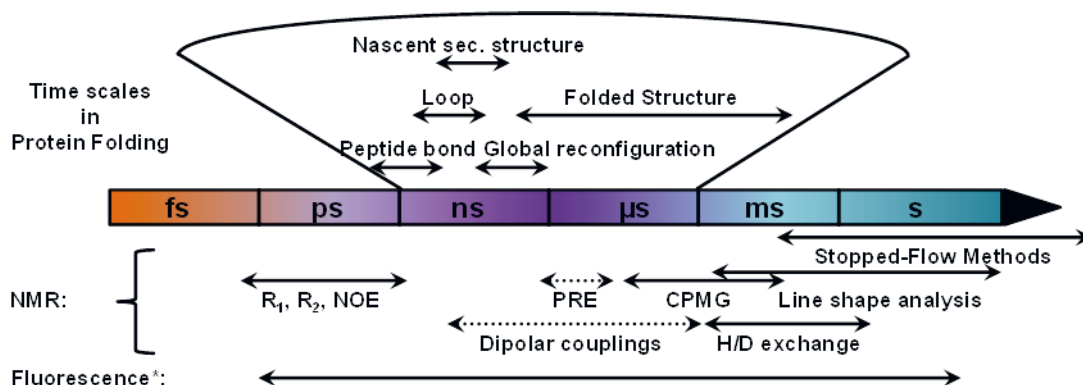


Figure 1.2.2 Time scales of and methods to study protein folding. The upper part shows the time scales of the diverse processes involved in protein folding according to (19). The lower part gives an overview of methods generally used to study the dynamics and folding of protein in NMR (Compiled from (23-25)). Additionally, time scales where fluorescence or stopped flow methods can be applied are shown. *Details on fluorescence methods are shown in Figure 2.3.1.

Yet, the lack of sufficient restraints is always a point of critic and might only be overcome by placing many fluorescence probes all over the protein to be able to probe or analyze each direction and motion. Also, here care has to be taken when inferring either the structural characteristics or the polymer characteristics of a protein. Some ensemble, average measures have been shown by simulation to be significantly sensible to the label positions within the protein (16).

1.3. Phage T4 Lysozyme – a model enzyme

The lysozyme produced from the Phage T4 (T4L) is used as model enzyme in biochemistry and structural biology for decades. It has already helped to understand the universal genetic code and recently been grouped among the 50 most important crystal structures in the protein database (PDB) (26). In fact, the PDB lists to date more than 500 structures containing T4L – mostly due to its amenability to crystallization even when hampered strongly with its original amino acid sequence or when attached as a “crystallization helper” to much larger, difficult to crystallize proteins (27). T4L has been the subject of more than a dozen studies on e.g. the mutational effects on protein stability by exchanging for example polar against non-polar residues or as many residues as possible against alanine, or on the effects of (large) insertions or deletions in the protein structure (28-31).

1.3.1. Native structure and function

T4L is a glycohydrolase (EC 3.2.1.17) produced from the T4 phage late in its infection cycle and lyses the bacterial cell wall of Gram negative bacteria while hydrolyzing the bond between the N-acetylmuramic acid (NAM) and N-acetylglucosamine (NAG) (32). It has 164 amino acids (18.3 kDa) and was firstly crystallized by Matthews in 1974 (33). Structurally T4L consist of an N-terminal subdomain which is made of both α -helices and β -sheets, whereas the C-terminal lobe consists only of α -helices. Both subdomains are connected by the long α -helix C. Interestingly helix A, comprising the first twelve residues in sequence, belongs structurally-wise to the C-terminus (Figure 1.3.1A).

When comparing structures which do not include too many mutations, they can be grouped in two conformers (34): an open (blue, Figure 1.3.1B) and a closed (violet) conformation. The latter structure was crystallized with its covalently bound product (35). EPR studies done by Mchaourab *et al.*, who labeled each of the two protein lobes and observed substrate induced signal changes, lead to the conclusion that the hinge-bending observed between the two conformers must be important for its biological function (35, 36).

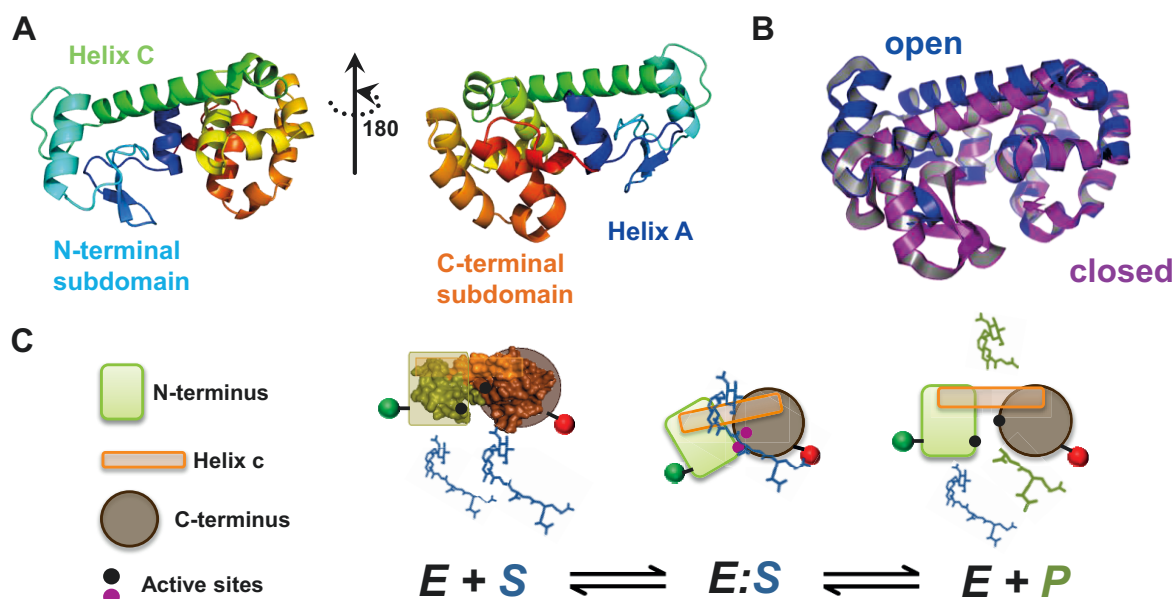


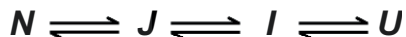
Figure 1.3.1 Structure and function of Phage T4 Lysozyme. (A) X-ray structure of T4L. (B) overlay of X-ray structures representative of the free (“open” conformation, 172L) and peptidoglycan bound (“closed” conformation, 148L) of T4L. (C) Simplified substrate cleavage scheme. T4L binds its substrate peptidoglycan, the gap between the N- and C-terminal closes to bring the active residues (E11, T26) closer, the bond between N-acetylmuramic acid and N-acetylglucosamine is hydrolyzed and the products are released.

Its catalytic residues T26 and E11 are located within the cleft between the two lobes, which close and reopen during substrate catalysis (36, 37) (highlighted Figure 1.3.1C). When the original threonine is mutated to glutamic acid, the reaction does not proceed to a full cleavage of the substrate but a covalent enzyme-substrate adduct is formed (35). If the residue 11 is replaced by an alanine (E11A), the first step in the reaction is blocked and T4L shows only binding- unbinding, but no catalytic activity (38).

1.3.2. Unfolding of T4L

The first attempt to characterize the unfolding of T4L was undertaken in 1975 by Elwell and Schellmann who found it behaving like a typical two-state folder under slightly acidic conditions where both the CD and the tryptophan (Trp) fluorescence signal are changing synchronously with increasing concentrations of GdmCl (39). However, this simple unfolding scheme was questioned already in 1984 by Desmadril and Yon, who determined a shift of change in Trp fluorescence to higher GdmCl concentrations under neutral conditions (40). This lead to first hints that (1) the unfolding of T4L might be more complex than assumed at first glance and involve possibly an intermediate where the three Trp’s in the C-terminal subdomain are still protected from solvent and (2) the observed results are highly dependent upon the conditions used for the unfolding experiment.

Currently, the existence of at least two intermediates (J , an unfolding intermediate, and I , a folding intermediate) has been shown by determining folding and unfolding rates of various mutants in stopped flow experiments (20):



with N being the native, folded conformations and U the ensemble of unfolded states. The time scale for the interconversion between $J \rightarrow I$ has been measured by stopped flow experiments to be in the milliseconds to seconds range, whereas the first ($N \rightarrow J$) and last step ($I \rightarrow U$) occur too fast ($< \text{ms}$) to be measured with the methods used (20).

Additionally, several studies (21, 41-44) were able to describe the structure of at least one intermediate. By assessing the stability of the single subdomains, circular permutants and the full-length protein with site-specific mutations and by performing native-state hydrogen exchange studies, they all agree on an intermediate consisting of a folded C-terminal and an unfolded N-terminal subdomain. Nonetheless, the structure of a second intermediate resembling a largely native-like structure has been proposed in 1992 by Lu and Dahlquist, whose relevance is at the moment under discussion as it might be an artefact due to the high pH treatment needed in the pulsed hydrogen labeling NMR (20, 43, 45).

1.4. Aims

The whole work in this thesis deals with the model enzyme Phage T4 Lysozyme and the application of ensemble and single molecule fluorescence spectroscopy methods to characterize its behavior in its folded (Chapter 3) and unfolded state (Chapter 5) via its unfolding transition (Chapter 4) (Figure 1.4.1). The main aim is to explore the possibilities fluorescence spectroscopy offers in the field of structural biology - which is traditionally occupied by crystallographers and NMR spectroscopists -, especially regarding the characterization of the proteins' dynamic behavior, transiently populated states and its denatured form.

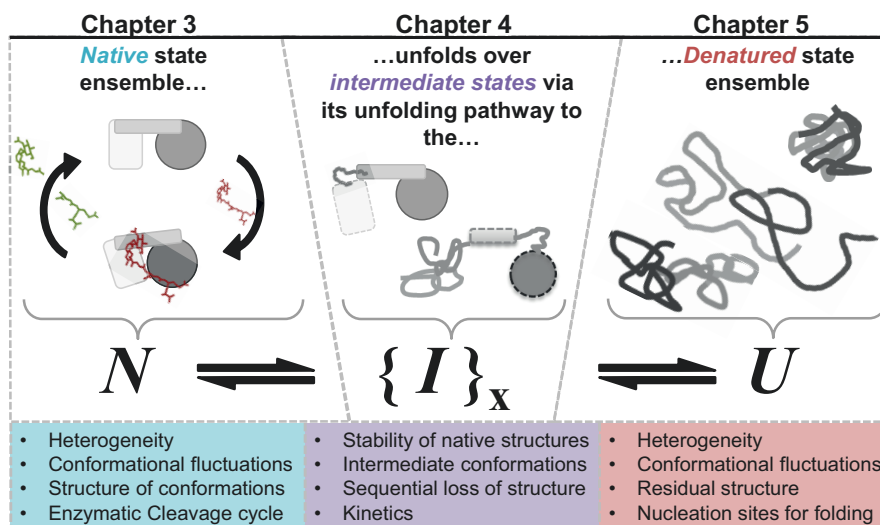


Figure 1.4.1 Thesis outline. First, the fluctuations between the different native states have been characterized and linked to the enzymatic cleavage cycle. Next, the unfolding of T4L by the denaturant urea via intermediate states was probed. Last, the ensemble of denatured states was characterized with respect to its heterogeneity and conformational fluctuations.

(1) Establishment of a workflow for a fluorescence spectroscopic toolkit

In the field of fluorescence spectroscopy a large variety of methods have been developed, many of such in the Seidel group (25, 46-49), which can be used singly or in combination to answer questions concerning the structure, dynamics or function of a biomolecule. The information obtainable can be increased by using FRET and/ or single-molecule spectroscopy. In this study selected methods were applied (Figure 1.4.2) and the first aim was to develop a workflow, with which (with minimal effort) a holistic understanding of the biomolecule's structure, function and dynamics could be obtained. Thus, more specific, which measurements do I need, how do I compare the results to another method and which combination of methods gives me the most information.

The methods applied within this study will be explained in more details in Chapter 2.

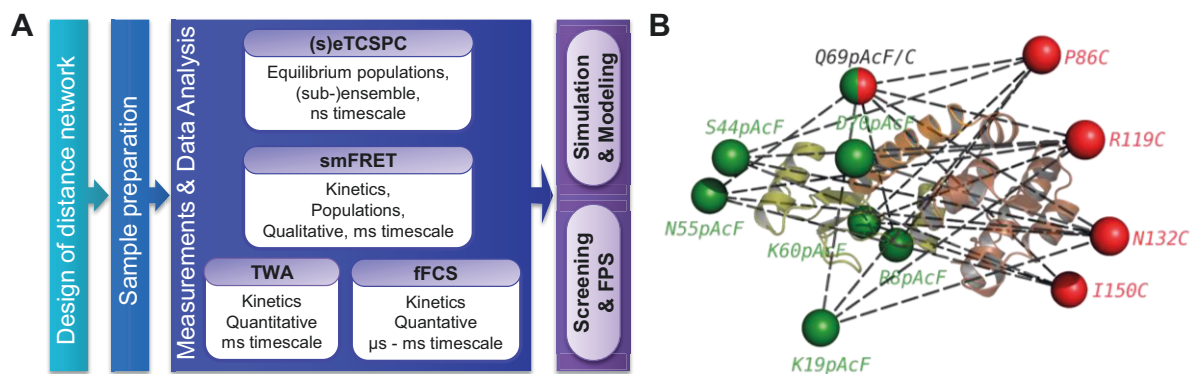


Figure 1.4.2 (A) Workflow of applied methods. (B) Network of 24 variants, which were used within this work.

(2) T4L under native conditions

It is well known that T4L exchanges between two conformations, an “open” and a “closed” state (34, 35), and also estimates for the exchange rate constants have been given (37). The goal in this interdisciplinary work was to generate first a set of 24 double mutants by simulation guided site-directed mutagenesis which could be labeled site-specifically with fluorescence dyes. The labeling positions were selected based on exemplary X-ray structures for the open (C_1) and closed conformations (C_2) by computational methods (Figure 1.3.1B). Secondly, ensemble (eTCSPC) and single molecule measurements (MFD) were used to generate a network of distance restraints for both conformers.

Additionally, two inactivating mutants that trap the enzyme at different steps in its cleavage reaction of peptidoglycan were prepared (Figure 1.3.1C). By a combination of ensemble and fluorescence correlation experiments (filteredFCS), the kinetics of the enzymatic cleavage were analyzed.

(3) Unfolding pathway of T4L

After the first hint of T4L not being a “simple” two-state folding protein in 1984, many studies have been undertaken to decipher its unfolding pathway and the structure of possible intermediates (20, 21, 40-44). Recently, native T4L (N) has been proven to unfold to its ensemble of unfolded states (U , DSE) via (at least) two intermediates I and J (20). However, for the methods used so far, NMR and stopped-flow CD, the first and last step of unfolding ($N \rightarrow I$, $J \rightarrow U$) are too fast to be probed. In contrast, the middle step ($I \rightarrow J$) occurs on the millisecond to second time scale and would thus be too slow to be able to probe with the spectroscopic tools used in this work.

In Chapter 4, the kinetics of the missing unfolding steps were characterized under equilibrium conditions with a combined fluorescence spectroscopic toolkit of several ensemble and single-molecule methods, which covers time scales which have not been accessible before in the conducted NMR and Stopped-flow CD measurements. By increasing the denaturant concentration gradually, we probed the stability of the native structures and observed the sequential loss of tertiary and secondary structure elements. Here, eight out of the 24 generated variants from the first part of this thesis were used.

Specifically, our goals were to examine the influence of an equilibrium of native states, containing also thermally excited states, on the unfolding pathway, to deduce the intermediate structures by probing the sequential, stepwise loss of structural elements and finally to

disentangle the fast folding-unfolding transitions, which are not accessible by conventional methods.

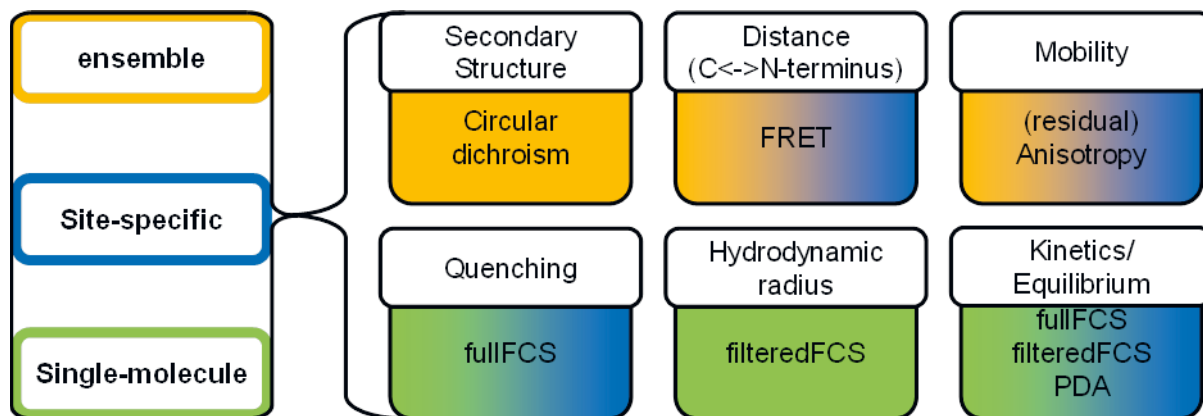


Figure 1.4.3 Optical microscopy tools

(4) The denatured state of T4L

The fact that denatured proteins may exhibit residual secondary structure or even tertiary interactions gets increasingly accepted within the field of structural biology (50). However, due the heterogeneous nature of the ensemble of denatured states (DSE) – rapid interconversion between many individual conformations – their characterization is a challenge (51).

Here, the full network of 24 variants generated in the first part of this thesis was exposed to high concentrations of denaturant (7.5 M urea). The DSE of T4L was then analyzed especially by means of its time-resolved distance distribution and inherent dynamics with ensemble and single-molecule fluorescence methods with the optical spectroscopy tools outlined in Figure 1.4.3. Thus, our goal was to describe the (polymer) behavior of T4L under highly denaturing conditions, to identify regions of residual structure (i.e. nucleation sites for folding) and to deduce structural features of the unfolded states. Additionally, the site-specific local mobility and global kinetics was analyzed to describe the rough energy landscape of the DSE.

2. Theoretical Background

2.1. Fluorescence

A molecule can be excited from its electronic ground state S_0 to an excited state S_1 - S_n by the absorption of a photon and within 10^{-13} to 10^{-11} s it relaxes back to the lowest excited state S_1 . This fast occurring process is called internal conversion. Relaxation from S_1 to S_0 , occurs more than 1000 fold slower and can happen either by emission of a photon (luminescence) or due to radiation-free processes, e. g. in form of vibrational energy (motion, collision with solvent molecules) (52). Luminescence can be further divided into two classes of processes, fluorescence and phosphorescence. In fluorescence, the photon is emitted directly from S_1 , while in phosphorescence the molecule first undergoes a transition from the singlet to the triplet state T_1 (intersystem crossing). For the photon to be emitted, the system has to return to the singlet state (forbidden spin reversion). The triplet state is thus a long living state and the emission of phosphorescence can take up to several minutes, whereas the fluorescence is generally emitted in the nanosecond range after excitation. The involved absorption and emission processes are summarized in the Jablonski-diagram (Figure 2.1.1).

Molecules emitting fluorescence typically used in experiments with biomolecules are named fluorophores and can be grouped in two classes: intrinsic/natural fluorophores, e. g. tryptophane, green fluorescent protein, and extrinsic/artificial fluorophores, e. g. Alexa488. The emitted light is of lower energy as compared to the absorbed light. Thus, the emission spectrum of a fluorophore is red shifted compared to its excitation spectrum.

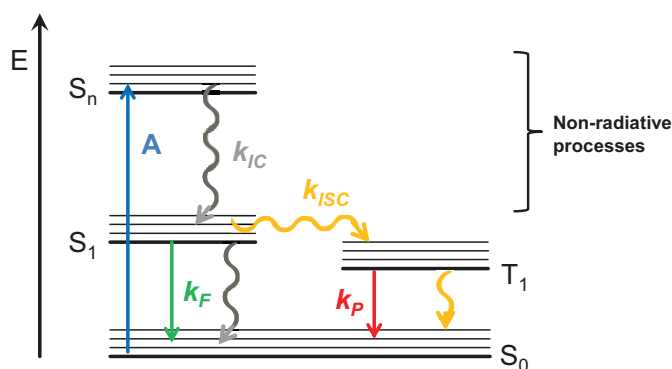


Figure 2.1.1 Jablonski Diagram. Possible relaxation processes with the rate constants k_x from the excited state S_1 into the ground state S_0 after absorption of a light photon (A, blue) of a fluorophore: Fluorescence (green, k_F) internal conversion (gray, k_{IC}) and phosphorescence (k_P , red) after intersystem crossing (k_{ISC} , yellow) took place.

Fluorescence is only one part of the total decay process from $S_1 \rightarrow S_0$. Its fraction is given by the fluorescence quantum yield Φ_F . Φ_F is calculated as the ratio of photons emitted by

fluorescence and the total number of absorbed photons and can be expressed in terms of the rate constants describing each of the competing deactivation processes (Eq. 2.1.1):

$$\Phi_F = \frac{\text{\# of emitted fluorescence photons}}{\text{\# of absorbed photons}} = \frac{k_F}{k_F + k_{IC} + k_{ISC}} \leq 1 \quad 2.1.1$$

2.1.1. Fluorescence Lifetime

The emission of fluorescence (photons) is a stochastic process and can be described as an exponential decay $F(t)$. The inverse of the characteristic decay time constant k_F is called the fluorescence lifetime τ of the respective fluorophore (Eq. 2.1.2).

$$F(t) = F(0)\exp(-t \cdot k_F) = F(0)\exp(-t / \tau) \quad 2.1.2$$

τ is defined as that time where the initial maximal intensity after excitation $F(0)$ dropped to $1/e$ of its original value. The fluorescence lifetime of a fluorophore depends on many factors, including the surrounding environment e. g. the solvent or vicinity to certain amino acid residues when coupled to a protein, which can quench the fluorescence lifetime. As the environment of a fluorophore coupled to a biomolecules is heterogenous, the fluorescence lifetime decay is thus no longer monoexponential but has to be described using a multi-exponential model (Eq. 2.1.3):

$$F(t) = \sum_i x_i \exp(-t / \tau_i) \quad 2.1.3$$

τ_i are fluorescence lifetimes associated with each (quenched or non-quenched) state and x_i the respective species fractions. The species-average fluorescence lifetime $\langle \tau \rangle_x$ and fluorescence-weighted lifetime $\langle \tau \rangle_f$ are calculated by Eq. 2.1.4a,b:

$$\langle \tau \rangle_x = \sum_i x_i \cdot \tau_i \quad \text{and} \quad \langle \tau \rangle_f = \frac{1}{\langle \tau \rangle_x} \sum_i x_i \cdot \tau_i^2. \quad 2.1.4a,b$$

The fluorescence lifetime is related to the fluorescence quantum yield according to the following formula (Eq. 2.1.5):

$$\frac{\tau}{\phi_f} = \frac{k_F + k_{IC} + k_{ISC}}{k_F \cdot (k_F + k_{IC} + k_{ISC})} = \frac{1}{k_F} \quad 2.1.5$$

However, in most cases, a reference sample is known and measured and the fluorescence quantum yield associated with the measured fluorescence lifetime is calculated by Eq. 2.1.6:

$$\frac{\langle \tau \rangle_{x,sample}}{\phi_{f,sample}} = \frac{\langle \tau \rangle_{x,reference}}{\phi_{f,reference}} \quad 2.1.6$$

2.1.2. Fluorescence Anisotropy

In general, the term anisotropy describes the directionality dependence of physical characteristics of a material. In fluorescence, this means the preferred interaction of fluorophores with light whose polarization is aligned parallel to the transition dipole moment of the electronic excitation. This results in the preferred excitation, i.e. a higher excitation probability, of those molecules in the isotropic mixture that are oriented in a certain angle to the incident, polarized light beam (photo selection). The emission of the fluorophore again occurs polarized, depending upon the transition dipole moment of the emission. Thus, the emitted light is depolarized with respect to the incident beam. This depolarization is described by the anisotropy r (Eq. 2.1.7):

$$r = \frac{F_{\parallel} - G \cdot F_{\perp}}{F_{\parallel} + G \cdot F_{\perp}} \quad \text{with} \quad G = \frac{F_{HV}}{F_{HH}} \quad 2.1.7$$

F_{\parallel} and F_{\perp} are the emission intensities measured parallel (0°) and perpendicular (90°) to the incident beam (0°) and G is a measurement device dependent factor which corrects for the instruments non-idealities, e. g. unequal detection efficiencies of differently polarized light. It is determined by measuring the fluorescence emission perpendicular (F_{HV} , 0°) and parallel (F_{HH} , 90°) to the incident beam after horizontal excitation (90°).

For static, non-rotating fluorophores, equation 2.1.8 yields the fundamental anisotropy r_0 – the maximum value this fluorophore can have - from which the angle between excitation and emission dipole β can be obtained:

$$r_0 = 0.4 \frac{3\cos^2(\beta) - 1}{5} \quad 2.1.8$$

Rotation of the fluorophores during their fluorescence lifetime leads to a decreased anisotropy, free small dye molecules of Rhodamin 110 in aqueous environment e. g. have an anisotropy close to 0, while fluorophores coupled to large biomolecules, e. g. proteins, have a hindered rotation and thus larger anisotropy. The maximal possible value is 0.4, the minimum -0.2. The anisotropy measured under this steady- state condition is termed r_{ss} or r_{∞} . However, fluorophores coupled to biomolecules via long and flexible linkers generally probe not only the slow rotation of the biomolecule but can also move freely, e. g. linker and side chain diffusion or dye rotation, to some extent. These processes can be quantified by measuring the anisotropy under time-resolved conditions. The anisotropy can then be described by the three exponential decay (53) in equation 2.1.9a. This can be simplified to equation 2.1.9b.

$$r(t) = r_0 \left[\frac{r_{fast}}{r_0} e^{-t/\rho_{fast}} + \frac{r_{slow}}{r_0} e^{-t/\rho_{slow}} + \frac{r_{\infty}}{r_0} \right] e^{-t/\rho_{global}} \quad 2.1.9a$$

$$r(t) = b_{protein} \exp(-t/\rho_{protein}) + b_{linker} \exp(-t/\rho_{linker}) + b_{dye} \exp(-t/\rho_{dye}) \quad 2.1.9b$$

ρ_i are the rotational correlation times associated with the fractions b_i . The sum of all b_i adds up to the fundamental anisotropy of the dye. From equation 10 it is evident that the rotation processes must occur in the same time scale as the fluorescence lifetime in order to be observed. It could be shown that the fluorescence anisotropy r for freely diffusing molecules is connected to the fluorescence lifetime by the Perrin-equation (54) (Eq. 2.1.10):

$$r = \frac{r_0}{1 + \frac{\tau}{\rho}} \quad 2.1.10$$

2.2. Förster Resonance Energy Transfer (FRET)

Förster Resonance Energy transfer (FRET) is the radiation free energy transfer of an excited donor fluorophore to an acceptor molecule. This results in a decreased fluorescence of the donor fluorescence and – if the acceptor is a fluorophore as well- the appearance of acceptor fluorescence. This energy transfer is highly dependent upon the distance and orientation between donor (D) and acceptor molecule (A) and was first described by Theodor Förster in 1948 (55). FRET relies on long-range dipole-dipole interactions between the donor and acceptor molecule and generally occurs between singlet states ($S_1, Donor$ to $S_0, Acceptor$). The main prerequisite for FRET is the spectral overlap of the donor emission spectrum with the acceptor absorption spectrum, quantified as the overlap integral J (Figure 2.2.1A) (Eq. 2.2.1):

$$J(\lambda) = \int_0^{\infty} F_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda \quad 2.2.1$$

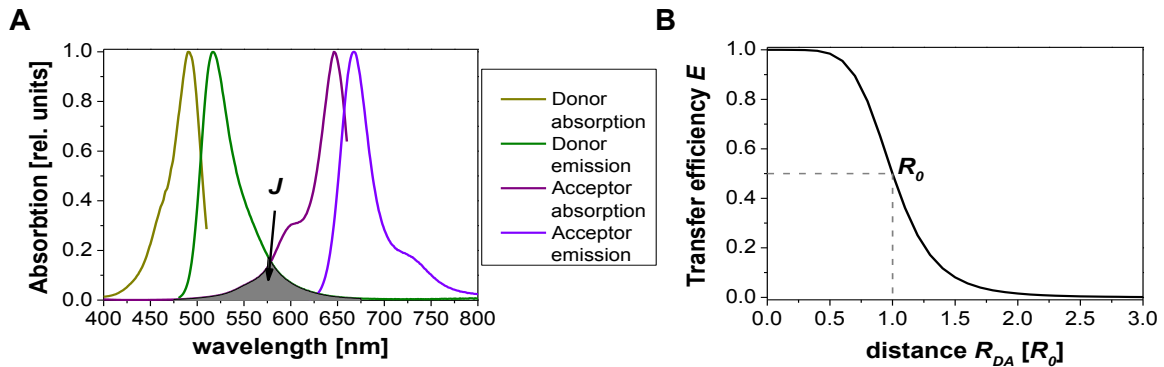


Figure 2.2.1 Overlap integral of Alexa488 and Alexa647 and distance dependence of the energy transfer. **(A)** Overlap integral J (gray) is the spectral overlap of the donor emission (green) and acceptor absorption (magenta) spectrum. **(B)** Distance dependence of the transfer efficiency in units of R_0 .

$F_D(\lambda)$ is the area-normalized donor emission and $\epsilon_A(\lambda)$ is the wavelength dependent extinction coefficient of the acceptor. A fluorophore pair, which is able to excerpt FRET, is called a FRET pair and the distance where 50 % of the energy is transferred to the acceptor molecule is defined as the Förster radius R_0 (Figure 2.2.1B) (Eq. 2.2.2):

$$R_0 = 0.2108 \cdot \left(\kappa^2 \cdot \Phi_{D(0)} \cdot J \cdot n_D^{-4} \right)^{1/6} [\text{\AA}] \quad 2.2.2$$

κ^2 is the orientation factor, $\Phi_{D(0)}$ is the fluorescence quantum yield of the donor in absence of acceptor and n_D is the refractive index of the surrounding media. When isotropic averaging of fluorophore motion during fluorescence emission can be assumed, κ^2 equals 2/3.

The energy transfer due to FRET E can be expressed analogous to the Φ_D as rate constant k_{FRET} (compare Eq. 2.1.1 and 2.1.5) (Eq. 2.2.3):

$$E = \frac{k_{FRET}}{k_{FRET} + k_F + k_{IC} + k_{ISC}} = \frac{k_{FRET}}{\tau^{-1} + k_{FRET}} \leq 1 \quad 2.2.3$$

and can be connected to the fluorescence lifetime ($\tau_{D(0)}, \tau_{DA}$), \sim intensities (F_D, F_A) or distance between the two fluorophores (R_{DA}) by the following relations (Eq. 2.2.4):

$$E = 1 - \frac{\tau_{DA}}{\tau_{D(0)}} = \frac{F_A}{F_D + F_A} = \frac{R_0^6}{R_0^6 + R_{DA}^6} \quad 2.2.4$$

2.3. Fluorescence spectroscopic methods

Many methods based on the characteristic properties of fluorescence have been developed to enable researchers to study the structure and function of biomolecules over a wide range of time scales from nanoseconds up to seconds (Figure 2.4.1).

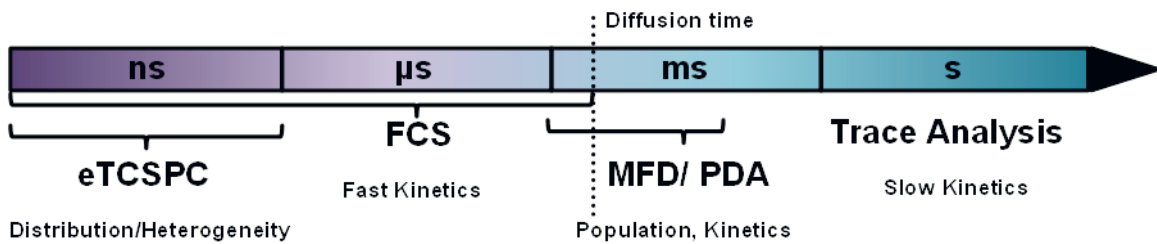


Figure 2.3.1 Timescales observable with different fluorescence methods. Various distinct fluorescence techniques can be applied to monitor biomolecular processes on different time scales with different time resolution (Modified from (25)). Details to the methods used in this work can be found in the respective section of this chapter.

Analysis of time-resolved fluorescence lifetime measurements ((s)eTCSPC) gives information about the sample heterogeneity within the nanosecond time scale (\sim fluorescence lifetime). Dynamics in the μ s to ms time range is observed with fluorescence correlation

methods (FCS) like full FCS or filtered FCS. Multiparameter Fluorescence Detection (MFD) reveals the different populations present in the sample whose kinetics can be analyzed further using Photon Distribution Analysis (PDA). For slower kinetics on time scales in the second to minutes range, the trace analysis can be used. Except the trace analysis all above-mentioned methods were used within the course of this work and will be described in more details below.

2.4. Single Molecule Multiparameter Fluorescence Detection (MFD)

Single-molecule fluorescence measurements offer the great advantage that only a single, unique molecule and its characteristic behavior, e. g. in terms of FRET-efficiency, donor fluorescence lifetime in presence of acceptor or anisotropy, can be observed at a time (56). This means that the average parameter detected in the ensemble measurement in a cuvette can be dissolved into its complete broadness of distribution and frequency of single values – displayed in a frequency histogram (57). The single-molecule experiments performed in the course of this work are based on the so-called Burst-Integrated Fluorescence Lifetime (BIFL) method developed by Keller (58). This method was further advanced within the group (49, 59-61) and the measurements were conducted on a home-built eight channel multiparameter fluorescence detection (MFD) setup (Figure 2.4.1).

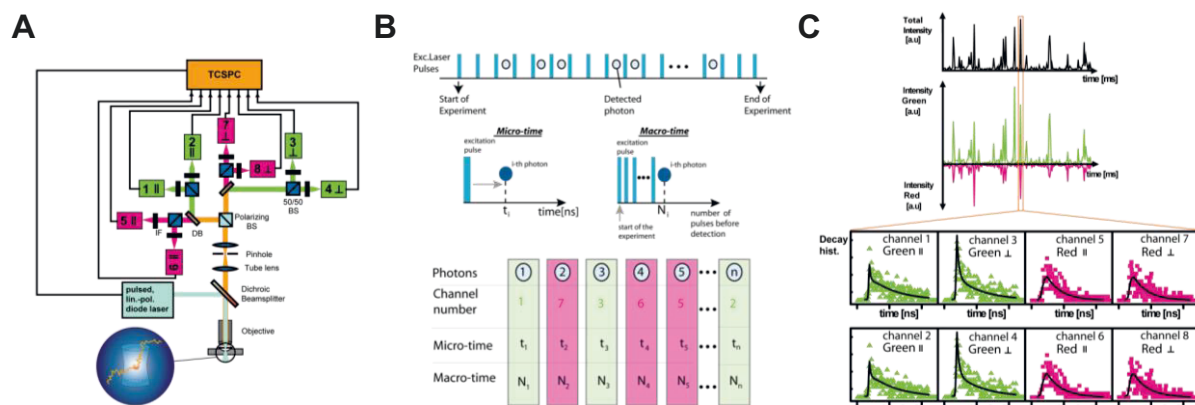


Figure 2.4.1 Experimental MFD setup and data acquisition. (A) A modified confocal microscope optimized for high-resolution single molecule spectroscopy. Instead of a typical MFD setup which consists in general of four detectors, we use a total of eight detectors connected to time correlated single photon counting electronics to assure dead-time free detection of photons. (B) In time correlated single photon counting, each photon is identified by three parameters: Micro-time or time after excitation pulse, Macro-time or number of excitation pulses from the start of the experiment and channel number. These three parameters are needed for further analysis. (C) Single molecules diffuse freely through the confocal volume. Photons are emitted leaving a burst of photons as a function of time. Each burst is selected and fitted accordingly.

The pM concentration of double-labeled protein samples are applied onto a confocal microscope and single molecules diffusing through the focus are excited with a pulsed, linear

polarized laser. After being excited, the molecules emit a short “burst” of fluorescence photons which are being collected and focused through a tube lens onto the pinhole ($\sim 100 \mu\text{m}$ in diameter). Subsequently, the photons are firstly split 50/50 onto both halves of the setup, then by color (“green” or “red”) through band pass filters and finally by polarization (“parallel” or “perpendicular”) due to polarizing beam splitters before they are allowed to hit the detectors. The signal from all detectors are connected to time-correlated single photon counting boards which store three parameters for each detected photon: (I) Micro-time or time after excitation pulse, (II) Macro-time or number of excitation pulses from the start of the experiment and (III) channel number (49). Bursts are then identified from background using software developed in the Seidel group (“Paris”) and fitted with the Maximum Likelihood Estimator (62).

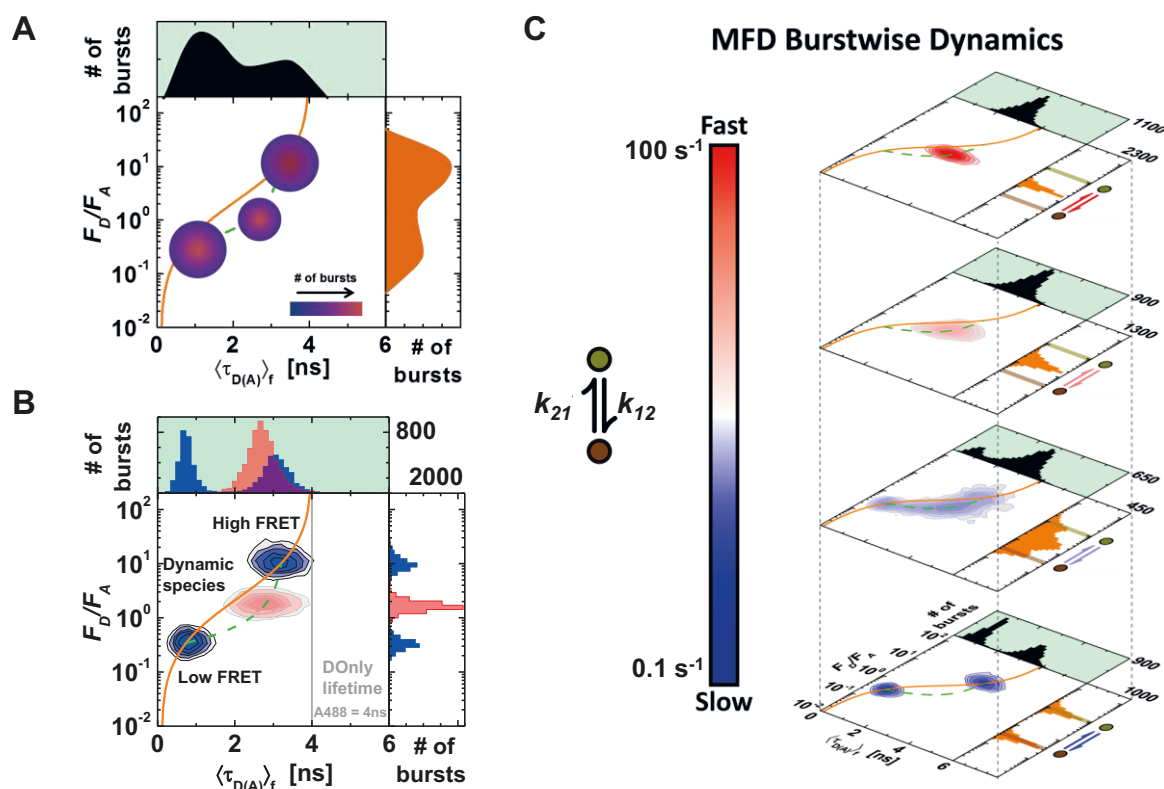


Figure 2.4.2 Schematic representations of MFD and dynamics. (A) Schematic diagram of a sample containing three species. In the 1D histograms on the top and side only the distribution around two mean values is observable. (B) In the 2D plot of the two fluorescence parameters donor fluorescence lifetime in presence of acceptor ($\langle \tau_{D(A)} \rangle_f$) and ratio of donor over acceptor fluorescence (F_D/F_A), static species are located on the so-called static FRET line (orange), whereas dynamic species are shifted towards the right, onto the dynamic FRET line (green dashes) which connects the two “border” states of the dynamic mixture. DOnly labeled species have a fluorescence lifetime of $\sim 4 \text{ ns}$ for the fluorophore Alexa488. The 2D plot was generated with Margarita and Origin 8.6. (C) Schematic visualization of the influence of dynamics on the shape and position of the 2D frequency histogram. Slow exchange, e.g. no exchange during burst duration ($\sim 1 \text{ ms}$) will lead to two separate species being visible, while an increasing exchange rate constant leads first to “smearing” between the two border states and finally the merge into only one population. Data of (B) and picture (C) is courtesy of Hugo Sanabria (University Clemson, USA).

Selected bursts can then be visualized in 2D frequency histograms (“MFD-plots”), where two parameter are plotted against each other, and further analyzed population specific (25). The setup of the MFD experiment allows obtaining many fluorescence reporters of a sample simultaneously, e.g. donor and acceptor fluorescence lifetime and anisotropy or the green-to-red signal ratio from which the FRET efficiency (or distance between fluorophores) can be inferred. MFD-plots offer both the advantage that otherwise potentially invisible populations get visible (Figure 2.4.2A) and that the dynamics of the system can be seen and its time scales estimated (Figure 2.4.2C). Depending upon the exchange rate constant between the two species shown in Figure 2.4.2B, the dynamic FRET population is to a greater/smaller extent elongated, and depending upon the equilibrium constant, the dynamic population is shifted towards the low/high FRET population (63).

2.5. Photon Distribution Analysis (PDA)

The quantitative analysis of the above described “by eye” recognizable dynamics in the MFD diagrams is done with Photon Distribution Analysis (PDA) (48, 63, 64). Under consideration of background count rate, spectral crosstalk and experimental shot noise (all Poissonian distributed), the minimal width of the experimentally acquired histogram is determined.

The resulting, best-fitting distance between the fluorophores is characterized by a (or several) Gaussian distribution with mean distance $\langle R_{DA} \rangle_E$ and apparent half-width h_{WDA} . It is important to note here that the half-width of distribution determined here cannot be compared with the width determined by the ps- resolution of ensemble measurements due to the application of burst-wise averaging in PDA: Additional broadening of the experimental histogram when compared to the theoretical determined width occurs due to sample heterogeneity and/or the dynamic exchange between more than one species in the sample. The effect of dynamics on the ratio of green to red signal S_g/S_r between two species is depicted schematically in Figure 2.5.1A using the same data as for Figure 2.4.2. If the exchange between two species is slower than the diffusion time of the molecules through the focus, two peaks will be visible - one for each species. If the exchange rate constant increase, firstly a “smear” between the two border peaks will appear and secondly they will become broader and move closer to each other before finally merging in one single, broad peak. If the exchange rate constant is even faster, only one narrow peak is recognizable.

The dynamic exchange is analyzed by an extension of the PDA theory: dynamic PDA (65). Here, the two border species $\langle R_{DA} \rangle_{E,1}$ and $\langle R_{DA} \rangle_{E,2}$ are modeled as Gaussian distributions

which exchange with rate constants k_{12} and k_{21} . The equilibrium constant K is the ratio k_{12}/k_{21} . For analysis, the data is generally analyzed globally integrating (or “binning”) of the experimental photon bursts in different time-intervals or time-windows (TW’s). If the dynamic exchange occurs within the binning of the time window, only one peak appears in the histogram (Figure 2.5.1B). Decreasing the bin width from 1 ms to 0.3 ms instead shows the two border states between which the exchange takes place (Figure 2.5.1C).

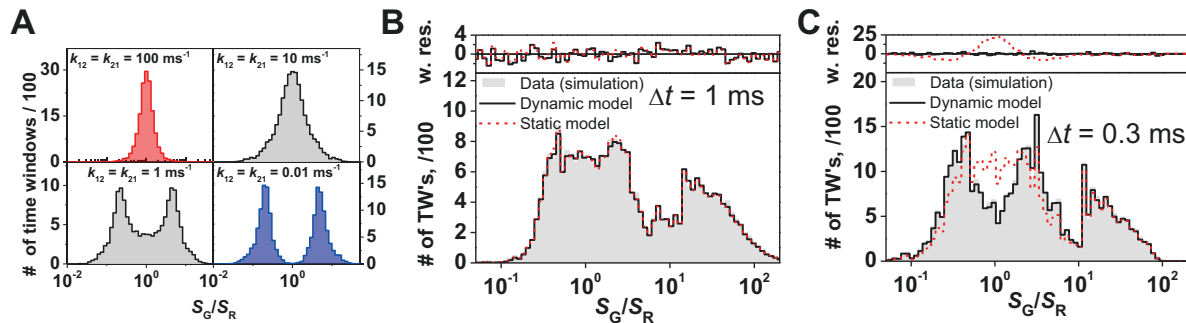


Figure 2.5.1 Schematic representation PDA and dynamics. (A) Simulation of a dynamic system with two different states which exchange with distinct exchange rate constants k_{12} and k_{21} . Time window is 1 ms. (B-C) Global analysis of simulated S_G/S_R histograms with two different time windows, 1 ms and 0.3 ms. (B) can be described well with both the dynamic and the static model, whereas global fitting of both time windows is only possible when using a model including dynamics (C). Data and picture is courtesy of Thomas Peulen.

Dynamic PDA can be applied to detect and describe the dynamics of a system if the time scale of its dynamics is on the order of the diffusion time of the molecules through the confocal volume, i. e. in the low millisecond range. Averaging photons over time scales much longer would lead to the possibility of merging the photons from two different single molecule events in one time window and thus distort the burst specific information. On the other hand, time windows cannot be decreased arbitrary as a minimum of photon is needed for a reliable analysis.

2.6. Fluorescence Correlation Spectroscopy (FCS)

In contrast to PDA, fluorescence correlation spectroscopy (FCS) can be used to detect and quantify dynamics of a system occurring faster than the diffusion time of a molecule down to the nanosecond range limited only by the fluorescence lifetime of the dye. FCS uses the fluctuations in the fluorescence signal (deviations from the time-averaged mean signal) occurring when low concentrations of labeled molecules diffuse freely through the focus of a microscope (66, 67). The used concentration range is ~ 1 -10 molecules in the focus at a time. A standard FCS curve is shown in Figure 2.6.1A; its amplitude is determined by the inverse of the number of particles N in solution and the characteristic decay time (inflection point of the curve) is given by the average diffusion time t_d , i.e. the time range molecules spent in

average inside the detection focus. At time scales much longer than diffusion the correlation curve decays to the baseline, i. e. the fluorescence signal/fluctuations, are no longer correlated. In general, all processes of the biomolecule, which lead to a fluctuation in the fluorescence signal, and occur at time scales shorter than diffusion and longer than the fluorescence lifetime of the dye can be detected and lead to the appearance of additional correlating terms (Figure 2.6.1B). For double-labeled variants, the signals from both the donor and the acceptor channel can be correlated (autocorrelation) and additionally the donor signal can be correlated with the acceptor channel (crosscorrelation) (68).

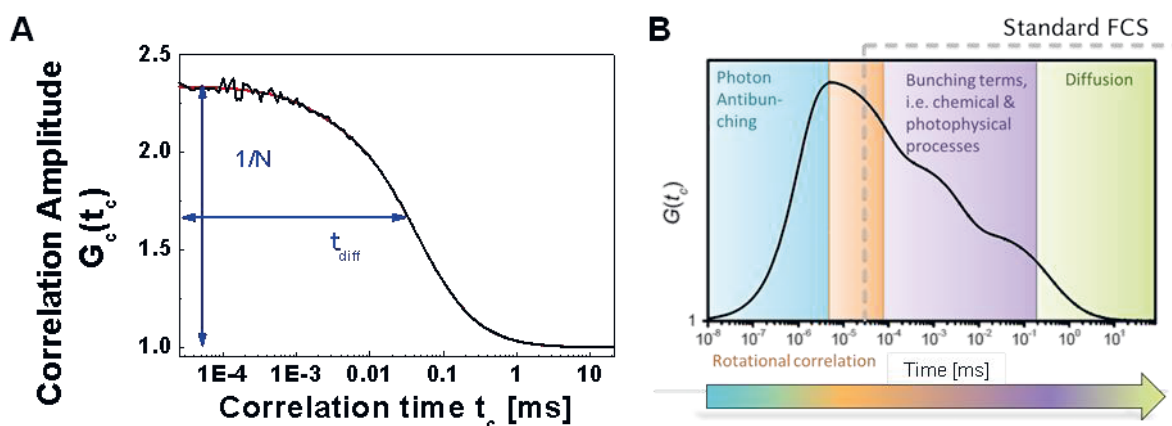


Figure 2.6.1 Fluorescence correlation curves. (A) When using a logarithmic time scale, the diffusion time of a molecule can be easily identified as the inflection point of the curve. The number of particles in the focus is inversely proportional to the amplitude at $\sim 50 \mu s$. (B) When the dyes exhibit photophysical processes and/ or additional protein specific fluorescence fluctuations due enzymatic reactions occur, this leads to the appearance of additional bunching terms. In high time-resolved measurement down to the ps time scale, also the photon antibunching can be observed.

Thus, here also FRET-induced processes can be monitored, which appear as anti-correlation terms in the correlation curves. However, in these cases it is important to measure also single labeled biomolecules to be able to differentiate dye or simple protein motions from FRET induced processes. Processes appearing in FCS curves in the μs to low ms time scale are e.g. domain motions occurring in enzymatic reactions due to rearrangements of the active site. In the low μs time, for autocorrelation curves of single labeled variants the triplet state formation and kinetics of the dye used can be observed ($\sim 5-10\%$ of total amplitude, depending upon the dye). When having a setup with polarized detection, also the rotational correlation time (time the molecule needs on average to rotate for one radian) of the molecule t_r can be observed (Figure 2.6.1B). t_r strongly depends upon the size of the molecule (~ 20 ns for small proteins) and is easily identified by its polarization depend amplitude. In high time-resolved setups with ps resolution, the photon antibunching term, i.e. the rapid decline of the correlation curve at time scales of the fluorescence lifetime of the dye can be observed (69).

A special variety of FCS is the *filtered* FCS method (47, 59, 70-72). Here, no longer channel/color specific fluctuations in the fluorescence signal are correlated. Instead “filter patterns” from two different species found in the single-molecule measurements are generated (Figure 2.6.2). These patterns are computed from burst ID’s and thus contain e.g. information about the fluorescence lifetime decay and time-resolved anisotropy decay in donor and acceptor channels. Imposing those filters while correlating, leads to a selective enhancement of transitions occurring between those two species while all pure dye related processes are suppressed (they are occurring in both species).

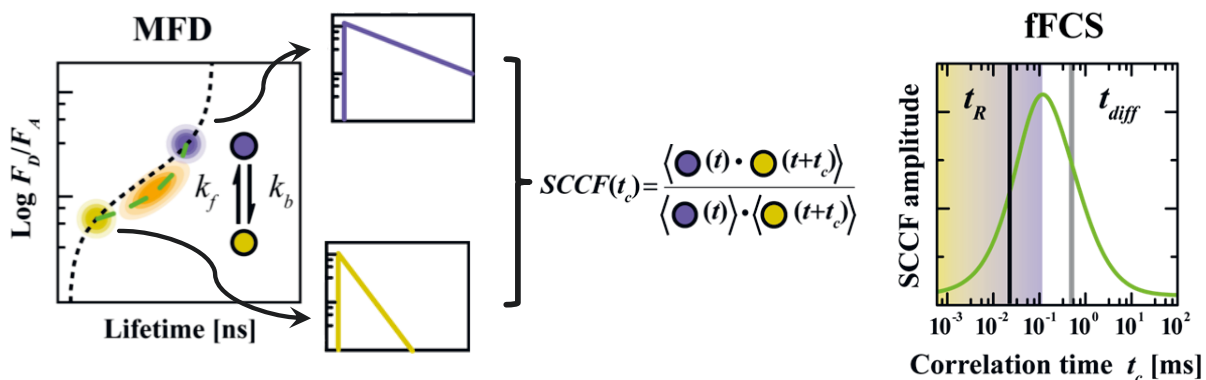


Figure 2.6.2 Schematic representation filtered FCS. Two species are identified in single molecule measurements and their burst ID’s containing e. g. the information about color-specific fluorescence lifetime and anisotropy are used to generate filters for correlation. The obtained relaxation time t_r is the inverse of the sum of the two reaction rate constants k_{12} and k_{21} . sCCF is the species-selective cross-correlation function.

2.7. Accessible Volume Simulation (AVSim)

If structures of the examined proteins have either been deposited to the PDB from previous X-ray or NMR experiments or a computed homology model is available, the expected FRET measurable distances can be estimated beforehand. This is especially useful to determine optimal labeling positions such that firstly, the distances between the donor and acceptor lie within the FRET- measurable range and secondly, the suspected motions/structural rearrangement of the biomolecules during e.g. an enzymatic reaction can be captured best. However, most fluorophores which are used for labeling are attached either at amino acid or nucleic acid site chains and contain additionally a flexible $(CH_2)_x$ -linker, which serves as a spacer between the surface of the biomolecule and the fluorophore. Instead of a single fixed measurable distance, thus, both dyes can explore a certain space volume – limited only by its linker length and steric hindrance of the biomolecules’ surface.

To take this free movement of the dyes into account, an “Accessible Volume Simulation” (AVSim) can be performed (53, 73). Here, the dyes are modeled as spheres with dye radius R_{dye} and the linker being a flexible cylinder of length L_{link} and width w_{link} (Figure 2.7.1).

In a geometrical algorithm all positions of the dye allowed by its dimensions and biomolecule surface are considered equally probable. The distances between random positions of donor and acceptor dye are collected and recalculated into experimental-like distance according to three methods. The distance between the mean positions of the donor and acceptor R_{mp} is not accessible by measurements (Eq. 2.7.1):

$$R_{mp} = \left| \left\langle \overrightarrow{R_D^{(i)}} \right\rangle - \left\langle \overrightarrow{R_A^{(j)}} \right\rangle \right| = \left| \frac{1}{n} \sum_{i=1}^n \overrightarrow{R_D^{(i)}} - \frac{1}{m} \sum_{j=1}^m \overrightarrow{R_A^{(j)}} \right| \quad 2.7.1$$

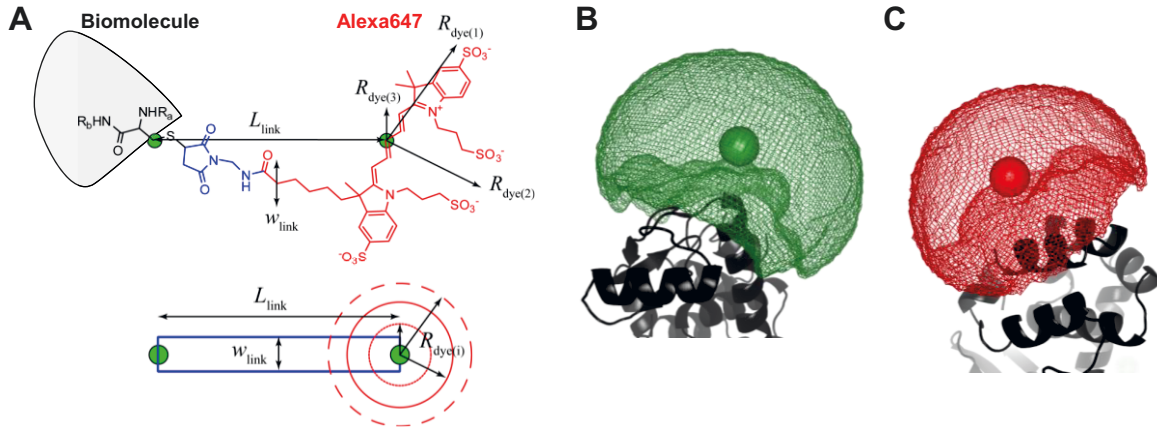


Figure 2.7.1 Schematic representation of a dye approximation with AVSim. (A) Structure of Alexa647-maleimide coupled to a biomolecule and the simplification used within AVSim. Fluorophores are approximated by a sphere with a defined radius $R_{dye(i)}$, a connecting linker with length L_{link} and width w_{Link} . Picture is modified from Sanabria *et al.* (In preparation) (B, C) AVSim for Alexa488-hydroxylamine at position 60 in T4L (B) and A647-maleimide at position 132 (C). The dye accessible space volume is shown as mesh, the mean dye positions as a sphere, created with PyMol 1.3 and calculated with FPS.

The mean donor-acceptor distance $\langle R_{DA} \rangle$ (Eq. 2.7.2) is obtained in time-resolved ensemble measurements with ns resolution, whereas in intensity based methods with limited time resolution, e.g. single-molecule MFD measurements only the FRET- efficiency weighted average distance $\langle R_{DA} \rangle_E$ (Eq. 2.7.3) can be determined:

$$\langle R_{DA} \rangle = \left| \left\langle \overrightarrow{R_D^{(i)}} - \overrightarrow{R_A^{(j)}} \right\rangle \right| = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \left| \overrightarrow{R_D^{(i)}} - \overrightarrow{R_A^{(j)}} \right| \quad 2.7.2$$

$$\langle R_{DA} \rangle_E = R_0 \left(\langle E \rangle^{-1} - 1 \right)^{1/6} \quad 2.7.3$$

The average efficiency $\langle E \rangle$ is defined as (Eq. 2.7.4):

$$\langle E \rangle = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \left(\frac{R_0}{R_0 + \left| \overrightarrow{R_D^{(i)}} - \overrightarrow{R_A^{(j)}} \right|^6} \right) \quad 2.7.4$$

3. Structure and function of T4L under native conditions

3.1. Dynamics & function of transient states in single enzyme molecules

The manuscript is based on collaborative work in a joint project of the 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: HS and **KH** purified and labeled the protein. HS, TP, **KH**, and DR measured and analyzed the FRET experiments. TP, MD and HG performed structural screening. MRF performed and analyzed EPR experiments. WH performed study design and EPR analysis. SF developed fluorescence analysis tools. FK and RK developed fluorescence instrumentation and gave technical support. All authors discussed the results and commented on the manuscript. **KH**, HS, WH and CAMS wrote the paper. CAMS supervised the project

HS: 20 %

DR: 20 %

KH: 20 %

TP: 12 %

MRF: 2 %

SF: 4 %

MD: 3 %

FK: 2 %

RK: 3 %

WH: 3 %

HG: 4 %

CAMS: 7 %

3.2. Manuscript

Abstract

To understand enzyme function it is key to visualize transiently populated conformational states and their corresponding exchange pathways. Here, we assess the conformational landscape of three enzyme states of T4 Lysozyme (T4L) (defined by substrate interaction) over the ns-ms timescale. To accomplish this, we introduce a hybrid approach involving FRET-based multiparameter fluorescence, EPR spectroscopy, mutagenesis, and FRET-positioning and screening. To build a model of enzyme function, we use a state matrix of three enzyme states and an equilibrium of three short-lived conformational states (open, closed, and compact) sampled at 4 μ s and 230 μ s. The 230 μ s bending motion has not been identified before and leads to a more compact state. After screening over 380 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 178L, for the closed state 1LWK, and no structure could be correlated with the new state. Furthermore, the closed state 1LWK differs by only 0.5 Å RMSD from the adduct form of T4L (148L). Moreover, the observed changes in equilibrium between transiently populated conformational states in variants that mimic the catalytic states suggest that the new 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. 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 could have an impact on understanding many non-ATP/GTP-hydrolyzing enzymes where stochastic product release is expected.

3.2.1. Introduction

Enzymes are biological catalysts that accelerate chemical reactions. During catalysis, enzymes sample defined structural conformations (74) including low populated states over a wide temporal domain (4); their structure and function are linked by conformational transitions (75) and some transient states have been linked to diseases (76). During the last decade, various experimental and theoretical tools were introduced to study the enzymes' dynamic behavior (4, 77), and our initial 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 remain hidden to classical methods used in structural biology. All spectroscopic techniques such as nuclear magnetic resonance (NMR) (78), electron paramagnetic resonance (EPR) (79), FRET (80), and lower resolution methods (4), 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, it is usually not ideal for evaluating large scale biomolecular dynamics. NMR spectroscopy provides detailed equilibrium dynamic information but 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 (81). 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.

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 3.2.1A, uses fluorescence lifetime measurements (ensemble Time Correlated Single Photon Counting; eTCSPC), confocal Multiparameter Fluorescence Detection (MFD) of single diffusing proteins (63), and filtered fluorescence correlation spectroscopy (fFCS; see 3.3.1.12 and (47, 71)). Moreover, we corroborated our results by the use of other biophysical tools, such as site-directed mutagenesis, chromatography, and EPR spectroscopy (79).

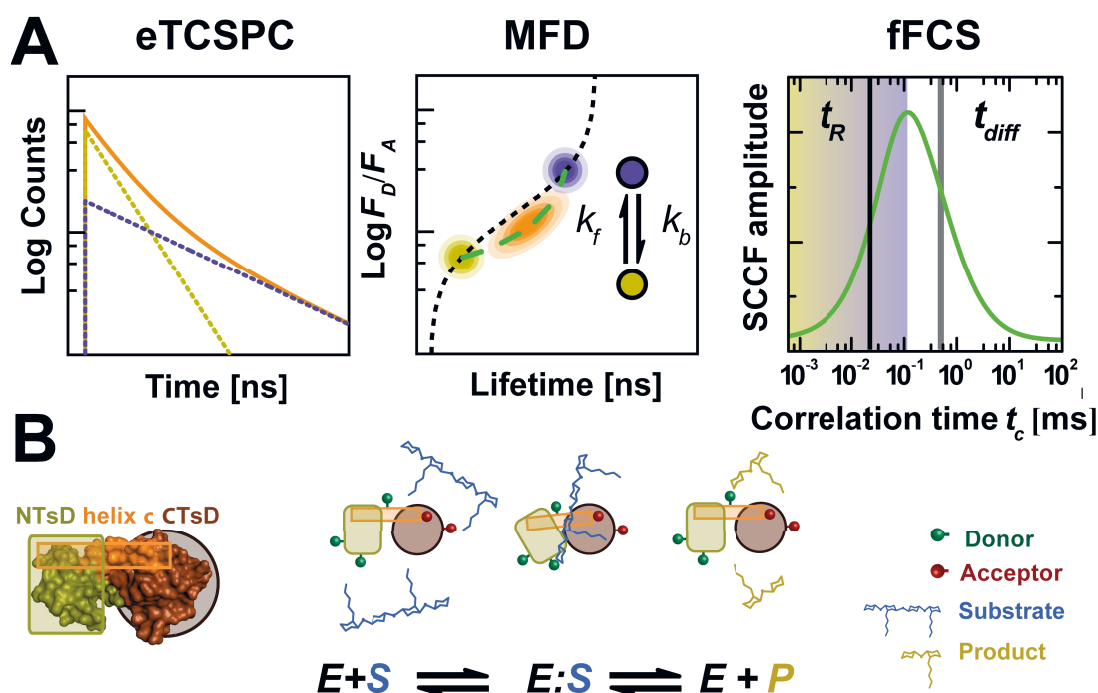


Figure 3.2.1 (A) Schematic of the high precision FRET and fluctuation analysis toolkit. eTCSPC resolves stable populations on the ns timescale. MFD experiments of freely diffusing single molecules are analyzed by the FRET indicator F_D/F_A (ratio of donor over acceptor fluorescence) accumulated during 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. 3.3.22). 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 (35, 82). This change in closure between the two lobes can be followed by the induced change in FRET.

In this study, we examine the hinge-bending dynamics of the bacteriophage T4 lysozyme (T4L) and determine the implications of this process to its enzymatic 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 (35). Structurally, T4L (33) consists of two interrelated subdomains (N-terminal subdomain, NTsD, and C-terminal subdomain, CTsD), which differ in their folding behavior and stability (42). The long α -helix (Helix c) mediates communication between the subdomains (Figure 3.2.1B). There are now more than 500 entries in the Protein Data Bank (PDB) after the original structure (PDB ID: 1LZM; (33)). 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 (34, 82). When substrate is bound, T4L adopts a closed conformation C_2 (35). 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 (Figure 3.2.1B) (37). Here, typically the product is expected to dissociate stochastically from enzymes. Recently, Lu and coworkers (83) proposed for Horse radish peroxidase that the product may be actively spilled out in the presence of strong enzyme-product interactions (83). Therefore, a modified

Michaelis-Menten model for enzyme kinetics with the inclusion of a product release state is debated (84, 85). Furthermore, 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 (37). Various computational simulations estimated a nanosecond hinge bending motion (86-88). NMR studies (82) suggested that inter-domain closure occurs faster than 170 μ s. EPR (79) and fluorescence correlation spectroscopy (37) 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 (89). From experiments on enzyme catalysis under various conditions (turbidity, FRET, and carbon nanotube currents (32, 90-92)), 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 and single enzyme nanotube-coupled experiments suggested that T4L exhibits complex conformational motions beyond the hinge-bending dynamics (92, 93).

To resolve how enzyme dynamics relates to function, we map the conformational landscape of T4L by using the above hybrid approach (Figure 3.2.1A). Our findings show a hidden third conformational state whose function in the enzymatic cycle is discussed. To support our results, we did a systematic study of 24 doubly labeled variants capturing all possible motions of the T4L backbone (Figure 3.2.2A) and applied our toolkit with three independent methods, which allowed us to resolve all conformational states, their equilibrium fractions, and probe their time evolution. To derive the energy landscape of T4L's cleavage cycle, mutations affecting two active site residues were introduced to capture the enzyme at various steps during substrate hydrolysis.

3.2.2. 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 3.3.1). 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 (94, 95) (Figure 3.2.2A). This strategy allowed us to address 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 chose representative structures for the known open and closed conformations of T4L (PDB ID: 172L and 148L, respectively), and used FPS (FRET positioning and screening) to add FRET label *in silico* and compute the expected experimental interdye distances (see methods 3.2.4 and 3.3.1.15). The mean interdye distance $\langle R_{DA} \rangle$ is directly measured from time-resolved FRET experiments (eTCSPC). The treatment of the directionality of the fluorophores is essential for proper structural interpretation. Hence, $\langle R_{DA} \rangle$ must not be confused with the backbone ($C\alpha$ - $C\alpha$) distance (Table 3.3.1D, E). To obtain a suitable conversion between $\langle R_{DA} \rangle$ and the backbone distance an appropriate structural model is required (53) as shown in section 3.3.1.13 and 3.3.1.15. Based on our initial network design, the hinge-bending motion should lead to a two-states 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 3.3.1) 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 the *open* and *closed* conformations were expected to yield the same distance. In this case, the data for a two states system would yield a single FRET quenching term of the donor. Thus, we have two classes of variants: (1) *open* and *closed* cannot be distinguished; (2) *open* and *closed* 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 (10) and to test if 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 a 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 3.2.1A and 3.3.1.12). 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.

3.2.2.1. Free Enzyme

Detection of a third conformational state by eTCSPC. To analyze the conformational ensembles of free T4L with high time resolution and precision (see 3.3.1.6), we performed eTCSPC measurements using an instrument with comparably narrow instrument response function (230 ps), a small time bin width (14.1 ps/ch), a very high number of collected photons (30×10^6 photons), and an additional instrument linearization procedure. We systematically tested our conclusions by statistical and global analysis using five different fit models of increasing complexity to demonstrate that a detailed description of the system is essential for a unified and meaningful description of all eTCSPC data (3.3.1.11). To minimize the number of fit parameters, we analyzed the data set of each variant (i.e. the decay of the D-only- and DA-labeled sample) by global fits (eq. 3.3.10, 3.3.13 and 3.3.1.11), which allow us to determine the donor fluorescence lifetimes and to extract the additional donor quenching by FRET. The quenching of the donor by FRET was described by distributions of FRET-rate constants (eq. 3.3.14) and the fluorescence decays of the donors in the absence of FRET (necessary reference samples) by multi-exponential decays (Table 3.3.1). Typical fluorescence decay curves of a DA- and the corresponding D-only-sample of the variant K19pAcF/R119C are shown in Figure 3.2.2B. In the first fit, we tried to describe the FRET related quenching by a single FRET-rate constant, which is formally expressed as single apparent distance (R_{DA}). This failed even for mutants where X-ray structures of the two known *open* and *closed* conformers predict comparable distances (Table 3.3.1B, C).

Next, we used models with two and three FRET-rate constants, respectively (Table 3.3.1B, C). Not surprisingly, we find better agreement using three FRET-rate constants. However, by inspecting all FRET-pairs for global consistency, clear features emerged that support the need for a third state. All variants, including those that were designed as controls with comparable distances for the *open* and *closed* conformation, needed at least two FRET-rate constants. To test whether the improvement in χ^2_r of the three FRET-rate constant model is statistically significant, we postulated the null hypothesis that a two FRET-rate constant 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 $p(2vs3)$. By plotting $p(2vs3)$ against the fitted distance difference ($\Delta R = |R_{DA}^{(1)} - R_{DA}^{(2)}|$), obtained by the three

FRET-rate constant model, we recognized the existence of three classes of variants of distinct statistical significance values: (1) $R_{DA}^{(1)}$ and $R_{DA}^{(2)}$ are indistinguishable (Figure 3.2.2C, black points); (2) $R_{DA}^{(1)}$ and $R_{DA}^{(2)}$ are distinguishable, but the distance difference is small ($\Delta R < 5$ Å, blue points); (3) all three distances 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 (3.3.2.1, Figure 3.3.2B) is clearly smaller than the F-value of our data (1.046); thus, we need at least three FRET-rate constants to describe our data.

Both donor and acceptor fluorophores are coupled via flexible linkers to T4L. Due to this conformational flexibility of the dyes (this assumption is verified by corresponding anisotropy measurements of each dye, see Table 3.3.3), distance distributions should be considered in the fluorescence decay analysis. Thus, for a later quantitative comparison of the known X-ray structures to the experimental data, we additionally analyzed our experimental fluorescence decays by distance distributions, which consider the distance broadening caused by the flexible dye-linkers. Initially, we used in the fourth fit a two state model with independent species fractions and compared the recovered distances to known limiting structural states (see Figure 3.3.2). This comparison resulted in an absolutely inconsistent assignment to the two known conformers (Figure 3.3.2C, D). Given this discrepancy, we decided to describe all decays with a three states model (Table 3.3.1D, E) in the fifth fit. Here, instead of a manual assignment of recovered distances to states, we performed a joint analysis of all 24 FRET experiments and their donor-reference measurements, in which the species fraction of the states were linked over the whole data set. By this unbiased analysis, we obtained a “major” (C_1), “middle” (C_2) and “minor” (C_3) population (0.44, 0.32 and 0.24 species fraction, respectively) that directly defines the corresponding distance sets (Table 3.3.1D, E). To conclude, only the fifth fit model taking conformational dye flexibility into account and considering the correct number of conformational states allows for a good and meaningful joint fit of all 24 FRET samples and their corresponding reference samples. The globally linked amplitudes stabilize the joint fit dramatically by reducing the number of free parameters and allow for an unbiased (automatic) grouping of the distances.

Based on structural pre-knowledge we expected only two states. Hence, our minor state, C_3 , can be considered a “hidden” state, because it has eluded identification based on other structural methods.

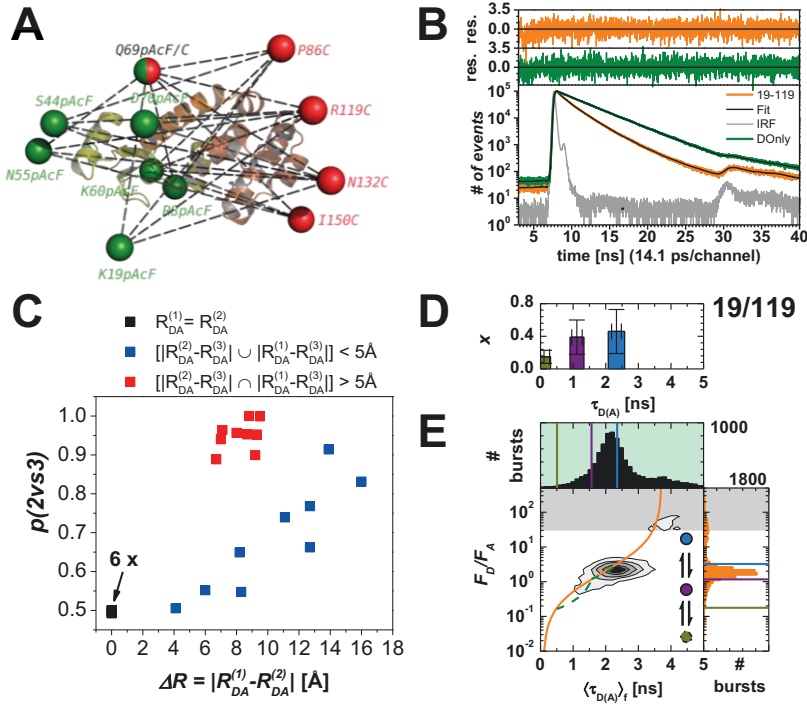


Figure 3.2.2 Uncovering a third conformational state. (A) Distance network of 24 DA FRET pairs. At donor positions (green), the inserted unnatural amino acid pAcF was labeled with Alexa488-hydroxylamine; the cysteine at the acceptor positions (red) was labeled with Alexa647-maleimide. Position 69 was used as both donor and acceptor position. (B) Time-resolved fluorescence decay for K19pAcF/R119C-(DA) (orange) and its respective D-only-decay (green) fitted with a three states model (black line), respectively. We determined the following values for R_{DAi} 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 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 $\chi^2_{rel} = \chi^2(2\text{ states}) / \chi^2(3\text{ states})$ is plotted against the absolute difference of the distances $R_{DA}^{(1)}$ and $R_{DA}^{(2)}$ as obtained from the three states model. For six variants, the distance 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. 3.3.9) 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 ($g_D/g_R = 0.75$). Shaded area in gray is the region of D-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. $\langle \tau_{D(A)} \rangle_f$.

Submillisecond conformational averaging detected by single-molecule MFD. Single-molecule experiments of freely diffusing molecules with burst analysis (3.3.1.8 and 3.3.1.10) provide direct evidence on the temporal evolution of the three found conformational states (25). 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 a surprising result (Figure 3.2.2E). The variant K19pAcF/R119C-(DA) exhibits only a single, slightly tailing population – apart from a small D-only-fraction in the grey shaded area. With the pre-knowledge that three different, well-

separated states are identified in ensemble conditions (Figure 3.2.2D), 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 (Figure 3.2.2E) 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 merging of different populations during burst duration (~ 1 ms) into only a single tailing one is a hallmark for dynamics at sub-millisecond timescales (25).

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 subdomain and the donor is located at the C-terminal subdomain.

The conformation-specific FRET species are depicted in the MFD plot in Figure 3.2.3A. To maximize time resolution in the single-molecule bursts, we used a time window of 0.5 ms for the analysis of the selected bursts. Faster dynamics were studied by fFCS, which maximizes the contrast in fluctuation analysis by using species-specific filters (Figure 3.3.3A-D). These filters depend on the species-specific fluorescence properties (fluorescence lifetime, spectral window, and fluorescence anisotropy) and are used to provide unique crosscorrelation functions, sCCF (Figure 3.2.3B and Figure 3.3.3E-H) (47, 70). To calculate the sCCFs by fFCS, filters for two species must be specified. 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 .

The normalized sCCFs of -(DA) and -(AD) labeled T4L variants agree extremely well showing two relaxation times at $t_{R1} = 4 \mu\text{s}$, and $t_{R2} = 230 \mu\text{s}$ in the global analysis of various mutants. This is indicative for at least three exchanging states (Figure 3.2.3B). The observed relaxation times in sCCF correspond to the eigenvalues of the kinetic matrix (3.3.1.12 and 3.3.2.5, Eq. 3.3.35). 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 (Figure 3.3.4D-G), which supports that three states interconvert at sub-millisecond timescale.

Particularly for the mutant S44pAcF/I150D-(DA), we observe a small but significant contribution of nearly static bursts ($\sim 5\%$) at the location of the C_3 state in the MFD plot (3.3.2.5 and Figure 3.3.6). This shows that the C_3 population actually consists of two sub-populations: (i) a dynamic fraction C_{3d} (overall $22\% \pm 10\%$) that exchanges with C_2 and C_1 on a sub-milli second timescale; and (ii) a trapped static fraction C_{3t} connected to C_{3d} .

Simulation of the kinetic network of T4 Lysozyme. Assuming a three-state system, there are four fundamental models ($C_1 \rightleftharpoons C_2 \rightleftharpoons C_3$; $C_2 \rightleftharpoons C_1 \rightleftharpoons C_3$; $C_1 \rightleftharpoons C_3 \rightleftharpoons C_2$; $C_1 \rightleftharpoons C_2 \rightleftharpoons C_3 \rightleftharpoons C_1$) to describe the connectivity of the states. Considering the microscopic reversibility of transitions we can discriminate the full cycle (fourth model), which could start in C_1 , C_2 or C_3 . Then, we are only left with three possibilities. Using dynamic FRET lines (3.3.1.8), which trace the connectivity between states, we clearly see that the transition C_1 - C_3 is forbidden (Figure 3.3.4, Figure 3.3.6 and Figure 3.2.2E), 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 the C_1 - C_2 transition or the C_2 - C_3 transition 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 (3.3.2.5). The state matrix, shown in 3.3.2.5, has two possible analytical solutions (Figure 3.2.3C-D). To take advantage of the various temporal domains of the transitions, we analyzed both solutions at various time windows (Figure 3.2.3C-D, Figure 3.3.6) (65). The visual comparison of both simulations (C_1 - C_2 fast vs. C_1 - C_2 slow) in Figure 3.2.3C, D with the experimental data shows that only the first option agrees well. This result is supported by the corresponding p -value for 1D and 2D-MFD histograms for both FRET-indicators (3.3.2.5, Table 3.3.2A, B).

The rate constants for the transition between C_{3d} and C_{3t} were selected such that their transition is slower than the diffusion of the molecules because the experimental data show that this state is reached by transitioning from C_{3d} and not from C_{3t} to C_1 - C_2 (otherwise we would observe a smear of bursts along the FRET lines between these states).

Next, we compared fFCS for simulated and experimental data, respectively, by selecting the mixture of C_1 - C_2 and cross-correlated it with the species C_3 . The sCCF of simulated data (Figure 3.3.6D) shows almost identical behavior as the experimental data (Figure 3.2.3B).

Therefore, we conclude that we are able to reproduce our experimental data by Brownian dynamics simulations (3.3.1.14, 3.3.2.5).

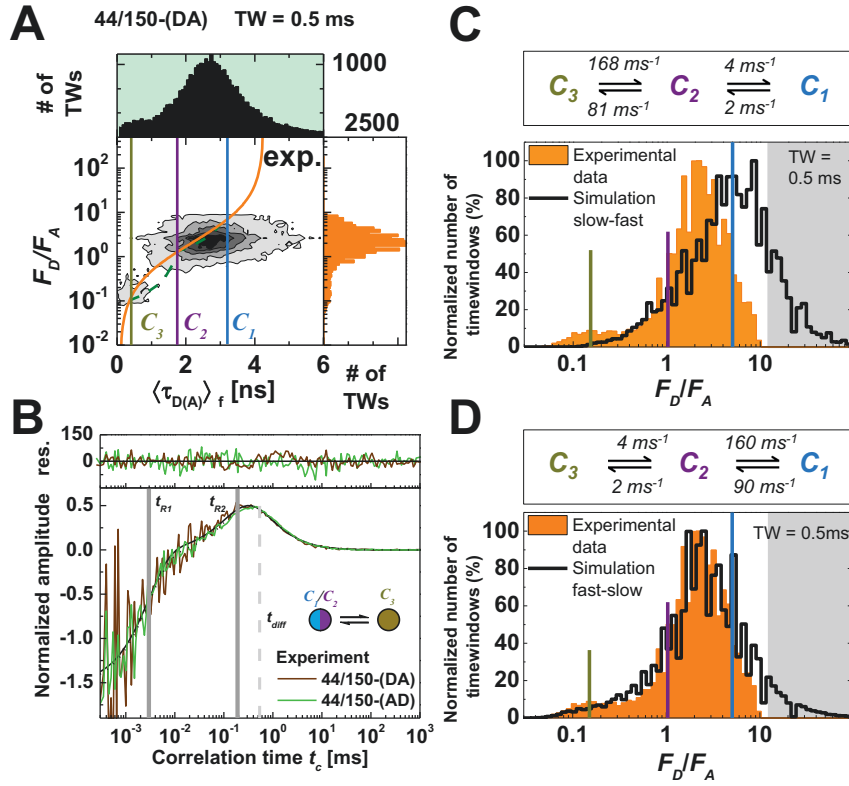


Figure 3.2.3 Kinetic connectivity. (A) MFD histogram of S44pAcF/I150C-(DA) labeled T4L analyzed in 0.5 ms timewindows. Two dimensional histogram F_D/F_A vs. lifetime of donor in the presence of acceptor $\langle \tau_{D(A)} \rangle_f$. One dimensional projections for F_D/F_A and $\langle \tau_{D(A)} \rangle_f$ are also shown. FRET line (Eq. 3.3.13) 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. 3.3.25) of S44pAcF/I150C-(DA) and S44pAcF/I150C-(AD). The global fit with other mutants shows two relaxation times ($t_{R1} = 4.0 \pm 2.4$ μ s, $t_{R2} = 230 \pm 28$ μ s) and diffusion time $t_{diff} = 0.54$ ms (C, D) Overlay of time window analysis, fast-slow in C and slow-fast in D, respectively. Details concerning the model and the simulation can be found in 3.3.2.5.

Detection of C_3 by EPR. EPR studies of a double spin labeled T4L mutant using Double Electron-Electron Resonance (DEER) provide additional support for the C_3 state. The variant S44pAcF/I150C was site-specifically labeled with the appropriate spin labels (HO-4120 and HO-225, respectively, SI section 3.3.1.5) to produce the variant 44K1/150R1, which was chosen because it contains spin labels in each domain. Under physiological pH conditions, the distribution of interspin distances is broad for this variant in the wt* background and for the variant with the covalently linked substrate adduct T26E(+) (Figure 3.2.4A, solid lines). Some of the most frequently observed distances fall within the range of 40 and 60 Å, and may correspond to the various substates of the open (C_1) and closed state (C_2), and expected

from the enzyme hinge-bending motion. In addition to these distances, a second population of distances ~ 35 Å was observed. To ensure the side state is not an artifact of the Tikhonov regularization algorithm (96, 97) or by rotamer populations of the spin label side chain, we decided to influence the conformational equilibria of the states by lowering pH (98). The FRET experiment with the mutant S44pAcF/I150C shows an increase in the population of C_3 at pH 2 (Figure 3.2.4B). The DEER experiment with the mutant S44pAcF/I150C shows an increase in the population of C_3 at pH 3 (Figure 3.2.4A). Thus, we concluded that in fact EPR and FRET show the compact state C_3 . The full pH dependence of T4L will be reported elsewhere.

3.2.2.2. Trapped reaction states of T4L

To look deeper into the functional implication of the new conformational state C_3 , we separately mutated residues E11 and T26 at the active site using the backbone of the S44pAcF/I150C variant referred to as wt**. We used wt** because of the advantage to resolve all three conformations of the free enzyme (E) clearly by FRET. Both additional mutations are known to alter the catalytic activity of T4L (14, 40) and should help to identify the role of the C_3 state during enzyme catalysis. The E11A mutation inactivates T4L (40). With this mutation, the enzyme can bind, but cannot catalyze (41); thus, it is used to mimic the enzyme-substrate complex (ES). We monitored the effect of the substrate binding for the E11A mutation by FCS and compared the characteristic diffusion times, t_{diff} , in absence and presence of substrate S (purified peptidoglycan from *Micrococcus luteus*). Without substrate t_{diff} is small (0.54 ms) (Figure 3.2.4C, green curve), whereas in presence of substrate it increased by several orders of magnitude (Figure 3.2.4C, yellow curve). Additional evidence for the binding is observed by the shift towards larger donor anisotropy values upon incubation with substrate (Figure 3.3.4H). The population of the conformational enzyme states was analyzed by sub-ensemble TCSPC of the E11A/wt**-(DA) FRET population in Figure 3.2.4D. The population of C_2 in the complex ES increased by 125 % as compared to the free enzyme E with a concomitant reduction of C_1 , while there is no effect of substrate binding for wt**-(DA) observable (Figure 3.2.4E). Note that mutations influence slightly the conformational equilibrium already in the absence of substrate for wt** and the E11A mutation. Therefore, we used the relative population changes to compute the conformational populations of the different enzyme states of wt** in the later discussion.

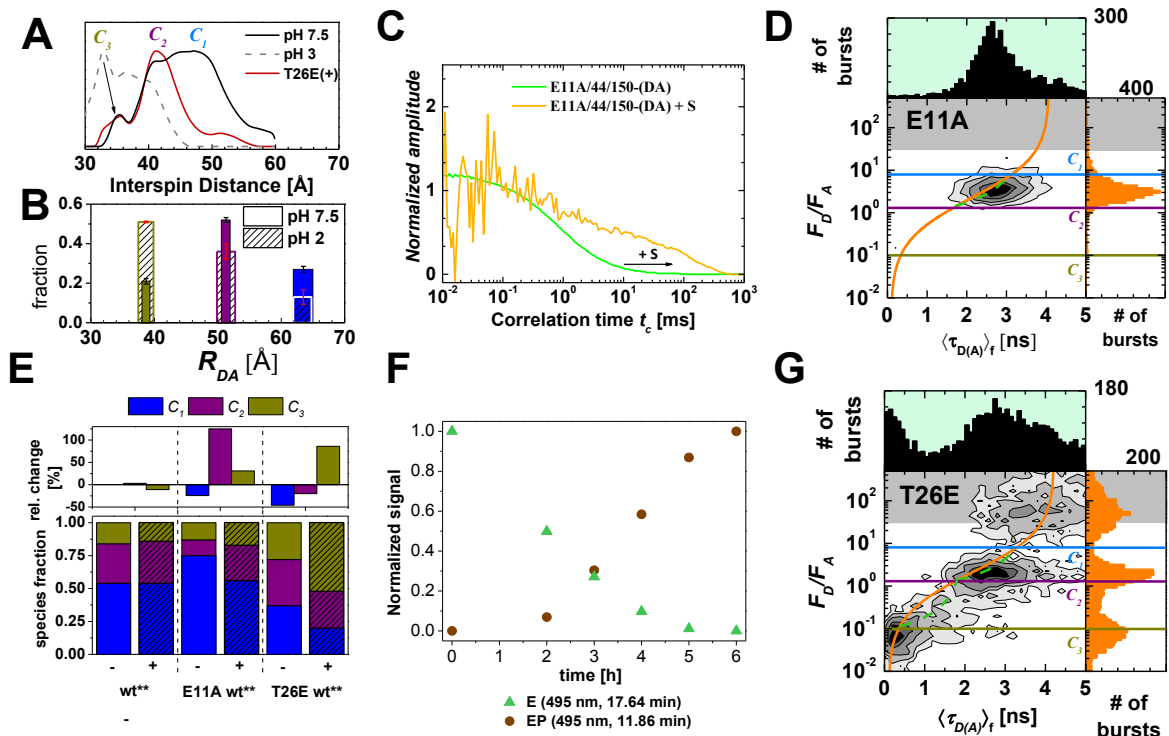


Figure 3.2.4 Functional states of T4L. (A, B) Normalized interspin distance distributions for S44(KI)/I150(RI), T26E/S44(KI)/I150(RI) and FRET distance distributions for S44pAcF/I150C-(DA) from eTCSPC. Due to linker effects the interlabel distances are not directly comparable. The shoulder, indicated by the arrow, of the distribution at short distance agrees with the short interdyne 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) without (-) and with (+) peptidoglycan. 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) The formation of the adduct of the labeled T4L with peptidoglycan was monitored by reverse phase HPLC by plotting 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. $\langle \tau_{D(A)} \rangle_t$ histogram of variant T26E/S44pAcF/I150C-(DA) incubated with peptidoglycan. An accumulation of the high FRET states (C_{3d}/C_{3t}) is observed.

The variant T26E is able to cleave the substrate but is unable to release the product due to the formation of a covalent adduct (14). In analogy to X-ray crystallography (PDB ID 148L) we used this variant to monitor the transition to the product bound enzyme state (EP) of T4L via this intermediate adduct. To confirm the formation of the adduct form under our measurement conditions, we monitored the adduct formation of labeled T4L (T26E/S44pAcF/I150C-(DA) variant) by HPLC (Figure 3.2.4F). T4L without substrate (E) is eluted at ~ 18 min. After incubation with the substrate, the peak of E drops, and a new elution peak at ~ 12 min, indicative of the adduct form of T4L (EP), was detected with increasing

incubation time (Figure 3.2.4F, Figure 3.3.5). Both, ensemble (Figure 3.2.4E) and single-molecule MFD measurements (Figure 3.2.4G) show a significant increase of the relative fraction of the C_3 state. This effect is also seen in the EPR measurements (Figure 3.2.4A). In the T26E variant the accumulation of the C_3 state is connected to the inability of this variant to release a part of the product (14). As the product release is an important step in the catalytic cycle of enzymes, we conclude that the new state must be involved in this step.

3.2.2.3. Structural features of the conformational states

We compared the three distance sets (Table 3.3.1D, E) obtained by eTCSPC with all known X-ray structures of T4L to find the structure that agrees best with C_3 . To realize an unbiased quantitative comparison, we used the results of the fifth fit model in 3.2.2.1 (analysis over all 24 variants with three sets of 24 distances (one for each conformer), global species fractions and distance distributions due to the flexible dye linkers (Table 3.3.1D, E). 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. Excluding the deposited structures with severe modifications, we applied FPS with the obtained FRET-restraints to finally screen 389 structural models where water molecules and ligands were stripped. FPS is used to label a known structure *in silico* for computing an

error function $\chi_{r,FPS}^2(j) = \frac{1}{N} \sum_{i=1}^N \frac{(R_{DA}^{(i)}{}_{\text{model},j} - R_{DA}^{(i)}{}_{\text{experiment}})^2}{\Delta R_{DA}^2(\kappa^2, k_{FRET})^{(i)}}$, where each measured distance

$R_{DA}^{(i)}{}_{\text{experiment}}$ is compared to the modeled distance $R_{DA}^{(i)}{}_{\text{model}}$ and weighted by the experimental uncertainties $\Delta R_{DA}^2(\kappa^2, k_{FRET})^{(i)}$ over the number of measurements ($i=1\dots N$; $N = 24$) for each conformational state (j) (3.3.1.15).

First, we used the two distance sets $\{C_1, C_2\}$ of the two major populations, to find a pair of PDB-deposited structures in a combined screening (389 x 389 matrix of structures), which simultaneously satisfies both sets. The screened pairs, sorted according to $\chi_{r,FPS}^2$, show a funnel towards the pair of structures that represent the experimental states best (Figure 3.2.5A).

We use a χ_r^2 criterion to exclude the pairs of very unlikely representatives. For both major states 54 conformer pairs fall below $\chi_{r,FPS,\text{threshold}}^2$. For the major state ($x_1=0.44$), we identify four conformers (172L, 151L, 178L, 189L), all are within 1.6 Å C α RMSD (Root Mean Squared Deviation) from 172L. Therefore, we assign the C_1 population to the *open* conformation. For the less populated state ($x_2 = 0.32$), we find 49 C_2 conformers, all are

within 0.73 Å C α RMSD from 148L. Hence, the population of C_2 represents the *closed* conformation. This assignment was done with a confidence level of 95 % (F-test, $\chi^2_{r,FPS,threshold} = 1.16$) and eliminates 151.267 of the 151.321 possible combinations (Figure 3.2.5A). The lowest $\chi^2_{r,FPS}$ is 1.07 and corresponds to PDB ID 178L (T4L wt* with two new cysteine residues) (34) for the state C_1 while C_2 can be described best by PDB ID 1LWK (99) – a T4L wt* variant containing multiple methionine substitutions. An alignment of both best crystal structures shows that in 1LWK the N- and C-terminal lobes are closer together as compared to 178L. Also, the structures 178L and 1LWK deviate only by 0.8 Å RMSD and 0.5 Å RMSD from our initial modelling structures 172L and 148L, respectively.

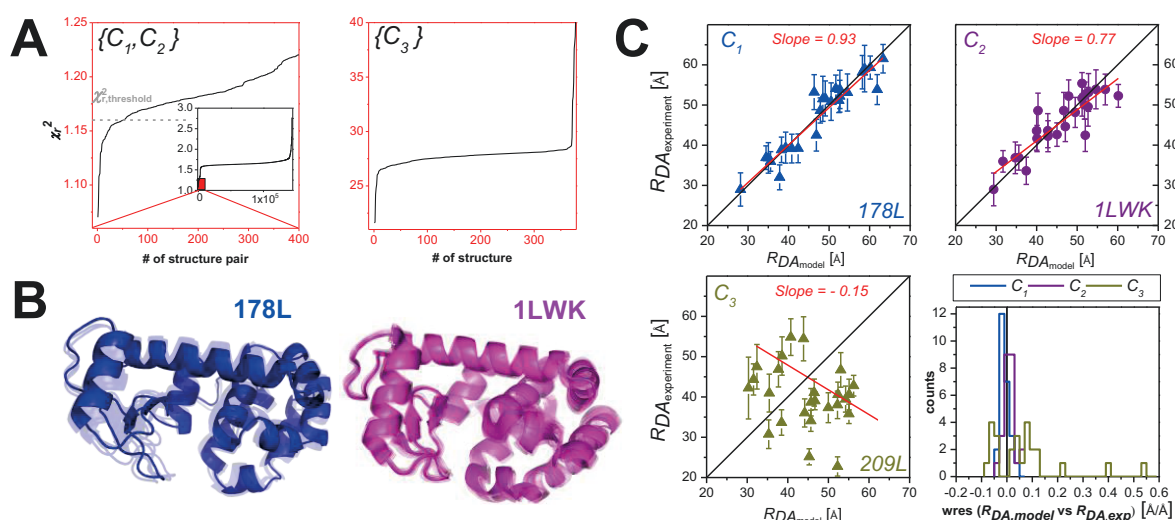


Figure 3.2.5 PDB screening. (A) A global screening containing the determined distances for C_1 and C_2 was performed against 389 out of 583 total structures deposited in the PDB for T4L. The resulting 389x389 structural combinations were sorted with increasing $\chi^2_{r,FPS}$. For C_1 and C_2 (left) the inset shows the development of the combined $\chi^2_{r,FPS}$ of all 151.321 possible combinations. Below a threshold of $\chi^2_r=1.16$ the structures 151L, 172L, 189L and 178L agree with C_1 and 49 structures agree with C_2 . The best structures for C_1 and C_2 are 178L and 1LWK, respectively. C_3 was screened separately. The best structure for C_3 is 209L. (B) The best structure of C_1 and C_2 are shown as cartoon representation in blue and magenta, respectively. All structures below $\chi^2_{r,threshold}$ aligned to the C α atoms of the best structure are shown in transparent. (C) For the best PDB structure the experimental distances $R_{DA,experiment}$ are plotted against the model distances $R_{DA,model}$ and fitted linearly (red line). The black lines show a 1:1-relationship. Weighted residuals (error weighted difference between model and experiment) divided by the experimental distances are shown in the histogram on the right.

To check the quality of the screening, we plotted the modeled distances $R_{DA,model}$ against the experimental distances $R_{DA,exp}$ (Figure 3.2.5B). Here, ideally a 1:1 relation is expected (shown as black line). The slope of the linear regression (red line) is close to one, which demonstrates the absence of significant systematic deviations.

To find the best structure for C_3 we separately screened all 389 structures of T4L with the third distance set. The structure with lowest $\chi^2_{r,FPS}$ found for C_3 is PDB ID 209L (100), which contains an insertions of three alanines in the middle of helix c (Figure 3.2.1B, orange).

Contrary to C_1 and C_2 , for C_3 we find no structure fulfilling our experimental restraints. For the best X-ray structure all distances disagree significantly within the estimated errors (Figure 3.2.5C). This is reflected by dramatically worse $\chi^2_{\text{r,FPS}} = 21.6$ and the absence of a correlation between model and experimental distances. This indicates that this structure does not correctly represent the C_3 state. We conclude that currently no T4L structure in the PDB represents the newly found state C_3 .

3.2.3. Discussion

3.2.3.1. The C_3 state and its structural properties

We present four distinct arguments that corroborate the existence of C_3 : error statistics of data analysis, structural validation of the obtained FRET parameters, kinetic behavior, and effects of specific mutations.

Error statistics. Key to the analysis and determination of C_3 was the use of global analysis of time-resolved ensemble fluorescence decays. It is important to recognize that oversampling by simultaneous consideration of 24 variants reduces the variability and uncertainty associated with FRET experiments. In addition, we used rigorous statistical analysis (*F-test*) comparing the differences between a two and a three states model, which allowed us to exclude the two states model with a confidence level of 99 % (Figure 3.2.1C, Figure 3.3.2B).

Structural validation. The global analysis of ensemble fluorescence decays using a two states model yielded two distance sets, which could not describe the expected interdye distances of the known conformers (C_1 (172L) and C_2 (148L)). In contrast to the three states model (Figure 3.2.5B top), when using the two states model we did not observe the expected linear correlation between modeled and experimental interdye distances in Figure 3.3.2C-D.

Kinetic behavior. Considering several variants, the single-molecule experiments directly showed the presence of an additional FRET population in the MFD plots, which differed significantly from C_1 and C_2 (Figure 3.2.3, Figure 3.2.4, Figure 3.3.4). This FRET population was populated or depopulated with specific mutations that alter T4L's overall catalytic activity. Moreover, global fluctuation analysis recovered at least two relaxation times that are shared throughout all studied variants (Figure 3.2.3, Figure 3.3.3). In kinetic theory, two relaxation times indicate at least three states in equilibrium, which were reproduced by simulations (Figure 3.3.6).

Specific mutations. The variant Q69pAcF/P86C-(DA) is especially informative, as the donor is placed in the middle of helix c (Orange Figure 3.2.1B) connecting both subdomains, while

the acceptor is located in the middle of helix d, which is part of the CTD (Brown Figure 3.2.1B). According to the AV calculations the interdye distances for C_1 (172L) and C_2 (1C67) states are hardly distinguishable by FRET, R_{DA} of 34 and 35 Å, correspondingly. Assuming that both subdomains preserve their secondary structure, the compaction of T4L in C_3 could proceed only by a kink in helix c so that the two subdomains approach each other, but given the location of the dyes and the extension of the dye linkers, their expected orientations lead to the increase of the interdye distance. Hence, we expect the interdye distance of the new state C_3 to be larger than the experimental distances for C_1 (38 Å) and C_2 (34 Å). The additional observed distance of 48 Å agrees with this hypothesis (Table 3.3.1C, Figure 3.3.2A and Figure 3.3.4F).

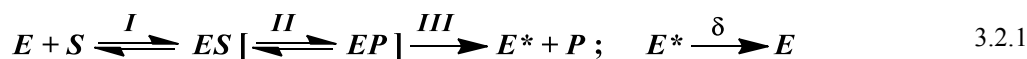
It is known that salt bridges can influence the hinge bending motion and T4L's enzymatic function. To test this we used the inactive mutant R137E (101, 102) where the salt bridge between residues 22 and 137 is disturbed. When this particular mutation is added to the variant S44pAcF/I150C-(DA) the C_3 population is reduced by ~ 50 % (Figure 3.3.4G, Table 3.3.1D), a phenotype also observed on the inactive mutant E11A/S44C/I150C-(DA).

To conclude, contrary to what has been traditionally indicated in literature (35) about T4L, where a two-state model with a single hinge-bending motion would suffice to describe its function, there is now broad evidence of the need for a 3rd state. We find that the domain motions of T4L are more complex than a single hinge-bending motion and involve more than two conformational states, which agrees with recent observations of other groups (37, 92, 93).

3.2.3.2. *Relationship between conformational states and its function*

T4L hydrolyses peptidoglycan in three steps, which involve E11 and T26 of our functional mutants as essential residues. First, the glycosidic bond of the substrate S is protonated by E11. Simultaneously, a nucleophilic attack of water molecules hydrogen bonded to residues D20 and T26 to the C-1 carbon of S occurs. This leads to the covalent adduct observed in (35). Second, the proton is presumably restored from D20 to E11 through solvent transfer and, finally, the product is released from the active site.

To link this functional cycle to our three observed conformations, we used an extended Michaelis-Menten (eMM) mechanism suggested by Kou *et al.* (85) (eq. 3.2.1).



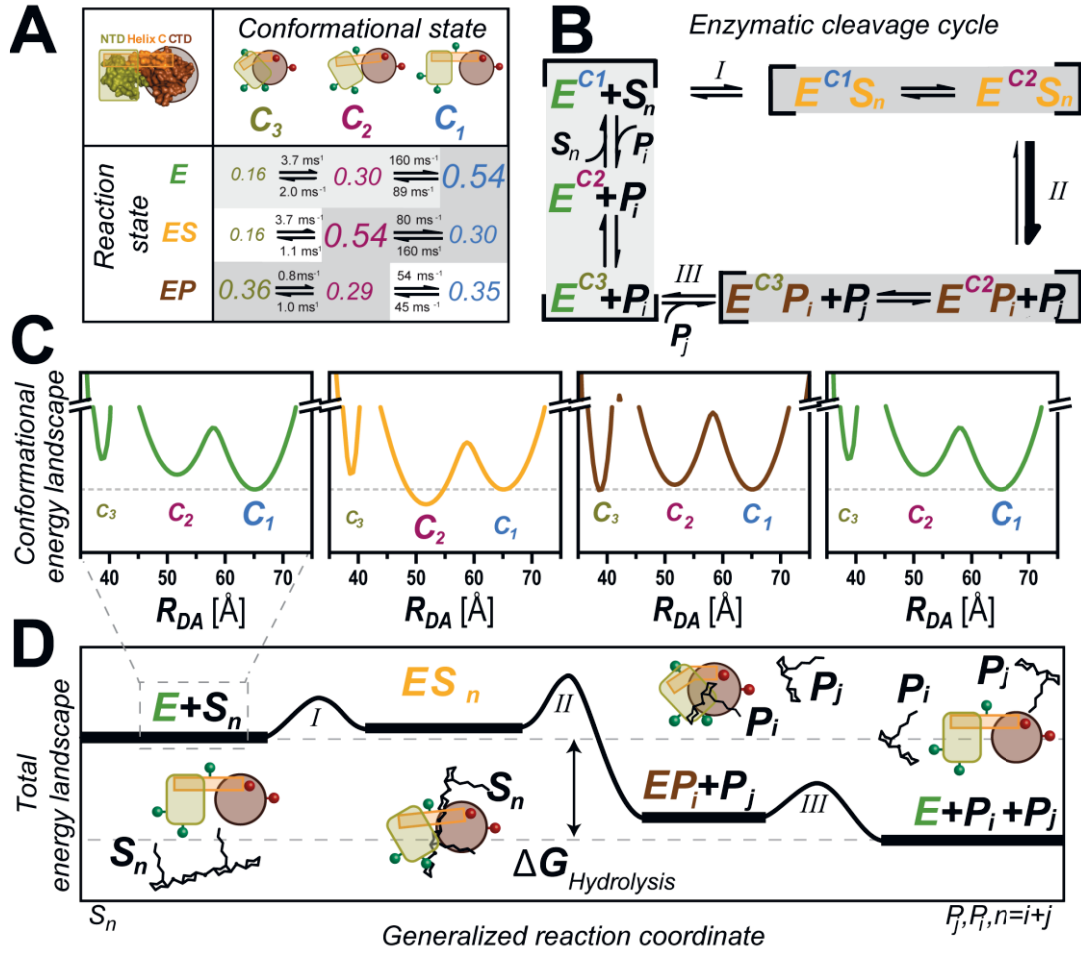


Figure 3.2.6 Energy landscape of T4L. (A) T4L interconverts between three major conformational states, C_1 , C_2 , and C_3 . Population fractions of C_1 , C_2 , and C_3 are normalized to the variant S44AcF/I150C-(DA) in the absence of substrate using the relative changes in Figure 3.2.4E to correct for influence by specific mutations in the absence of substrate. The different font sizes represent the species fractions x_i for each conformer according to Table 3.3.1F and satisfy $x_1 + x_2 + x_3 = 1$. 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. Reaction rate constants were calculated according to Figure 3.2.3B. Confidence intervals on the reaction rate constants are shown in Table 3.3.2C. (B) The peptidoglycan chain with n subunits (S_n) is cleaved into two products (P_i and P_j with $n = i + j$) by T4L, both of which can be further processed by T4L until only the dimer of N -acetylglucosamine and N -acetylmuramic acid is left. Gray shaded steps indicate the conformational/reaction states observed. (C) Relative Gibbs free energy landscapes were calculated using $\Delta G^0 = -k_B T \ln \left(\frac{k_{ji}}{k_{ij}} \right)$, 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 $\Delta G^{0+} = -k_B T \ln \left(\frac{k_{ji}}{k_0} \right)$

assuming $k_0 = 10^3 \text{ ms}^{-1}$ as an arbitrary constant. Solid line distributions consider C_1 , C_2 , and C_3 to follow a Gaussian distribution as a function of the interdye distance R_{DA} . Gaussian widths (w_{DA}) were adjusted to satisfy the energy differences and calculated activation energies. (D) Total energy landscape of the hydrolysis of T4L on a generalized reaction coordinate.

Here, the substrate S binds reversibly to the enzyme E to form an enzyme-substrate complex ES . Then, in the ES complex S is converted to the product P , resulting in the EP complex

with the product still bound to the enzyme. P is released from the complex via a transition of E to an excited state E^* . Finally, the free enzyme E^* relaxes to E .

By considering the eMM framework and using the S44/I150C-(DA) backbone and two key functional T4L variants (E11A, T26E), we present the relation between the conformational (C_1 , C_2 , C_3) and the above described reaction states (E , ES , EP) and elucidate the functional role of E^* (Figure 3.2.6). We observed that the populations of the conformers (Figure 3.2.4) differ significantly between the three reaction states.

To connect conformational equilibria with catalysis, we analyzed the relative changes in population fractions observed across the functional mutants in the absence and presence of substrate (upper panel of Figure 3.2.4E) by generating a 3 x 3 state matrix (Figure 3.2.6A). The matrix shows that in each enzyme reaction state specific conformational states are favored (Figure 3.2.6A, B).

In the free enzyme state, E , the open conformation C_1 is mostly populated to enable substrate binding, which starts the catalytic cycle by the formation of ES . Now, the closed conformation C_2 is the most abundant conformation as found also by (35). In this conformation the glycosidic bond is cleaved, such that C_2 connects ES and EP . Our studies indicate that product release occurs in the compact conformation C_3 , which is majorly increased in EP . Thus C_3 links EP and E so that the original enzyme E is regenerated from EP and the enzymatic cycle is closed. In consequence, the compact state C_3 would correspond to the excited conformational state E^* (eq. 3.2.1) (Figure 3.2.6A, C) the function of which is to spill out the product (83, 93) (Figure 3.2.6A, C). Note that, even under saturating conditions, which favor ES and EP states, the enzyme is still in dynamic equilibrium between conformational states.

To visualize the energetic changes of the enzyme at various steps in the catalytic cycle, we used the obtained population fractions and reaction rate constants to compute the relative energy landscape based on the Arrhenius equation (Figure 3.2.6C). Each energy landscape is independently normalized to C_1 . To plot these relative energy landscapes, we assumed that each local basin could be represented by a Gaussian distribution of energies where the widths were adjusted such that the barrier height reflects the activation energies along the interdyer distances as the selected reaction coordinate.

We demonstrate the importance of describing the function of thermally fluctuating enzymes by a state matrix. This approach allows us to assign a functional role to conformational fluctuations involving short-lived states, which will foster understanding of enzyme

dynamics. The enzymatic pathway highlighted in grey in the eMM (Figure 3.2.6A, B) allows us to identify key conformational states in the reaction cycle and to reconstruct the free energy landscape (103) of an enzyme for the reaction steps I, II, and III (Figure 3.2.6D). Our observations demonstrate that a fine-tuned shift of the conformational equilibrium favors motions of active product release in T4L where the energy of product formation in step II defines the directionality of the reaction (104). This hydrolysis reaction is irreversible and thus can be denoted as “ratchet mechanism” (105).

In T4L, the new transient C_3 state has a lifetime of approximately 230 μ s. All our evidence indicates that the conformational state C_3 , which is more compact than any other structure known of T4L, is needed after catalysis and not before. Its compactness suggests a functional role that is related to product release via an excited state E^* (eq. 3.2.1). This mechanism can be an evolutionary advantage to release the product when processivity is required for function. 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. 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 (106).

In view of the Michaelis-Menten based description of enzyme kinetics, our results stress the importance of considering also short-lived functional states in kinetic descriptions, i.e. we speculate that short-lived (excited states) product release states could exist for many enzymes. As shown for T4L and horseradish peroxidase (83, 93), the increased time resolution of confocal FRET experiments promises the detection and structural characterization of many more excited states of enzymes. The dynamic equilibrium and exchanges between conformational states departs from the rigid view of enzymes as being locked in particular states; however, this effect is difficult to observe by other methods. It is our next challenge to determine the structure of this hidden state C_3 to validate the possibility of this product release step based on its three-dimensional structure. It remains to be seen whether such compaction and motions are common for other proteins with comparable subdomain architecture.

3.2.4. Materials and Methods

Sample preparation. Mutagenesis of the pseudo-wild-type (wt*) with the mutations C54T and C97A resulting in 24 T4L mutants was carried out as previously described (95). 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 Alexa488 was used as donor and Alexa647 as acceptor dye with a Förster Radius $R_0 = 52 \text{ Å}$. 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 (69) (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 (95). FRET positioning and screening was done according to Kalinin *et al.* (46). More detailed description of all materials and methods is given in SI (3.3.1).

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 Alexa488 and Alexa647 respectively) and length (l : 20 Å and 22 Å for Alexa488 and Alexa647 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) (53). 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 (3.3.1.13). From all donor and acceptor positions ($\overline{R}_D^{(i)}$ and $\overline{R}_A^{(j)}$) the interdye distance $\langle R_{DA} \rangle$, measured by TCSPC, is the average

$$\langle R_{DA} \rangle = \left\langle \left| \overline{R}_D^{(i)} - \overline{R}_A^{(j)} \right| \right\rangle = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \left| \overline{R}_D^{(i)} - \overline{R}_A^{(j)} \right|.$$

Data analysis. Fluorescence lifetime measurements, confocal Multiparameter Fluorescence Detection experiments, and filtered fluorescence correlation spectroscopy curves were analyzed by in house written programs, which can be downloaded from <http://www.mpc.uni-duesseldorf.de>. Details on the methodology are found in SI.

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 (SI).

1. The labeling does not alter enzyme function. Labeled T26E mutant shows expected adduct formation (Figure 3.2.3, Figure 3.3.5).
2. Local quenching of donor dye is considered when calculating distances and cross correlations (Table 3.3.1).
3. Triplet state quenchers do not affect the observed relaxation times and amplitudes on the species cross-correlation (Figure 3.3.7).
4. Acceptor cis-trans isomerization does not contribute to the signal on the species correlation analysis (Figure 3.3.7).
5. κ^2 distributions show that our assumption of $\kappa^2 = 2/3$ is valid (Table 3.3.3). Table 3.3.3 summarize the residual anisotropies (r_∞) of D - donor, A - acceptor and A(D) by the FRET sensitized emission of acceptor used for calculating κ^2 distributions (25).
6. The existence and the population fraction of the new conformational state C_3 is consistent across our library of mutants (Table 3.3.1B, C), $x_3 = 28 \pm 10$ %. 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 (20). 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 structures determined by X-ray crystallography of T4L without outliers (Figure 3.2.5, Table 3.3.1C).
8. We oversample FRET restraints to reduce uncertainty introduced by each point mutation (Figure 3.2.5, Table 3.3.1C).

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 (3.3.1.11).

Acknowledgements

We thank Marina Rodnina, Philipp Neudecker, and M. Neal Waxham for comments and suggestions on the manuscript; and Daniel Rohrbeck for support on the mutagenesis. We thank Evan Brooks for generating some T4L mutants and for helping with DEER sample preparation. This work was supported by the Deutsche Forschungsgemeinschaft (SE 1195/13-1) and EU (FP7-Health-2007-A-201837) (to CAMS) and by NIH grant R01EY05216 (to WLH) and the Jules Stein Professorship Endowment (to WLH). HS acknowledges support from the Alexander von Humboldt foundation and Clemson Start-up funds. KH is supported by the NRW Research School BioStruct. HG acknowledges support from the “Fit for Excellence” initiative of Heinrich-Heine-University.

Author Information

H.S. current address: Clemson University, SC, USA; M.R.F. current address: Avanir Pharmaceuticals Inc., Aliso Viejo, California, USA; correspondence and requests should be addressed to C.A.M.S. (cseidel@hhu.de) and H.S. (hsanabr@clemson.edu).

3.3. Supplementary Information

3.3.1. Materials and Methods

3.3.1.1. Experimental Design

To address how T4L dynamics relates to its function, we used a hybrid approach that combines 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 timewindows 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 Figure 3.3.1.

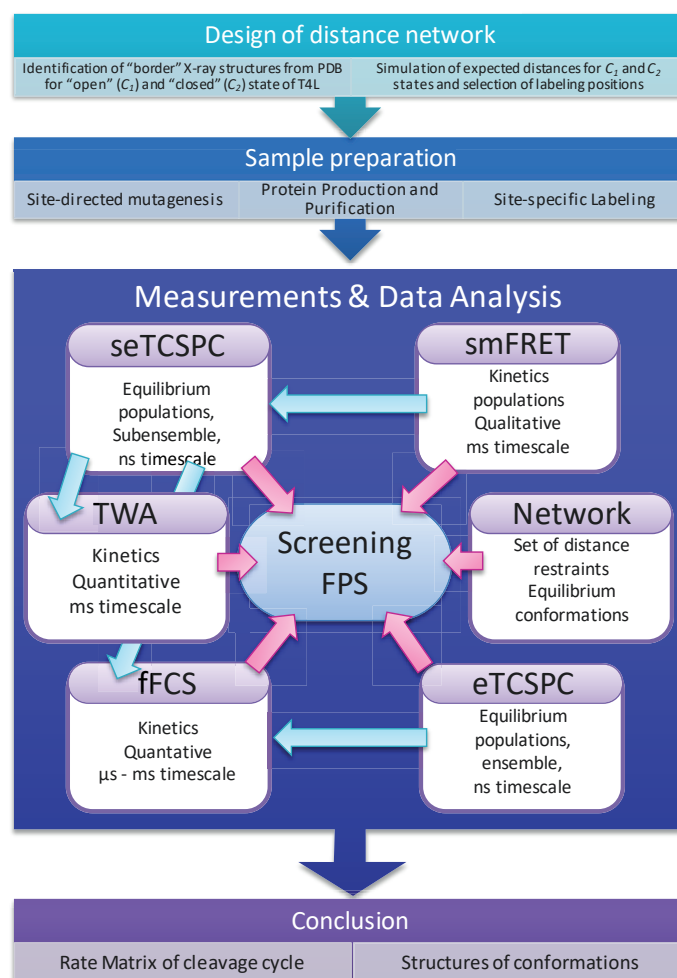


Figure 3.3.1 Flowchart of data acquisition and data analysis. After identification of border X-ray structures, a distance network spanning the whole T4 lysozyme is built. T4L variants are generated via site-directed mutagenesis, produced in *E. coli*, purified and site-specifically labeled. Of all variants the following experiments are conducted: ensemble time correlated single photon counting (eTCSPC), multiparameter fluorescence detection (MFD) and filtered fluorescence correlation spectroscopy (fFCS). Of single molecule data, population-specific data can be cut and sub-ensemble TCSPC (seTCSPC) can be done. Time-window analysis (TWA) is used to discriminate between different kinetic models.

3.3.1.2. 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.) (94, 95, 107). The plasmid containing the gene with the desired mutant was co-transformed with pEVOL (107) 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 OD₆₀₀ 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 at 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.

3.3.1.3. 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, Dinslaken, Germany). 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.

3.3.1.4. 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 Alexa488 and Alexa647 (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 (95). Double mutants were labeled sequentially - first thiol and second the keto handle, as suggested by Brustad *et al.* (94). 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.).

3.3.1.5. EPR spectroscopy

For double electron electron resonance (DEER) measurements of doubly spin labeled proteins, ~ 200 μ M 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 (π) 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 (96).

3.3.1.6. 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* (~230 ps FWHM) 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.

3.3.1.7. 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 μ W) 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 μ m 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 μ M TROLOX and 1 μ M 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 (Immersionol, Carl Zeiss Inc., Germany). NUNC chambers (Lab-Tek, Thermo Scientific, Germany) were used with 500 μ L 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.

3.3.1.8. MFD burst analysis: Multiparameter FRET histograms and FRET lines

Bursts were selected by 2σ 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 (62) using in house developed programs (LabVIEW, National Instruments Co.). Fluorescent bursts were plotted in 2D histograms (Origin 8.6, OriginLab Co).

The relation of donor fluorescence over the acceptor fluorescence F_D/F_A and the species weighted average donor lifetimes $\langle \tau \rangle_x$ depends on specific experimental parameters such as fluorescence quantum yields of the dyes ($\Phi_{FD(0)}$ and Φ_{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 (α):

$$\left(\frac{F_D}{F_A} \right) = \frac{\Phi_{FD(0)}}{\Phi_{FA}} \cdot \left(\frac{\langle \tau_{D(0)} \rangle_x}{\langle \tau_{D(A)} \rangle_x} - 1 \right)^{-1} \quad 3.3.1$$

Here, the brackets $\langle \dots \rangle_x$ describe the averaging over all lifetime components $\tau^{(i)}$ weighted $x^{(i)}$, their respective species fractions:

$$\langle \tau_{D(0)} \rangle_x = \sum_i x^{(i)} \tau_{D(0)}^{(i)} \quad \text{and} \quad \langle \tau_{D(A)} \rangle_x = \sum_i x^{(i)} \tau_{D(A)}^{(i)} \quad 3.3.2$$

In sm-experiments only ~ 100 green photons per burst are detected. Hence, only the average time since excitation can be determined reliably by the maximum likelihood estimators (MLE). This experimentally determined time is weighted by the fluorescence intensity and relates to the lifetime components by:

$$\langle \tau_{D(0)} \rangle_f = \frac{\sum_i x^{(i)} \tau_{D(0)}^{(i)2}}{\sum_i x^{(i)} \tau_{D(0)}^{(i)}} \quad \text{and} \quad \langle \tau_{D(A)} \rangle_f = \frac{\sum_i x^{(i)} \tau_{D(A)}^{(i)2}}{\sum_i x^{(i)} \tau_{D(A)}^{(i)}} \quad 3.3.3$$

We call this lifetime the fluorescence weighted average lifetime.

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

$$F_D = \frac{S_G - \langle B_G \rangle}{g_G} \quad 3.3.4$$

$$F_A = \frac{S_R - \alpha F_G - \langle B_R \rangle}{g_R} \quad 3.3.5$$

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.

The two averaged observables F_D/F_A (in eq. 3.3.1) and the fluorescence weighted average lifetime $\langle \tau_{D(A)} \rangle_f$ can be related two each other. We call a line which relates these two observables a “FRET-line”. These FRET-lines are projections of a parametrization of a multi-dimensional lifetime distribution to a two dimensional plane using either the transfer-efficiency E or F_D/F_A as one and $\langle \tau_{D(A)} \rangle_f$ as second axis.

The fluorophores are moving entities coupled to specific places via flexible linkers. Therefore, already for single protein conformations a distance distribution has to be considered. For simplicity, we use normal distributions to describe this dependency. If the donor and acceptor in average are separated by $\overline{R_{DA}}$ the corresponding distance distribution is approximated by:

$$p(R_{DA}) = \frac{1}{w_{DA} \sqrt{\pi/2}} \exp \left(-2 \left[\frac{R_{DA} - \langle R_{DA} \rangle}{w_{DA}} \right]^2 \right) \quad 3.3.6$$

Here, w_{DA} is half-width of the distance distribution attributed to the broadening due to the linker-flexibility set to a physical meaningful value of 12 Å (53). Using the Förster-relationship $\tau_{D(A)}(R_{DA}) = \tau_{D(0)} \cdot \left(1 + (R_0/R_{DA})^6\right)^{-1}$ and the following integrals:

$$\langle \tau_{D(A)} \rangle_x = \int \tau_{D(A)}(R_{DA}) p(R_{DA}) dR_{DA} \quad 3.3.7$$

$$\langle \tau_{D(A)} \rangle_f = \frac{\int (\tau_{D(A)}(R_{DA}))^2 p(R_{DA}) dR_{DA}}{\langle \tau_{D(A)} \rangle_{x,L}} \quad 3.3.8$$

This (eq. 3.3.6) distribution can be projected to a point in the $F_D/F_A - \langle \tau_{D(A)} \rangle_f$ plane. If now the average donor-acceptor distance $\overline{R_{DA}}$ is varied within a given range (i.e. $[0, \infty]$) a line within the $F_D/F_A - \langle \tau_{D(A)} \rangle_f$ plane is obtained. Such a line we call a static FRET-line, as it is valid for all molecules with a given (single) conformation, irrespectively of the mean donor-acceptor separation, $\overline{R_{DA}}$.

To describe the position of a molecule on the $F_D/F_A - \langle \tau_{D(A)} \rangle_f$ plane which is interconverting between two states with mean distances $\overline{R_{DA}}^{(1)}$ and $\overline{R_{DA}}^{(2)}$ and fractions $x^{(1)}$ and $x^{(2)} = 1 - x^{(1)}$, we use the following distance distribution:

$$p(R_{DA}) = \frac{x_{DA}^{(1)}}{w_{DA} \sqrt{\pi/2}} \exp\left(-2 \left[\frac{R_{DA} - \langle R_{DA}^{(1)} \rangle}{w_{DA}} \right]^2\right) + \frac{1 - x_{DA}^{(1)}}{w_{DA} \sqrt{\pi/2}} \exp\left(-2 \left[\frac{R_{DA} - \langle R_{DA}^{(2)} \rangle}{w_{DA}} \right]^2\right) \quad 3.3.9$$

To obtain a line, which is valid for a molecule in exchange between these two states, the fraction x_I is varied within the range $[0, 1]$ and the position in the $F_D/F_A - \langle \tau_{D(A)} \rangle_f$ plane is calculated using the eq. 3.3.7, 3.3.8 and eq. 3.3.9.

3.3.1.9. Donor and Acceptor fluorescence 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 $\Phi_{FD(0)}$, Φ_{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 $\Phi_{FA} = 0.32$ (80). The obtained donor and acceptor fluorescence quantum yields are presented in Table 3.3.1A. This FRET pair has a Förster distance of 52 Å.

3.3.1.10. *Guideline 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 Alexa488). *II)* High FRET appears at shorter lifetimes when the fluorescence of acceptor is high ($F_D/F_A \rightarrow 0$). *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 (63). *IV)* Static FRET states follow a theoretical line that accounts for dye linker mobility called "static FRET line" (63). *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.

3.3.1.11. *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.

$$F_{D(0)}(t) = \sum_i x_{D(0)}^{(i)} \exp(-t/\tau_{D(0)}^{(i)}) \quad 3.3.10$$

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_{RET} . This is true if quenching does not change the donor radiative lifetime and if FRET is uncorrelated with quenching of the donor by its local environment. Given these assumptions the donor fluorescence intensity decay in the presence of FRET $F_{D(A)}(t)$ can be factorized into the donor fluorescence decay in absence of FRET and the time-resolved FRET-induced donor quenching $\varepsilon_{D(A)}(t)$:

$$F_{D(A)}(t) = F_{D(0)}(t) \cdot \varepsilon_{D(A)}(t) \quad 3.3.11$$

We relate the FRET-induced donor decay to the distribution of distances by the rate-constant of energy transfer as defined by Förster:

$$k_{RET} = k_F \cdot \kappa^2 \cdot \left(\frac{R_{0J}}{R_{DA}} \right)^6 \quad 3.3.12$$

Here, R_{0J} is a reduced Förster-radius, k_F is the radiative rate constant of fluorescence and κ^2 is the orientation-factor. The reduced Förster-radius is given by:

$$R_{0J} = \left[\frac{9(\ln 10)}{128\pi^5 \cdot N_A} \cdot \frac{J}{n^4} \right]^{\frac{1}{6}} = 0.2108 \cdot \text{\AA} \cdot \left[\frac{1}{n^4} \cdot \left(\frac{J(\lambda)}{\text{mol}^{-1} \cdot \text{dm}^3 \cdot \text{cm}^{-1} \cdot \text{nm}^4} \right) \right]^{\frac{1}{6}} \quad 3.3.13$$

where N_A is the Avogadro's constant, n is the refractive index of the medium and $J = \int f_D(\lambda) \cdot \varepsilon_A(\lambda) \cdot \lambda^4 d\lambda$ is the overlap integral between $f_D(\lambda)$, the donor emission spectrum and $\varepsilon_A(\lambda)$, the acceptor absorption spectrum. This reduced Förster-radius stresses that the FRET-rate constant is independent of quenching and specific for the dye-pair under the condition that the spectral overlap is independent of dynamic quenching. Under the assumption that dynamic quenching is uncorrelated with FRET the FRET-induced donor decay relates to the distance distribution $p(R_{DA})$ by:

$$\varepsilon_{D(A)}(t) = \int p(R_{DA}) \cdot \exp\left(-t \cdot \langle \kappa^2 \rangle \cdot k_F \left[1 + (R_{0J} / R_{DA})^6 \right]\right) dR_{DA} \quad 3.3.14$$

Assuming the orientation factor can be approximated by an average $\langle \kappa^2 \rangle \approx 2/3$ (justified by the anisotropy studies compiled in Table 3.3.3), this time-dependent quantifier is directly related to the donor-acceptor distance distribution. We used a reduced Förster-radius of $R_{0J} = 56.4 \text{ \AA}$ which was determined for the donor with a radiative rate constant $k_F = 0.224 \text{ ns}^{-1}$. As previously described (46) we propagate potential errors of the $\langle \kappa^2 \rangle \approx 2/3$ approximation to our experimental distances (Table 3.3.3D).

In fit models 1 to 3 we use discrete distance distributions:

$$p(R_{DA}) = \sum_{i=1}^N x_{DA}^{(i)} \cdot \delta(R_{DA} - R_{DA}^{(i)}) \quad 3.3.15$$

where δ is the delta-function, N is the number of states (1 to 3) with the apparent DA-distances $R_{DA}^{(i)}$ and $x_{DA}^{(i)}$ species fraction of the state. In fit models 4 and 5 we used continuous distance distributions which are described by a superposition of normal distributions:

$$p(R_{DA}) = \sum_{i=1}^N x_{DA}^{(i)} \frac{1}{w_{DA} \sqrt{\pi/2}} \exp\left(-2 \left[\frac{R_{DA} - \langle R_{DA}^{(i)} \rangle}{w_{DA}} \right]^2\right) \quad 3.3.16$$

Here, $\langle R_{DA}^{(i)} \rangle$ is the mean of the state (i) distance distribution with species fraction $x_{DA}^{(i)}$ and a width w_{DA} set to a physical meaningful value of 12 Å (flexible dye-linkers) (53).

The final analysis model is obtained by substituting either eq. 3.3.15 (fit-model: 1, 2, 3) or eq. 3.3.16 (fit-model: 4, 5) into eq. 3.3.14. Next, eq. 3.3.14 is inserted into eq. 3.3.13. Finally, the fluorescence intensity decays of the donor in presence and absence of FRET (eq. 3.3.17 or eq. 3.3.18) are globally analyzed. By this procedure the photon counting statistics of both the reference- and fluorescence-decay in presence of FRET is preserved. Thus, the counting statistics are clearly defined (Poisson distribution). This allows for an analysis with proper error-estimates. By the global (joint) analysis of the reference sample and the FRET-sample the photophysical properties (dynamic quenching) are taken into account.

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 *Lin*. (108) Given these corrections the experimental time-resolved fluorescence intensities of the FRET-sample and the donor reference sample are proportional to:

$$\begin{aligned} F_{\text{FRET}}(t) &= (N_0 \cdot [(1 - x_{\text{DOnly}}) F_{\text{D(A)}}(t) + x_{\text{DOnly}} F_{\text{D(0)}}(t)] \otimes \text{IRF} + sc \cdot \text{IRF} + c) \cdot \text{Lin} \\ F_{\text{Ref}}(t) &= (N_0 \cdot F_{\text{D(0)}}(t) \otimes \text{IRF} + sc \cdot \text{IRF} + c) \cdot \text{Lin} \end{aligned} \quad 3.3.17$$

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. 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:

$$\begin{aligned} F_{\text{FRET}}(t) &= N_F \cdot [(1 - x_{\text{DOnly}}) F_{\text{D(A)}}(t) + x_{\text{DOnly}} F_{\text{D(0)}}(t)] \otimes \text{IRF} + N_{BG} \cdot \text{IRF} \\ F_{\text{Ref}}(t) &= N_F \cdot F_{\text{D(0)}}(t) \otimes \text{IRF} + N_{BG} \cdot \text{IRF} \end{aligned} \quad 3.3.18$$

Summary fit models

In total, we used five different fit models to describe our data. They differ in their number of states, complexity (discrete vs continuous distributed states) and the number of joint (global) and free parameters, which are given below:

Model	Fit model	N-states	parameter		Con-straints	free parameter		Table	Average χ^2_r
			local	global		per sample	total		
Discrete FRET-states eq. 3.3.15, 3.3.17	1	1	$R_{DA}^{(1)}$	-	-	1	24 (24*1)	3.3.1 B	-
	2	2	$R_{DA}^{(1)}, R_{DA}^{(2)}$ $, x_{DA}^{(1)}, x_{DA}^{(2)}$	-	$x_{DA}^{(1)} +$ $x_{DA}^{(2)} = 1$	3 =(4-1)	72 =(24*3)	3.3.1 B	1.173
	3	3	$R_{DA}^{(1)}, R_{DA}^{(2)}$ $, R_{DA}^{(3)}, x_{DA}^{(1)}$ $, x_{DA}^{(2)}, x_{DA}^{(3)}$	-	$x_{DA}^{(1)} + x_{DA}^{(2)} +$ $x_{DA}^{(3)} = 1$	5 =(5-1)	120 =(24*5)	3.3.1 C	1.106
Normal distributed FRET-states with $w_{DA} = 12 \text{ \AA}$ eq. 3.3.16, 3.3.18	4	2	$\langle R_{DA}^{(1)} \rangle$, $\langle R_{DA}^{(2)} \rangle$	$x_{DA}^{(1)}$, $x_{DA}^{(2)}$	$x_{DA}^{(1)} +$ $x_{DA}^{(2)} = 1$	2	49 =(24*2+2-1)	3.3.1 C	1.171
	5	3	$\langle R_{DA}^{(1)} \rangle$, $\langle R_{DA}^{(2)} \rangle$, $\langle R_{DA}^{(3)} \rangle$	$x_{DA}^{(1)}$, $x_{DA}^{(2)}$, $x_{DA}^{(3)}$	$x_{DA}^{(1)} + x_{DA}^{(2)} +$ $x_{DA}^{(3)} = 1$	3	74 =(72+3-1)	3.3.1 D, E	1.106

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 χ^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 $w = p_{min} * r * w_0$ (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:

$$\chi^2_{r,max}(P) = \chi^2_{r,min} \cdot [1 + n / \nu \cdot \text{cdf}^{-1}(F(n, \nu, P))]$$
3.3.19

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 χ^2_r (109).

3.3.1.12. Filtered Fluorescence Correlation Spectroscopy

In fluorescence correlation spectroscopy (FCS) information on fluctuating systems is obtained by calculation of the correlation function (66, 110)

$${}^{A,B}G(t_c) = 1 + \frac{\langle \delta^A S(t) \cdot \delta^B S(t+t_c) \rangle}{\langle {}^A S(t) \rangle \cdot \langle {}^B S(t) \rangle}$$
3.3.20

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 $\langle {}^{A,B}S(t) \rangle$. 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 (111, 112).

$$G(t_c) = 1 + \frac{1}{N} \cdot \frac{\sum_i^n x^{(i)} \cdot (Q^{(i)})^2 \cdot G_{diff}^{(i)}(t_c)}{\left(\sum_i^n x^{(i)} \cdot Q^{(i)} \right)^2}$$
3.3.21

where

$$G(t_c) = 1 + \frac{1}{N} \cdot \frac{\sum_i^n x^{(i)} \cdot (Q^{(i)})^2 \cdot G_{diff}^{(i)}(t_c)}{\left(\sum_i^n x^{(i)} \cdot Q^{(i)} \right)^2}$$
3.3.22

A 3-dimensional Gaussian shaped volume element with spatial distribution of the detection probabilities: $W(x, y, z) = \exp(-2(x^2 + y^2)/\omega_0^2) \cdot \exp(-2z^2/z_0^2)$ is assumed. The $1/e^2$ radii in x and y or in z direction are denoted by ω_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 $x^{(i)} = \frac{c^{(i)}}{\sum_i c^{(i)}}$ has a concentration

$c^{(i)}$, and brightness $Q^{(i)}$.

To separate species, we use filtered FCS (fFCS) (47, 71). fFCS differs from standard FCS (66) and FRET-FCS (70) by interrogating the “species” (conformational states) fluctuations instead of photon count rates (70). We define the species cross- correlation function (sCCF) as

$$G^{(i,m)}(t_c) = \frac{\langle F^{(i)}(t) \cdot F^{(m)}(t+t_c) \rangle}{\langle F^{(i)}(t) \rangle \cdot \langle F^{(m)}(t+t_c) \rangle} = \frac{\left\langle \left(\sum_{j=1}^{d \cdot L} f_j^{(i)} \cdot S_j(t) \right) \cdot \left(\sum_{j=1}^{d \cdot L} f_j^{(m)} \cdot S_j(t+t_c) \right) \right\rangle}{\left\langle \sum_{j=1}^{d \cdot L} f_j^{(i)} \cdot S_j(t) \right\rangle \cdot \left\langle \sum_{j=1}^{d \cdot L} f_j^{(m)} \cdot S_j(t+t_c) \right\rangle} \quad 3.3.23$$

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 $j = 1 \dots L$ 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 (47). 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. 3.3.24.

$$\left\langle \left(\sum_{j=1}^{d \cdot L} f_j^{(i)} \cdot H_j - w^{(i)} \right)^2 \right\rangle \rightarrow \min \quad 3.3.24$$

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 (47). One major advantage of sCCF is that, if photophysical properties are decoupled from species selection, the intercrossing dynamics (70) is recovered with great fidelity.

To properly fit the species cross-correlation function, we used (47)

$$G(t_c) = 1 + \frac{1}{N} \cdot G_{diff}(t_c) \cdot [1 - G_K(t_c)] \quad 3.3.25$$

where $G_K(t_c)$ is

$$G_K(t_c) = \sum_{t_{Ri}}^{t_{Rn}} A_{Ki} \exp(-t_c / t_{Ri}) \quad 3.3.26$$

In Eq. 3.3.26 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. 3.3.22. The normalized correlation function is presented as:

$$g(t_c) = N \cdot (G(t_c) - 1) \quad 3.3.27$$

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

$$\begin{aligned} F_{\parallel}(t) &= F(t) \cdot (1 + (2 - 3l_1) \cdot r(t)) / 3 \\ F_{\perp}(t) &= F(t) \cdot (1 - (1 - 3 \cdot l_2) \cdot r(t)) / 3 \end{aligned} \quad 3.3.28$$

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 (113, 114). The anisotropy decay $r(t)$ is given by

$$r(t) = r_{0,ov} \exp(-t / \rho_{overall}) + r_{0,ba} \exp(-t / \rho_{backbone}) + r_{0,li} \exp(-t / \rho_{linker}) \quad 3.3.29$$

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

3.3.1.13. Accessible volume (AV) model and interdyer distance

The accessible volume considers dyes as hard sphere models connected to the protein via flexible linkers (modeled as a flexible cylindrical pipe) (53). The overall dimension (width and length) of the linker is based on their chemical structures. For Alexa488 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 (46, 73, 115).

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})

$$R_{mp} = \left| \left\langle \vec{R}_{D(i)} \right\rangle - \left\langle \vec{R}_{A(j)} \right\rangle \right| = \left| \frac{1}{n} \sum_{i=1}^n \vec{R}_{D(i)} - \frac{1}{m} \sum_{j=1}^m \vec{R}_{A(j)} \right| \quad 3.3.30$$

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:

$$\langle R_{DA} \rangle = \left| \left\langle \vec{R}_{D(i)} - \vec{R}_{A(j)} \right\rangle \right| = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \left| \vec{R}_{D(i)} - \vec{R}_{A(j)} \right| \quad 3.3.31$$

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 related to the distance between atoms on the backbone (C α -C α), except through the use of a structural model.

3.3.1.14. Single-molecule Brownian dynamics simulator

Simulations of single-molecule measurements were done via Brownian dynamics (116-119). 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

$$P(t_i) = k_i^{-1} \exp(-k_i t_i) \text{ and } P(t_j) = k_j^{-1} \exp(-k_j t_j) \quad 3.3.32$$

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

3.3.1.15. FRET positioning and screening (FPS)

FPS is done in four steps, and its flow is based on the recent work by Kalinin *et al.* (46). In order to do our experimental design and a combinatorial screening of C_1 and C_2 using 389

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

$$\chi_{r,FPS}^2(j) = \frac{1}{N} \sum_{i=1}^N \frac{\left(\langle R_{DA} \rangle_{\text{experiment}}^{(i)} - \langle R_{DA} \rangle_{\text{model},j}^{(i)} \right)^2}{\Delta R_{DA}^2(\kappa^2, k_{FRET})^{(i)}} \quad 3.3.33$$

where $N = 24$ is the total number of FRET distances ($\langle R_{DA} \rangle$) and the overall uncertainty

$\Delta R_{DA}^2(\kappa^2, k_{FRET})$ is determined by the error-propagation rule of $\Delta R_{DA}^2(\kappa^2, k_{FRET}) = \Delta R_{DA}^2(\kappa^2) + \Delta R_{DA}^2(k_{FRET})$. $\Delta R_{DA}^2(\kappa^2)$ is the uncertainty from the mutual orientation of donor and acceptor (κ^2 errors) and can be estimated by analyzing anisotropy decays accessible in MFD (120). We assume a weak dependence of R_θ on κ^2 (121) (Table 3.3.3D). As a consequence, correcting $\langle R_{DA} \rangle$ values for the mean κ^2 for all allowed orientations of D and A yields almost the same final screened structures as with the assumption of $\langle \kappa^2 \rangle = 2/3$. $\Delta R_{DA}^2(k_{FRET})^{(i)}$ 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. 3.3.16- 3.3.17). The error propagation was applied to all measured FRET distances (Table 3.3.1).

3.3.2. Supporting results

3.3.2.1. Fluorescence decay analysis of single and double labeled T4 Lysozyme

Selected mutants were labeled in two configurations (DA) and (AD), D for donor (Alexa488) and A for acceptor fluorophore (Alexa647). 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 subdomain and the second position corresponds to 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 Φ_{FA} , respectively. A summary table of these is shown in Table

3.3.1A. A summary of the fits using the two and three state models are found in Table 3.3.1B-F, decays are shown in Figure 3.3.2A.

Table 3.3.1A. Multi exponential fit of Donor-only labeled samples, fluorescence quantum yields of the donor (Alexa488) and acceptor (Alexa647) fitted with Eq. 3.3.10. 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. Empty cells represent parameters that are not available.

Samples	τ_1 [ns]	x_1	τ_2 [ns]	x_2	τ_3 [ns]	x_3	$\langle \tau \rangle_{x,D(0)}$ [ns]	$\Phi_{FD(0)}$	$\langle \tau \rangle_{x,A}$ [ns]	Φ_{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
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	4.41	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
Q69pAcF/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 3.3.1B. Table of determined discrete FRET rate constants (eq. 3.3.15, 3.3.17).

Measurement number	Variant	1 state (N=1)			2 states (N=2)					
		R_{DA} [Å]	x_{DOnly}	χ_r^2	R_{DA1} [Å]	x_1	R_{DA2} [Å]	x_2	x_{DOnly}	χ_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	12.4	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/I150 C-(DA)	52.0	0.12	8.12	55.4	0.70	42.0	0.30	0.06	1.11
11	N55pAcF/Q69 C-(DA)	40.4	0.33	3.57	36.8	0.74	46.2	0.26	0.29	1.24
12	N55pAcF/R11 9C-(DA)	52.0	0.51	3.36	57.0	0.68	41.0	0.32	0.42	1.26
13	N55pAcF/N13 2C-(DA)	50.0	0.12	10.9	53.7	0.63	41.1	0.37	0.07	1.26
14	N55pAcF/I150 C-(DA)	50.6	0.62	1.57	53.4	0.76	42.4	0.24	0.59	1.19
15	K60pAcF/P86 C-(DA)	47.9	0.47	2.61	50.7	0.66	38.8	0.34	0.41	1.12
16	K60pAcF/R11 9C-(DA)	50.2	0.14	10.1	53.6	0.65	40.7	0.35	0.09	1.23
17	K60pAcF/N13 2C-(DA)	45.6	0.27	6.25	49.3	0.53	38.0	0.47	0.21	1.05
18	K60pAcF/I150 C-(DA)	44.4	0.31	5.20	37.4	0.52	48.2	0.48	0.25	1.08
19	Q69pAcF/P86 C-(DA)	41.1	0.55	1.94	36.8	0.63	47.4	0.37	0.47	1.10
20	Q69pAcF/R11 9C-(DA)	43.9	0.40	4.32	38.1	0.59	48.9	0.41	0.34	1.12
21	Q69pAcF/N13 2C-(DA)	43.7	0.36	3.79	37.7	0.54	48.8	0.46	0.27	1.15
22	Q690pAcF/I15 0C-(DA)	45.6	0.53	2.25	48.7	0.59	37.8	0.41	0.47	1.11
23	D70pAcF/R11 9C-(DA)	42.7	0.21	7.51	36.1	0.61	46.5	0.39	0.15	1.15
24	D70pAcF/N13 2C-(DA)	42.8	0.26	8.11	36.3	0.64	47.5	0.36	0.18	1.20

Table 3.3.1C. Table of determined discrete FRET rate constants ($N=3$, eq. 3.3.15, 3.3.17) and state-specific average distances for the distribution fits ($N=2$, eq. 3.3.16, 3.3.17).

Measurement		discrete FRET rate constants fit					distribution fit			
		R_{DA} [Å] (x_i)			x_{DOnly}	χ_r^2	$\langle R_{DA} \rangle$ [Å] (x_i)		x_{DOnly}	χ_r^2
		1	2	3			1	2		
1	R8pAcF/Q69C-(DA)	39.7 (0.51)	31.0 (0.39)	51.2 (0.10)	0.19	1.03	37.1 (0.96)	60.0 (0.04)	0.19	1.06
2	R8pAcF/P86C-(DA)	51.9 (0.29)	36.0 (0.36)	43.4 (0.35)	0.20	1.1	37.3 (0.57)	48.7 (0.43)	0.18	1.11
3	K19pAcF/Q69C-(DA)	33.7 (0.26)	41.7 (0.26)	68.7 (0.47)	0.24	1.06	73.7 (0.50)	37.3 (0.50)	0.14	1.05
4	K19pAcF/P86C-(DA)	55.4 (0.46)	46.2 (0.36)	36.2 (0.17)	0.08	1.14	51.8 (0.75)	36.1 (0.25)	0.07	1.14
5	K19pAcF/R119C-(DA)	55.0 (0.46)	46.3 (0.39)	35.2 (0.15)	0.10	1.13	50.9 (0.80)	31.0 (0.20)	0.08	1.15
6	K19pAcF/N132C-(DA)	54.9 (0.26)	45.4 (0.39)	35.7 (0.34)	0.08	1.14	49.8 (0.52)	35.9 (0.48)	0.07	1.15
7	S44pAcF/P86C-(DA)	61.0 (0.19)	49.9 (0.54)	38.8 (0.27)	0.23	1.16	52.7 (0.64)	41.1 (0.36)	0.23	1.16
8***	S44pAcF/R119C-(DA)	55.6 (0.69)		41.3 (0.31)	0	1.11	55.8 (0.72)	40.0 (0.28)	0.22	1.15
9***	S44pAcF/N132C-(DA)	57.7 (0.68)		40.9 (0.32)	0.12	1.21	60.7 (0.61)	42.0 (0.39)	0.38	1.13
10	S44pAcF/I150C-(DA)	65.5 (0.20)	52.8 (0.56)	40.9 (0.23)	0.02	1.07	54.9 (0.75)	39.6 (0.25)	0.03	1.10
11	N55pAcF/Q69C-(DA)	32.7 (0.32)	39.4 (0.56)	49.4 (0.23)	0.27	1.17	37.4 (0.99)	37.4 (0.01)	0.25	1.21

12	<i>N55pAcF/R119C-(DA)</i>	62.4 (0.45)	49.6 (0.34)	38.3 (0.21)	0.38	1.24	58.9 (0.63)	41.3 (0.37)	0.38	1.24
13	<i>N55pAcF/N132C-(DA)</i>	64.7 (0.18)	50.8 (0.55)	39.4 (0.27)	0.03	1.18	53.4 (0.64)	41.0 (0.36)	0.04	1.20
14** *	<i>N55pAcF/I150C-(DA)</i>	53.4 (0.76)		42.4 (0.24)	0.47	1.19	52.1 (0.94)	38.8 (0.06)	0.58	1.19
15** *	<i>K60pAcF/P86C-(DA)</i>	50.7 (0.66)		38.8 (0.34)	0.26	1.12	49.2 (0.78)	36.8 (0.22)	0.39	1.19
16	<i>K60pAcF/R119C-(DA)</i>	46.4 (0.33)	55.7 (0.47)	37.1 (0.20)	0.08	1.14	52.6 (0.71)	37.7 (0.29)	0.06	1.15
17	<i>K60pAcF/N132C-(DA)</i>	49.8 (0.47)	41.4 (0.22)	36.5 (0.31)	0.21	1.04	38.3 (0.53)	49.5 (0.47)	0.17	1.06
18	<i>K60pAcF/I150C-(DA)</i>	49.3 (0.35)	43.3 (0.22)	36.5 (0.43)	0.25	1.07	45.3 (0.55)	33.9 (0.45)	0.20	1.09
19	<i>Q69pAcF/P86C-(DA)</i>	37.7 (0.51)	33.6 (0.15)	47.7 (0.34)	0.47	1.10	37.1 (0.78)	47.2 (0.22)	0.44	1.13
20	<i>Q69pAcF/R119C-(DA)</i>	36.9 (0.47)	45.1 (0.38)	53.6 (0.15)	0.33	1.10	39.7 (0.82)	51.6 (0.18)	0.31	1.12
21** *	<i>Q69pAcF/N132C-(DA)</i>	48.8 (0.54)		37.7 (0.44)	0.27	1.15	37.0 (0.55)	47.3 (0.45)	0.24	1.17
22** *	<i>Q69pAcF/I150C-(DA)</i>	48.7 (0.58)		37.9 (0.42)	0.47	1.12	45.9 (0.79)	33.9 (0.21)	0.45	1.17
23	<i>D70pAcF/R119C-(DA)</i>	38.9 (0.40)	32.0 (0.33)	47.7 (0.27)	0.13	1.07	27.7 (0.65)	42.1 (0.35)	0.07	1.07
24	<i>D70pAcF/N132C-(DA)</i>	39.2 (0.40)	32.0 (0.36)	48.8 (0.24)	0.16	1.10	33.4 (0.67)	45.4 (0.33)	0.14	1.09

*** fitting with the three FRET rate constant model was not successful.

Table 3.3.1D. Table of determined state-specific average distances for the global distribution fit ($N=3$, eq. 3.3.16, 3.3.17) for C_I including statistical uncertainties as described in sections 3.3.1.11, 3.3.1.13 and 3.3.1.15. Distances were used for FPS. We present the weighted residuals for the lowest χ_r^2 for the conformation C_I and the Ca- Ca distances from the best model structure.

Measurement		$C_I, x_1 = 0.441$			Ca- Ca X-ray* [Å]	$\langle R_{DA} \rangle$ X-ray [Å]	$w.res.$ [Å/Å]	χ_r^2	$xDOnly$
		$\langle R_{DA} \rangle$ [Å]	err-[Å]	err+[Å]					
1	<i>R8pAcF/Q69C-(DA)</i>	36.0	2.44	2.4	14.5	35.6	-0.2	1.08	0.49
2	<i>R8pAcF/P86C-(DA)</i>	48.6	3.13	3.1	26.9	47.8	-0.3	1.07	0.47
3	<i>K19pAcF/Q69C-(DA)</i>	32.0	3.09	3.1	19.7	37.8	1.9	1.06	0.48
4	<i>K19pAcF/P86C-(DA)</i>	51.1	3.01	3.0	35.7	52.4	0.4	1.13	0.07
5	<i>K19pAcF/R119C-(DA)</i>	52.2	3.96	4.0	37.5	52.6	0.1	1.19	0.08
6	<i>K19pAcF/N132C-(DA)</i>	51.5	5.61	5.6	35.3	48.5	-0.5	1.16	0.22
7	<i>S44pAcF/P86C-(DA)</i>	54.1	4.14	4.1	39.0	51.7	-0.6	1.12	0.41
8	<i>S44pAcF/R119C-(DA)</i>	58.2	4.16	4.2	44.0	58.2	0.0	1.18	0.23
9	<i>S44pAcF/N132C-(DA)</i>	59.3	2.86	2.9	44.9	60.2	0.3	1.08	0.48
10	<i>S44pAcF/I150C-(DA)</i>	59.1	5.79	5.8	38.9	58.8	-0.1	1.08	0.75
11	<i>N55pAcF/Q69C-(DA)</i>	36.8	3.80	3.8	20.7	35.0	-0.5	1.18	0.56

12	N55pAcF/R119C-(DA)	61.5	3.55	3.6	46.7	63.3	0.5	1.19	0.73
13	N55pAcF/N132C-(DA)	53.9	3.73	3.7	45.5	61.9	2.2	1.06	0.22
14	N55pAcF/I150C-(DA)	53.1	4.61	4.6	38.4	54.5	0.3	1.19	0.58
15	K60pAcF/P86C-(DA)	51.6	4.36	4.4	32.9	49.1	-0.6	1.11	0.66
16	K60pAcF/R119C-(DA)	53.9	4.76	4.8	36.2	52.7	-0.3	1.10	0.17
17	K60pAcF/N132C-(DA)	50.9	4.67	4.7	34.9	50.5	-0.1	1.05	0.40
18	K60pAcF/I150C-(DA)	39.4	3.85	3.9	26.5	39.3	0.0	1.06	0.45
19	Q69pAcF/P86C-(DA)	36.9	2.94	2.9	22.1	34.4	-0.8	1.13	0.44
20	Q69pAcF/R119C-(DA)	39.2	3.58	3.6	27.9	42.4	0.9	1.06	0.60
21	Q69pAcF/N132C-(DA)	42.5	3.94	3.9	30.4	46.9	1.1	1.12	0.53
22	Q69pAcF/I150C-(DA)	53.2	4.35	4.3	24.5	46.3	-1.6	1.06	0.67
23	D70pAcF/R119C-(DA)	38.9	3.17	3.2	25.8	38.3	-0.2	1.04	0.30
24	D70pAcF/N132C-(DA)	39.2	3.58	3.6	27.7	40.8	0.5	1.05	0.40

* C α -C α distance and R_{DA} of best fit model (178L for C₁), ** definition of $w.res = \frac{R_{DA,model} - R_{DA,experiment}}{\Delta R_{DA}^2(\kappa^2, k_{FRET})}$

Table 3.3.1E. Table of determined state-specific average distances for the global distribution fit ($N=3$, eq. 3.3.16, 3.3.17) for C₂ and C₃ including statistical uncertainties as described in sections 3.3.1.11, 3.3.1.13 and 3.3.1.15. Distances were used for FPS. We present the weighted residuals for the lowest χ_r^2 for each conformation C₂ and C₃ and the C α -C α distances from the best model structure: 1LWK for C₂, 209L for C₃

Measurement		C ₂ , x ₂ = 0.318			C α -C α X-ray * [Å]	$\langle R_{DA} \rangle$ X-ray [Å]	$w.res$ [Å/Å]	C ₃ , x ₃ = 0.241			C α -C α X-ray * [Å]	$\langle R_{DA} \rangle$ X-ray [Å]	$w.res$ [Å/Å]
		$\langle R_{DA} \rangle$ [Å]	err- [Å]	err+ [Å]				$\langle R_{DA} \rangle$ [Å]	err- [Å]	err+ [Å]			
1	R8pAcF/Q69C-(DA)	36.0	2.68	2.7	13.7	31.7	-1.6	44.3	3.87	3.9	13.7	31.6	-3.3
2	R8pAcF/P86C-(DA)	42.6	2.92	2.9	26.3	45.0	0.8	34.1	2.62	2.6	29.3	45.7	4.4
3	K19pAcF/Q69C-(DA)	41.6	3.26	3.3	20.3	40.1	-0.5	54.8	4.54	4.5	20.7	40.7	-3.1
4	K19pAcF/P86C-(DA)	52.5	3.21	3.2	33.8	51.9	-0.2	35.8	2.41	2.4	37.8	55.1	8.0
5	K19pAcF/R119C-(DA)	52.2	4.05	4.0	32.5	47.9	-1.1	41.1	3.36	3.4	32.5	46.5	1.6
6	K19pAcF/N132C-(DA)	43.6	4.83	4.8	27.6	40.0	-0.7	30.8	3.55	3.5	24.2	35.3	1.3
7	S44pAcF/P86C-(DA)	48.1	3.75	3.7	37.4	49.5	0.4	39.5	3.23	3.2	41.0	54.2	4.6
8	S44pAcF/R119C-(DA)	50.1	3.57	3.6	40.0	51.8	0.5	25.1	2.00	2.0	37.3	45.3	10.1
9	S44pAcF/N132C-(DA)	55.4	2.69	2.7	38.8	51.2	-1.5	39.0	2.20	2.2	34.2	46.7	3.5
10	S44pAcF/I150C-(DA)	49.6	4.83	4.8	35.1	52.7	0.6	22.8	2.38	2.4	35.3	52.3	12.4

11	N55pAcF/Q69C-(DA)	36.8	3.94	3.9	21.4	34.7	-0.5	41.0	4.69	4.7	21.3	35.4	-1.2
12	N55pAcF/R119C-(DA)	52.3	2.84	2.8	44.1	60.2	2.8	42.8	2.57	2.6	42.4	56.2	5.2
13	N55pAcF/N132C-(DA)	53.9	3.84	3.8	41.7	56.9	0.8	40.6	3.09	3.1	37.1	52.4	3.8
14	N55pAcF/I150C-(DA)	53.1	4.70	4.7	36.7	52.8	-0.1	46.7	4.29	4.3	36.6	53.0	1.5
15	K60pAcF/P86C-(DA)	51.7	4.44	4.4	34.9	51.2	-0.1	38.0	3.42	3.4	37.9	52.2	4.1
16	K60pAcF/R119C-(DA)	53.9	4.85	4.8	37.7	54.6	0.2	40.6	3.81	3.8	39.2	55.1	3.8
17	K60pAcF/N132C-(DA)	42.5	4.03	4.0	36.2	52.0	2.4	37.3	3.70	3.7	33.0	50.0	3.4
18	K60pAcF/I150C-(DA)	43.6	4.24	4.2	28.4	42.8	-0.2	54.4	5.47	5.5	28.8	43.9	-1.9
19	Q69pAcF/P86C-(DA)	36.9	3.12	3.1	22.7	35.6	-0.4	46.9	4.38	4.4	25.6	37.7	-2.1
20	Q69pAcF/R119C-(DA)	42.2	3.86	3.9	28.0	42.8	0.2	50.2	4.77	4.8	27.5	38.6	-2.4
21	Q69pAcF/N132C-(DA)	48.6	4.52	4.5	30.1	46.7	-0.4	36.0	3.49	3.5	25.9	44.2	2.3
22	Q69pAcF/I150C-(DA)	44.6	3.74	3.7	24.0	47.1	0.7	38.7	3.41	3.4	23.9	45.7	2.1
23	D70pAcF/R119C-(DA)	33.6	3.39	3.4	25.4	37.4	1.1	47.5	5.53	5.5	25.0	32.4	-2.7
24	D70pAcF/N132C-(DA)	48.6	4.35	4.3	26.7	40.3	-1.9	33.7	3.16	3.2	22.4	38.5	1.5

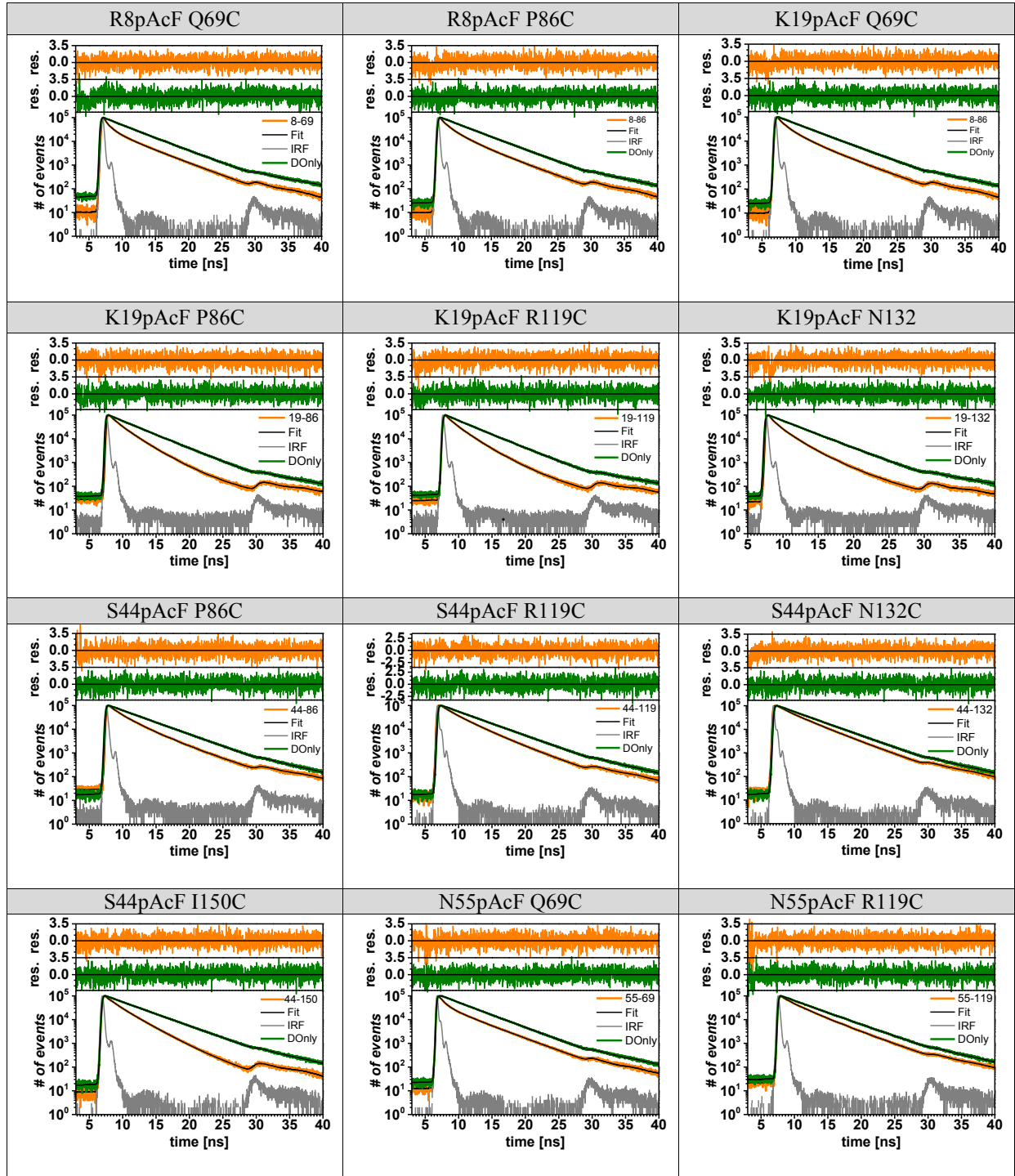
Table 3.3.1F. Results of discrete distance fit model for functional variants of the S44/I150C FRET pair.
Globally fitted parameters are highlighted in gray cells (Eq. 3.3.15, 3.3.17 and 3.3.18).

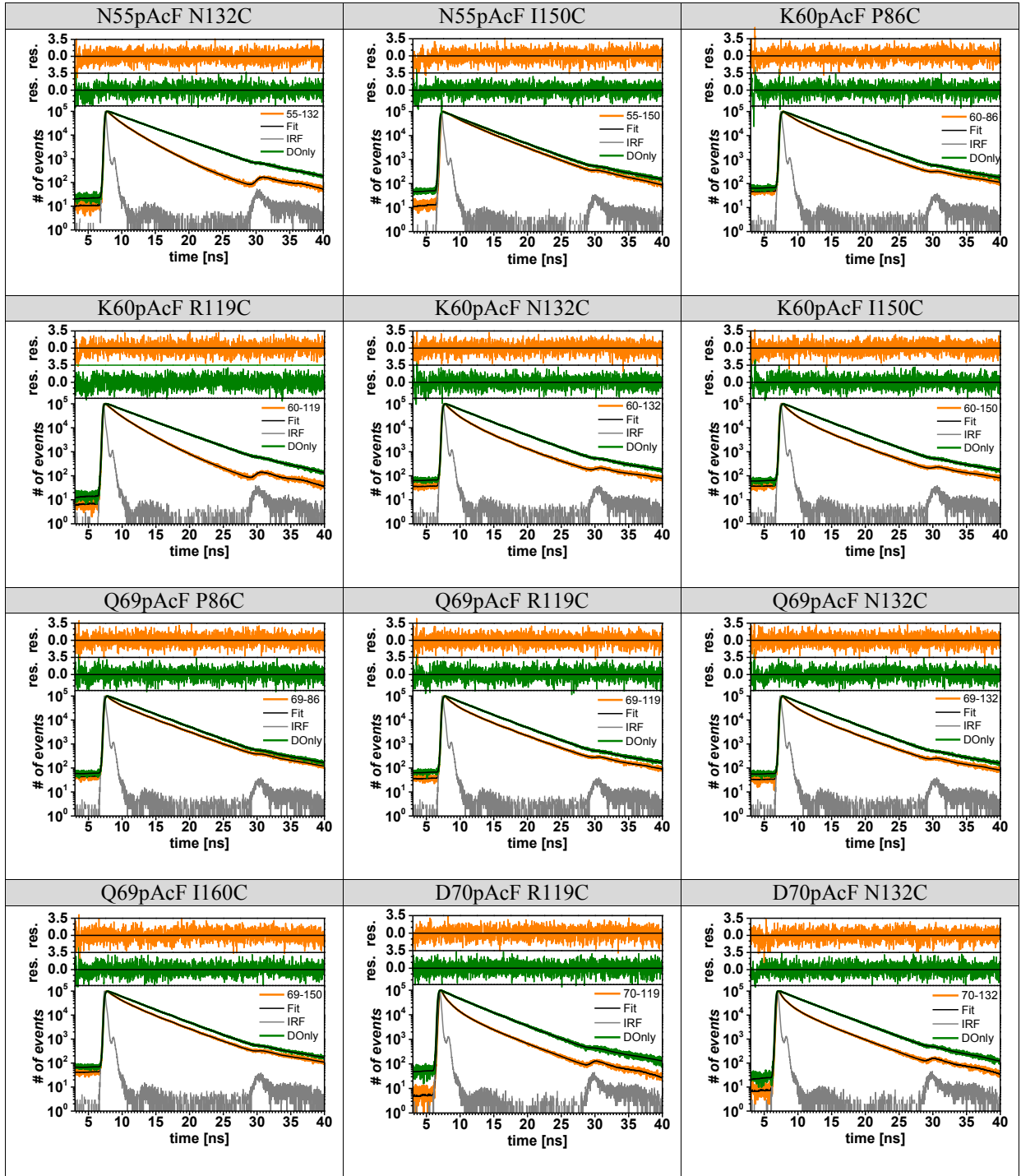
Samples (showing aa #'s)	R_{DA1} [Å]	x_1^\dagger	\pm err. (%)	R_{DA2} [Å]	x_2^\dagger	\pm err. (%)	R_{DA3} [Å]	x_3^\dagger	\pm err. (%)	$x_{D(0)}^*$	χ^2_r
S44pAcF/I150C-(DA)(-)	65.1	0.25	8.0	51.7	0.55	8.0	38.8	0.20	3.0	0.01	1.06
S44pAcF/I150C-(DA)(+)**	65.1	0.25	12.3	51.7	0.57	12.3	38.8	0.18	3.9	-	1.52**
T26E/S44pAcF/I150C-(DA)(-)**	65.1	0.37	4.7	51.7	0.35	3.0	34.9	0.28	9.0	0.62	1.21
T26E/S44pAcF/I150C-(DA)(+)**	65.1	0.20	1.9	51.7	0.28	1.0	34.9	0.52	13.0	0.74	1.08
E11A/S44C/I150C-(DA/AD)(-)**	65.1	0.75	7.8	51.7	0.12	7.8	38.8	0.13	11.8	-	1.98**
E11A/S44C/I150C-(DA/AD)(+)**	65.1	0.56	4.9	51.7	0.27	4.9	38.8	0.17	17.4	-	2.00**
R137E/S44pAcF/I150C-(DA)	59.3	0.52	7.3	49.3	0.37	2.6	36.2	0.11	9.1	0.24	1.07

[†] 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 χ^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).

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 (Figure 3.3.2C, D).

A





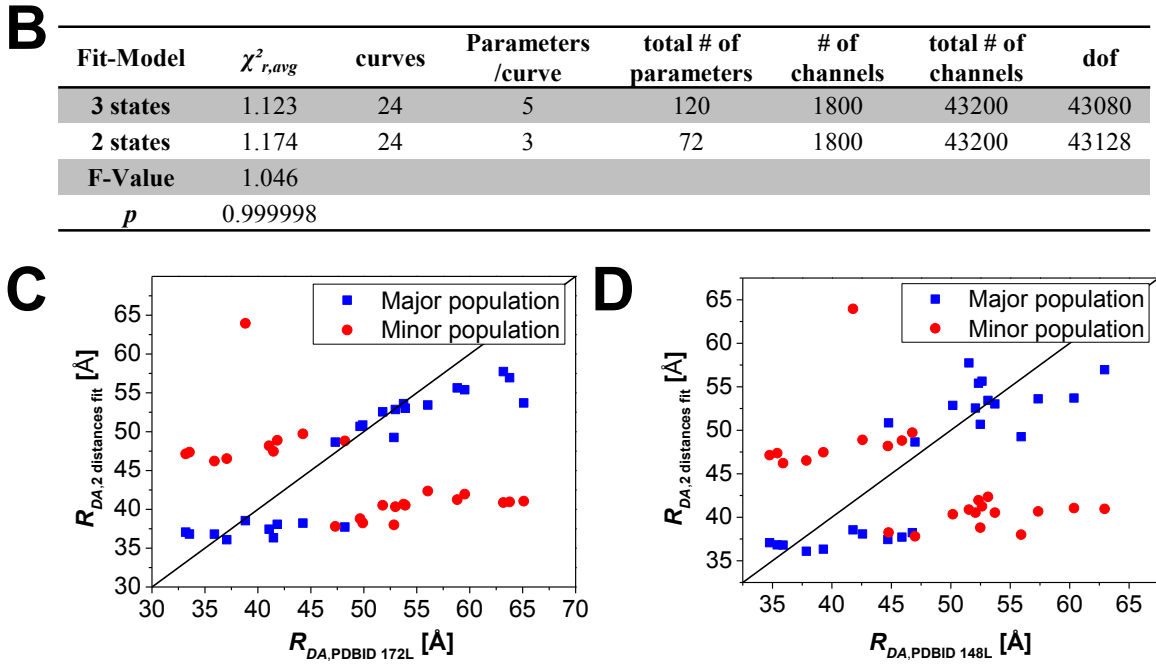


Figure 3.3.2 eTCSPC results of T4L. (A) Fit (black) of the experimental data of double-labeled sample (orange) and respective Donor-only labeled sample (green), weighted residuals are shown on top. Fit parameters are given in Table 3.3.1A (Only) and 3.3.1D-E (double-labeled sample). Instrument response function (IRF) is shown in gray. (B) Values used and obtained for the statistical evaluation (C) Fitted distances of two distributed states 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 (D) Same as (C), only for the model X-ray structure of the closed state (PDBID 148L).

To evaluate the statistical significance in the difference on χ^2_r between the two states and three states fit, we calculated the ratio $\chi^2_{r-rel} = \chi^2_r(3states) / \chi^2_r(2states)$ (*F-value*); then, we calculated the probability p , which determines the odds that the three states model describes the data better than the two states model. To do so, we calculated the cumulative distribution function (CDF) for each single data set by Excel (Microsoft):

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

The degrees of freedom (dof) per data set are three (two states fit) and five (three states fit), respectively (Figure 3.3.2C).

In a second step, we treated the whole data as a single set and used the global, average χ^2_r to determine the confidence level with which the two states model has to be refused. The values are summarized in Figure 3.3.2B. 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):

$$\text{F.INV} = [0.99, \text{dof}(2 \text{ states}), \text{dof}(3 \text{ states})] = 1.046$$

3.3.2.2. *Species Cross Correlation function –(DA) and –(AD) labeled samples*

Theoretically, the species cross-correlation function (sCCF), as defined in Eq. 3.3.22, 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 (47). 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. 3.3.28-3.3.29, 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 Figure 3.3.3A, C. The parameters used on the decay generation are shown in the caption. From these patterns, the filters $f_j^{(i)}$ (Eq. 3.3.23) were calculated. These are shown in panels B and D of Figure 3.3.3. These filters are then used to compute the sCCF by multiplying each photon and weighting its contribution to each state as in Eq. 3.3.22. The patterns that are shown in Figure 3.3.3 correspond to only 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 (Figure 3.2.3B, 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 (Figure 3.3.3E). For the functional mutants (E11A, T26E, R137E), we show the sCCF (Figure 3.3.3F-H).

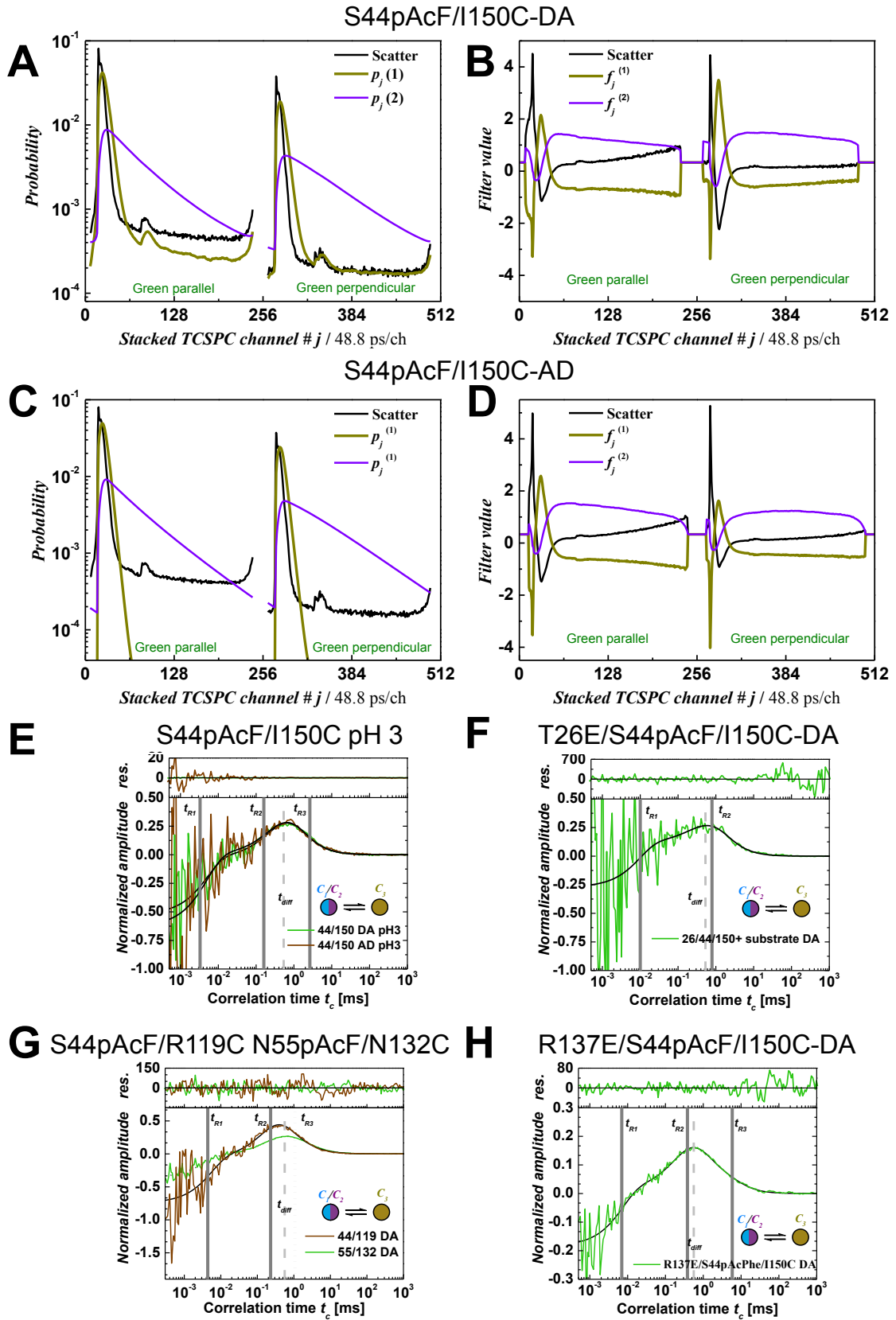


Figure 3.3.3 fFCS results. (A) Generated decays for two pseudo-species presented for the S44pAcF/I150C-DA in addition to the scatter profile. The parameters of the decay generation for the first pseudo-species were $\tau_1 = 0.25$ ns lifetime, and rotational correlation of $\rho_1 = 3.3$ ns. The second 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. 3.3.24

using the decays from graph (A). A similar process was done for each fFCS data set. First, the decays are generated; then, 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 (Figure 3.2.3B, 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. (3.3.25-3.3.26) 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 ($t_{R1} = 10$ μ s, 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 ($t_{R1} = 4 \pm 2.4$ μ s, $t_{R2} = 230 \pm 28$ μ s). The variant N55pAcF/N132C-(DA) requires an additional rate $t_{R3} \sim 1.1$ ms. **(H)** sCCF for variant R137E/S44pAcF/I150C-(DA). Three relaxation times were needed to fit the curve ($t_{R1} = 7$ μ s, $t_{R2} = 0.38$ ms and $t_{R3} = 5.84$ ms). The diffusion time was fixed to $t_{diff} = 0.54$ ms.

3.3.2.3. Additional SMD and fFCS

To test for possible influences of the dyes on the protein, two distinct labeling configurations (DA) and (AD) were prepared as previously described. In the sub- μ s 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 Figure 3.3.6A, to the –(AD) labeling scheme shown in Figure 3.3.4, 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 (Figure 3.3.4B). This resembles the accumulation found in the sample T26E/S44pAcF/I150C–(DA) (Figure 3.2.4G, 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 (Figure 3.2.3B, 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- μ s 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.

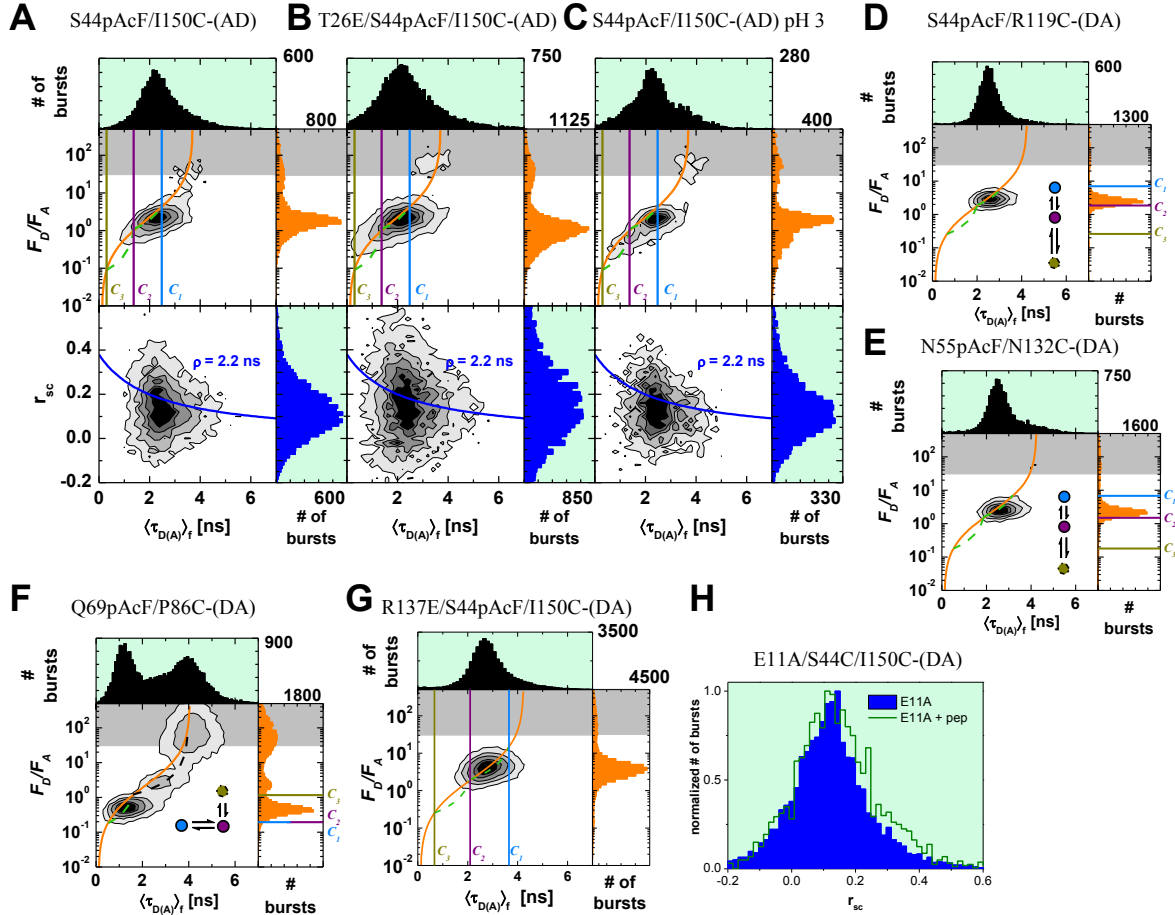


Figure 3.3.4 MFD analysis of selected samples. (A) MFD histogram of S44pAcF/I150C- $-(AD)$ labeled T4L. Two dimensional histogram F_D/F_A vs. lifetime of donor in the presence of acceptor $\langle \tau_{D(A)} \rangle_f$, and anisotropy vs. $\langle \tau_{D(A)} \rangle_f$ for S44pAcF/I150C- $-(AD)$. One dimensional projections for F_D/F_A , $\langle \tau_{D(A)} \rangle_f$ and anisotropy are also shown. Static FRET line 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- $-(DA)$ an accumulation of burst at high F_D/F_A ratio is seen. This can be expected when the acceptor is no longer functional. (G) **Functional mutant of T4L.** MFD histogram for R137E/S44pAcPh/I150C- $-(DA)$. (H) **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.

Additional MFD histograms for other mutants mentioned in the main text are shown in Figure 3.3.4D-G. A summary of the ensemble or sub-ensemble fits for these mutants is shown in Table 3.3.1.

3.3.2.4. 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. Prior to use, the purchased peptidoglycan (Sigma-Aldrich, Switzerland) was purified as described by Maeda in 1980 (122) to remove minor fluorescent impurities. Double-labeled mutants (1 μ M) were incubated with 3 mg/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 3.3.5. Figure 3.3.5A 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 Alexa488 (495 nm). This is shown in Figure 3.3.5B. After 260 minutes this mutant is fully saturated with the substrate.

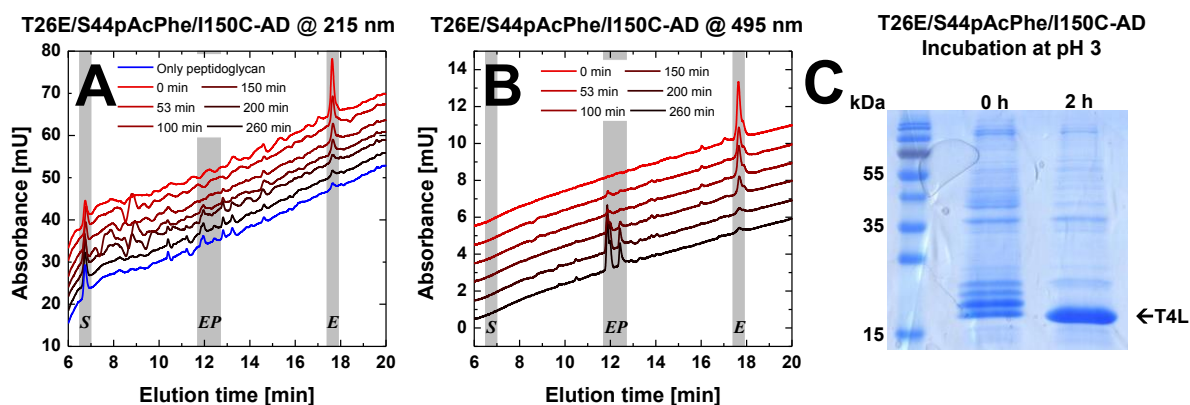


Figure 3.3.5 T4L binding to peptidoglycan as observed by reverse phase HPLC and cleavage at low pH. (A) The elution profile measured at 215 nm of reverse phase chromatography as described in Section 3.3.1.3 for the T26E/S44pAcF/I150C-(AD). Samples were taken at different times during the incubation with 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 Alexa488. 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.

3.3.2.5. Analyzing the kinetic network of conformational states in T4L

Detection of distinct C_3 species

The eTCSPC fluorescence decay of S44pAcF/I150C-(DA) was fit with a model containing three different conformational states (C_1 , C_2 and C_3). The total population of C_3 corresponds to 23 % (Table 3.3.1F). From the single molecule MFD histograms it was clear to observe burst accumulation at the location where C_3 lies. To quantify the amount of bursts corresponding to this, we computed the area under the curve for the region of $10^{-2} < F_D/F_A < 0.3$ (Figure 3.3.6E) corresponding to 564 burst of the total 10139 burst of all single molecule events (subtracting 893 bursts from molecules missing an active acceptor). Thus, the burst accumulation of this state is 5.3% of the total number of bursts. We called this population C_{3t} because it is a static accumulation of the population C_3 observed by eTCSPC. In order to account for the missing 18 % of C_3 , there has to be an additional population, which exchanges with C_1 and C_2 at timescales faster than the burst duration. We called this population C_{3d} . Therefore, the total contribution equals the sum of the static plus the dynamic subpopulations of C_3 ($C_3 = 23 \% = C_{3t} + C_{3d} = 5 \% + 18 \%$). Because in fFCS we only observe two relaxation times from μ s 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.

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 ($\sim 5\%$) 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 $\sim 4 \mu$ s and $\sim 230 \mu$ s.

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



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



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. 3.3.34.

$$K = \begin{pmatrix} -k_{12} & k_{21} & 0 \\ k_{12} & -(k_{21} + k_{23}) & k_{32} \\ 0 & k_{23} & -k_{32} \end{pmatrix} \tag{3.3.36}$$

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

$$\frac{1}{t_{R1,2}} = \frac{1}{2} \cdot (k_{12} + k_{21} + k_{23} + k_{32}) \pm \sqrt{(k_{12} + k_{21} + k_{23} + k_{32})^2 - 4 \cdot (k_{12}k_{23} + k_{12}k_{32} + k_{21}k_{32})} \tag{3.3.37}$$

The time evolution of the system on Eq. 3.3.34 is defined by

$$\frac{d}{dt}[C_i](t) = K_{ij}[C_j](t) \tag{3.3.38}$$

Equation 3.3.37 has an analytical solution on the form of

$$[C_i](t) = C_0 \exp(K_{ij}t) \tag{3.3.39}$$

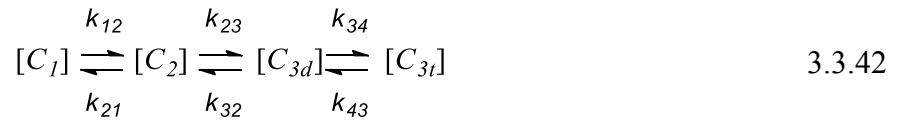
where C_0 is the i -th eigenvector. At equilibrium, or $t \rightarrow \infty$, the equilibrium fractions for each conformer can be obtained analytically and are given by

$$[C_i] = \begin{pmatrix} \frac{k_{21} \cdot k_{32}}{k_{21} \cdot k_{32} + k_{12} \cdot (k_{23} + k_{32})} \\ \frac{k_{12} \cdot k_{32}}{k_{21} \cdot k_{32} + k_{12} \cdot (k_{23} + k_{32})} \\ \frac{k_{12} \cdot k_{23}}{k_{21} \cdot k_{32} + k_{12} \cdot (k_{23} + k_{32})} \end{pmatrix} \tag{3.3.40}$$

Note that $[C_3] = 1 - ([C_1] + [C_2])$. These fractions are obtained by fluorescence decay analysis as done in Section 3.3.1.11. 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:

$$\begin{aligned}
 k_{12}^{(\pm)} &= \frac{[C_2] \cdot \left(\frac{1}{t_{R1}} + \frac{1}{t_{R2}} \right) \pm \left([C_2] \cdot \left(\frac{4 \cdot ([C_1] - 1) \cdot ([C_1] + [C_2])}{t_{R1} \cdot t_{R2}} + [C_2] \cdot \left(\frac{1}{t_{R1}} + \frac{1}{t_{R2}} \right)^2 \right) \right)^{1/2}}{2 \cdot ([C_1] + [C_2])} \\
 k_{12}^{(\pm)} &= \frac{[C_2] \cdot \left(\frac{1}{t_{R1}} + \frac{1}{t_{R2}} \right) \pm \left([C_2] \cdot \left(\frac{4 \cdot ([C_1] - 1) \cdot ([C_1] + [C_2])}{t_{R1} \cdot t_{R2}} + [C_2] \cdot \left(\frac{1}{t_{R1}} + \frac{1}{t_{R2}} \right)^2 \right) \right)^{1/2}}{2 \cdot ([C_1] + [C_2])} \\
 k_{12}^{(\pm)} &= \frac{[C_2] \cdot \left(\frac{1}{t_{R1}} + \frac{1}{t_{R2}} \right) \pm \left([C_2] \cdot \left(\frac{4 \cdot ([C_1] - 1) \cdot ([C_1] + [C_2])}{t_{R1} \cdot t_{R2}} + [C_2] \cdot \left(\frac{1}{t_{R1}} + \frac{1}{t_{R2}} \right)^2 \right) \right)^{1/2}}{2 \cdot ([C_1] + [C_2])} \\
 k_{32}^{(\pm)} &= \frac{-[C_2] \cdot \left(\frac{1}{t_{R1}} + \frac{1}{t_{R2}} \right) \pm \left([C_2] \cdot \left(\frac{4 \cdot ([C_1] - 1) \cdot ([C_1] + [C_2])}{t_{R1} \cdot t_{R2}} + [C_2] \cdot \left(\frac{1}{t_{R1}} + \frac{1}{t_{R2}} \right)^2 \right) \right)^{1/2}}{2 \cdot ([C_1] - 1)}
 \end{aligned} \tag{3.3.41}$$

To complete the model we need to add the static fraction of $\sim 5\%$. 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



Where $k_{34} = 0.003 \text{ ms}^{-1}$ and $k_{43} = 0.008 \text{ ms}^{-1}$ 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 simulated data are shown in Figure 3.3.6 and correspond to the experimental data shown in Figure 3.2.3.

Simulation of the FRET data in complex kinetic schemes

To describe the experimental 2D histogram a four-state scheme was used. First, we calculated FRET histograms (123) where a kinetic model with discrete conformations was assumed. The transition between the states is described by rate equations. The probability for the system to

be in state i at time t , $\mathbf{p}_i(t)$, satisfies a set of rate equations, which can be written in matrix notation as:

$$\frac{d\mathbf{p}}{dt} = \mathbf{K} \cdot \mathbf{p} \quad 3.3.43$$

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

$$\langle E \rangle = \frac{\sum t_i \cdot E_i}{t_{burst}} \quad 3.3.44$$

and

$$\langle E \rangle = \frac{\sum t_i \cdot E_i}{t_{burst}} \quad 3.3.45$$

where $t_i(\mathbf{K})$ is time spent by a molecule in state i within the duration of the burst and depends on the transition rate matrix \mathbf{K} ; E_i is the FRET efficiency of the i -th state; t_{burst} is the duration of the burst and τ_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 by each molecule in different states were calculated using Gillespie algorithm for continuous-time Markov Chain. Then, the 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 \mathbf{p} and the rate matrix \mathbf{K} (resulting into the equilibrium fractions for the state i , $\mathbf{p}_{eq,i}$) and the experimental observables, E and τ , used in the simulations are shown in the Table 3.3.2A. For plotting, E was 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 (Figure 3.3.6).

The resulting 1D and 2D histograms were compared to the experimental data, yielding a χ^2 parameter for each simulation and histogram. To test the significance of the difference in χ^2 , we performed F-test as described above. The resulting values are combined in the Table 3.3.2B.

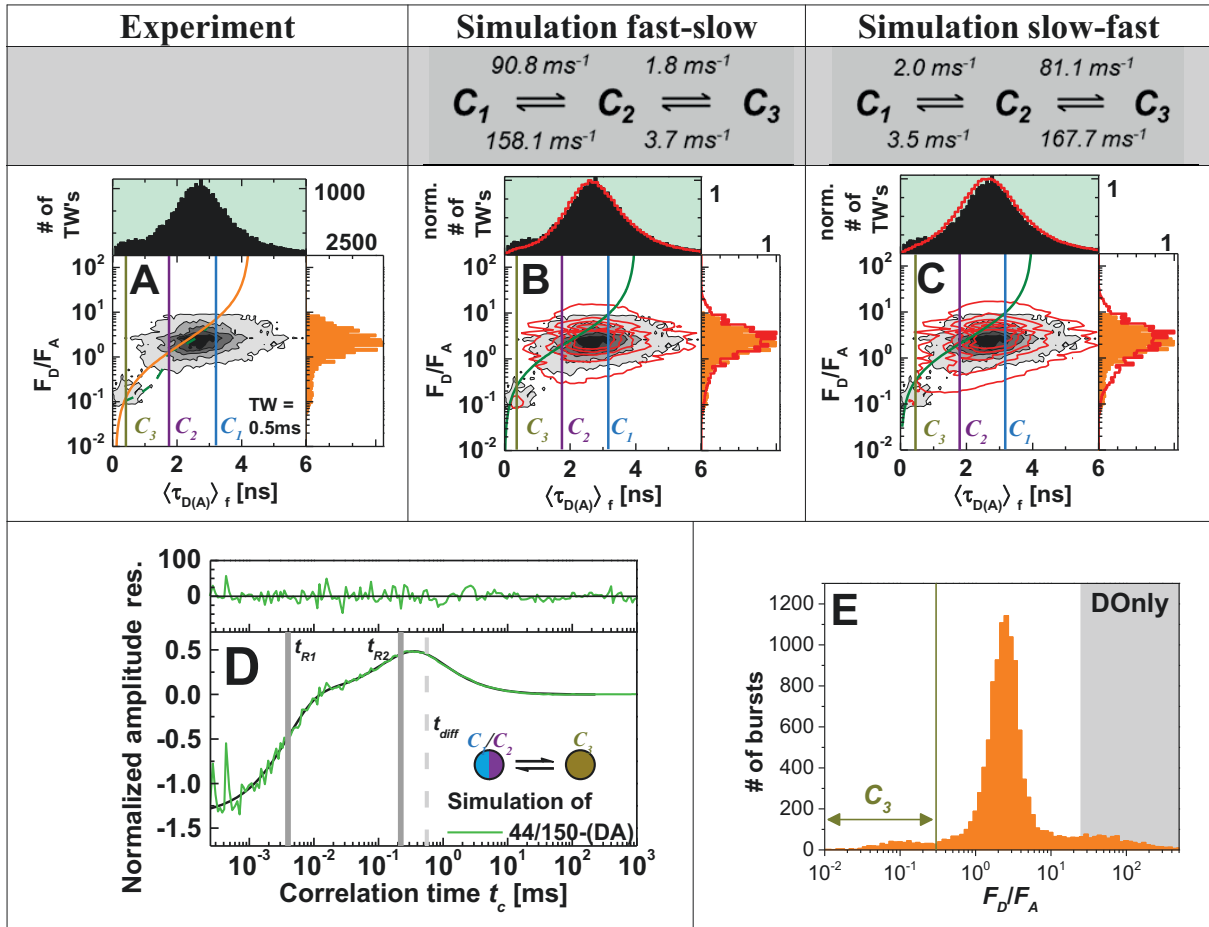


Figure 3.3.6 Brownian Dynamic Simulations of S44pAcF/I150C-(DA). (A) Two dimensional histograms from smFRET analysis (F_D/F_A vs. fluorescence lifetime ($\langle \tau_{D(A)} \rangle_f$) and anisotropy r_{sc} vs. $\langle \tau_{D(A)} \rangle_f$) of the raw data from the S44pAcF/I150-(DA) variant at pH 7.5 selected with 0.5 ms time-windows (TW's). FRET lines, static and dynamic are shown as orange solid and 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 Figure 3.2.3B, C was processed as the experimental data. Simulated parameters ($\langle B_G \rangle$, $\langle B_R \rangle$, α , 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). (D) 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 μ s and 220 μ s, consistent with our input parameters. (E) One dimensional histogram of the raw data from the S44pAcF/I150-(DA) variant at pH 7.5 analyzed in burstwise mode to illustrate the region of C_3 , DOnly and the dynamically mixed state.

Table 3.3.2A States description for the vector \mathbf{p} , equilibrium fraction vector \mathbf{p}_{eq} and the rate matrix \mathbf{K} .

Kinetic state i	State name	Efficiency E	Fluorescence lifetime τ_i 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

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

Table 3.3.2B. Parameters used in the evaluation of the statistical significance of different simulations.

Type of analysis	χ^2 fast-slow	χ^2 slow-fast	Degrees of freedom	F-value	$p(C_1-C_2$ fast vs. C_1-C_2 slow)
1D $\langle\tau_{D(A)}\rangle_f$ histogram	68.6	187.2	10	2.72	1
1D E histogram	85.1	95.5	10	1.12	0.734
2D $\langle\tau_{D(A)}\rangle_f$ vs E histogram	0.4551	0.6069	10	1.33	1

Table 3.3.2C. Calculated reaction rates for several variants using Eq. 3.3.37. Confidence intervals (2σ) 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: $t_{R1} = 4 \pm 2.3$ μ s; $t_{R2} = 230 \pm 28.4$ μ s for 44/150-(DA) and 11/44/150-(DA)+pept. $t_{R1} = 10$ μ s; $t_{R2} = 790$ μ s (Figure 3.3.4F) 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 (Figure 3.2.4E). Only the amino acid number of the mutagenesis is shown.				

To estimate our errors on determining the rates we considered the 2σ confidence interval in determining the population fractions (Table 3.3.1) and the 2σ confidence interval in determining the relaxation times by ffCS (Table 3.3.2). Taking those extremes we estimated the error and computed the reaction rate constant for Figure 3.2.6 in main text according to Eq. 3.3.41.

3.3.2.6. Challenges of smFRET measurements and their solutions

Most of the potential problems with smFRET come from the complexities associated with the labels. 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 fluorescence 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 fluorescence 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} (Figure 3.3.7) 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 (Figure 3.3.7).

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 (120). Table 3.3.3A-C summarizes the residual anisotropies (r_∞) of D - donor, A - acceptor and A(D) - the sensitized by FRET emission of acceptor that were used to calculate dye order parameters and κ^2 distributions (Table 3.3.3) according to refs. (53, 120) (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 (69).

6) The existence and the population fraction of the new conformational state C_3 is consistent across our library of mutants (Table 3.3.1) $x_3 = 27 \pm 10$ %. 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 restraints are within the expected uncertainty of our tools. In addition, the kinetics found in all our variants are consistent with two global relaxation times ($t_{R1} = \sim 4$ μ s, $t_{R2} = \sim 230$ μ s) 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.

Table 3.3.3A Analysis of time-resolved fluorescence anisotropies $r(t)$ for donor only labeled samples [a] obtained by ensemble time-resolved fluorescence decays as described in 3.3.1.6

Samples (showing aa #'s)	$r_{1,D}$	$\rho_{1,D}$	$r_{2,D}$	$\rho_{2,D}$	Σr
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:

$r(t) = r_1 \exp(-t / \rho_1) + r_2 \exp(-t / \rho_2)$ with $r_1 + r_2 \leq r_0$. For Alexa488 the fundamental anisotropy $r_{0,D}$ is 0.375, for Alexa647 $r_{0,A}$ is 0.39 and for FRET-sensitized anisotropy decay $r_{0,A(D)}$ is 0.38.

Table 3.3.3B 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 3.3.1.6.

Samples (showing aa #'s)	$r_{1,A}$	$\rho_{1,A}$	$r_{2,A}$	$\rho_{2,A}$	Σr
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 3.3.3C 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 3.3.1.6 except for ¹ and ².

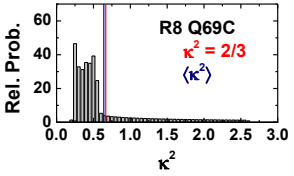
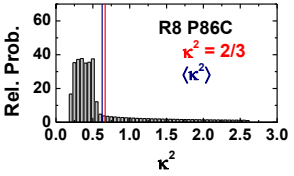
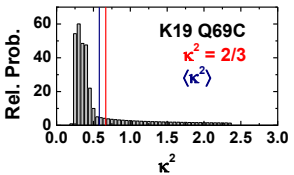
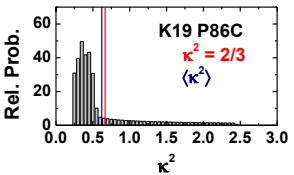
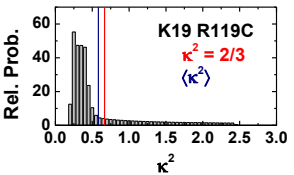
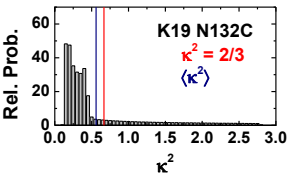
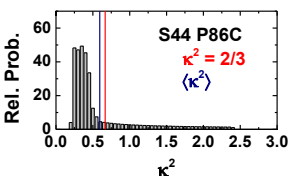
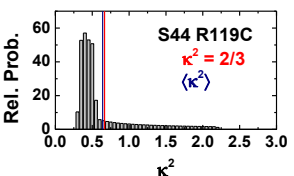
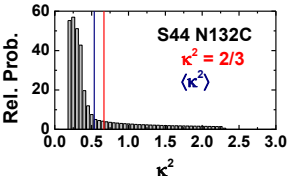
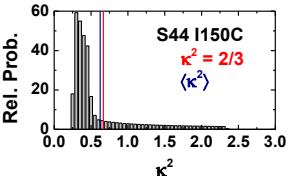
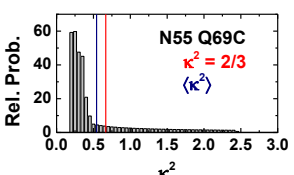
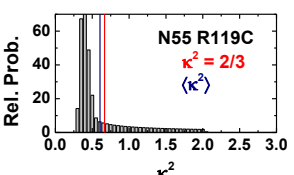
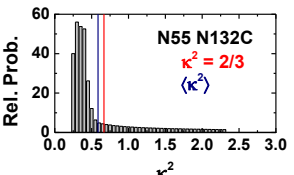
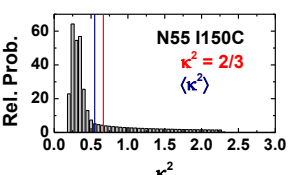
Samples (showing aa #'s)	$r_1, A(D)$	$\rho_{1,A(D)}$	$r_2, A(D)$	$\rho_{2,A(D)}$	Σr
8/69-(DA)	0.025	0.509	0.041	∞	0.066
8/86-(DA)	0.032	0.438	0.049	380	0.081
19/69-(DA)	0.081	0.398	0.105	48.063	0.186
19/86-(DA) ¹	0.209	0.756	0.0561	19.901	0.2651
19/119-(DA)	0.041	0.512	0.091	202	0.132
19/132-(DA)	0.1	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	∞	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.

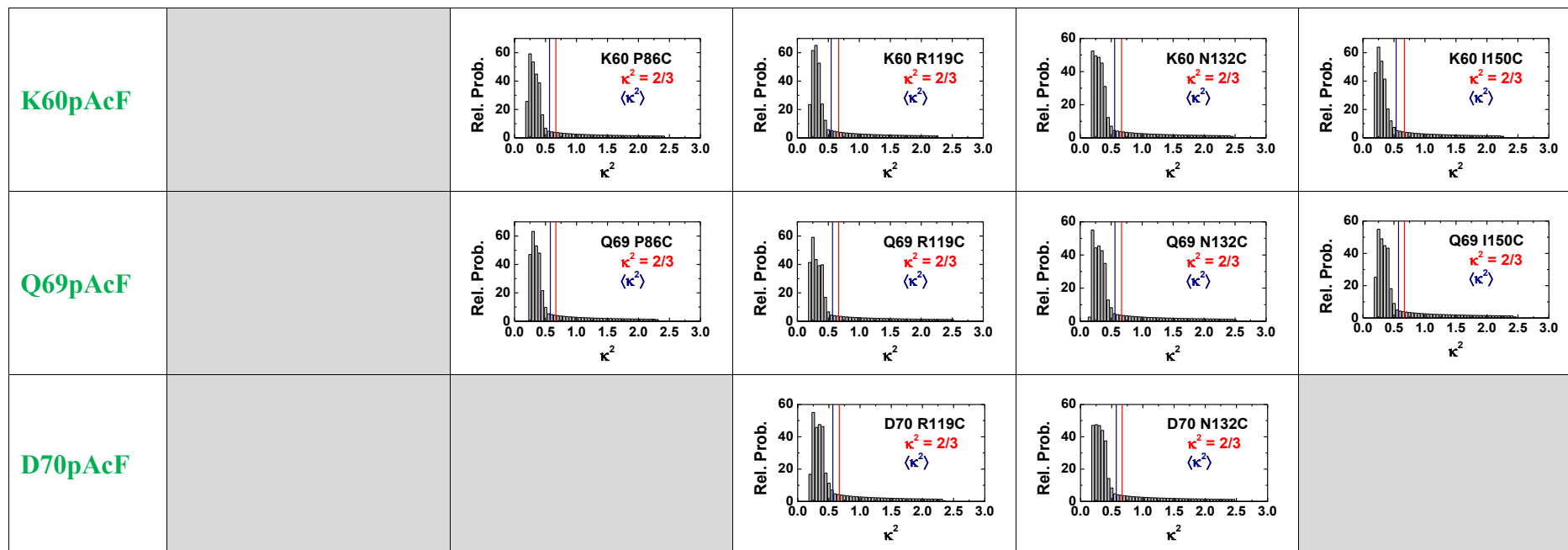
² eTCSPC data not available. Considering variants with a very high FRET efficiency, no satisfactory anisotropy decays from subensemble single-molecule MFD data were obtainable due the short donor fluorescence lifetime. Here, steady-values anisotropies were taken as upper limit from single-molecule MFD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

STRUCTURE AND FUNCTION OF T4L UNDER NATIVE CONDITIONS

Table 3.3.3D κ^2 distributions for the 24 DA samples. Donor positions are labeled on green and acceptor positions on red. The mean κ^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 κ^2 distribution adds to the uncertainty on our distances, which is considered as previously described in Eq. 3.3.32.

Acceptor / Donor	Q69C	P86C	R119C	N132C	I150C
R8pAcF					
K19pAcF					
S44pAcF					
N55pAcF					

STRUCTURE AND FUNCTION OF *T4L* UNDER NATIVE CONDITIONS



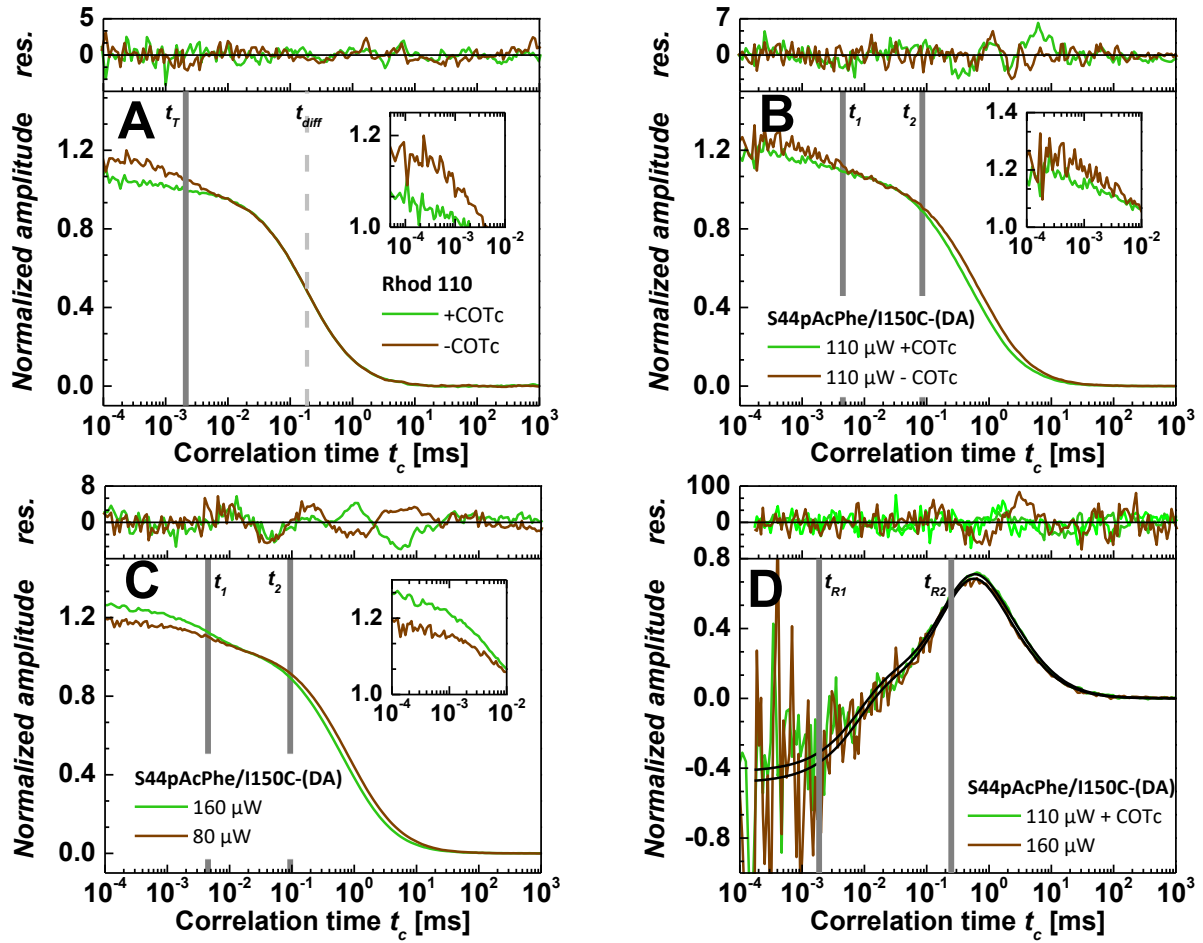


Figure 3.3.7 Triplet or dark states do not influence the sCCF on the variant S44pAcF/I150C-(DA). (A) The addition of the triplet quencher COTc into Rhod110 solution significantly reduces triplet fraction (see in inset). (B) Overlay of the standard auto/cross-correlation curves from signals in the green 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 μ W and at 160 μ W power at objective. Two bunching terms are needed to fit the data ($t_T = 4.5$ μ s, and $t_b = 60$ μ s). The triplet fraction changes from 10% at 80 μ W to 15% at 160 μ W power at objective. Also changes in diffusion times are observed from 0.8 ms at 80 μ W to 0.6 ms at 160 μ W 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 μ W and at 160 μ W power at the objective. The relaxation times fitted globally are $t_{R1} = 6$ μ s and $t_{R2} = 240$ μ s, that are within the errors presented on Table 3.3.2C. Note that the amplitudes do not change as in the case of the standard auto-correlation.

3.4. Outlook

In the study presented in this chapter we identified the third conformational state C_3 as thermally excited state with a so far unknown structure. Its determination is of particular interest, since it can provide deeper insight into the state's functional relevance. Given the sparsity of data provided by FRET experiments, identification of such a conformation will require a novel approach to allow achieving a high resolution for the state's structure.

The measured transition time for C_2 to C_3 is $\sim 300 \mu\text{s}$, which makes conventional MD inapplicable. Additionally, MD force fields are not optimized for the simulation of thermally excited states, in which frequently partially less structured regions within the protein's structure could occur. These features turn the generation of a simulated structure of C_3 into a challenge. At the moment we are developing a new hybrid method, which is based on a combination of computer simulations and the experimental FRET data. Incorporation of experimental data in simulations is implemented in a form of restraints. Here, the Accessible Volume Simulation (AVSim) of our FRET Positioning and Screening program (FPS) (46) was shown to be an effective way for incorporation in the FRET-restrained all-atom MD simulations.

In a collaborative project*, we are running right now FRET-restrained MD simulations and hope to gain a high-resolution structure of C_3 soon.

*Joint project “*FRET restrained structural modelling resolves an excited state in the catalytic cycle of T4L*” done by collaboration of: Mykola Dimura, Thomas Peulen, Katherina Hemmen, Hugo Sanabria, Dmitro Rodnin, Suren Felekyan, Ralf Kühnemuth, Holger Gohlke and Claus A.M. Seidel.

4. Unfolding pathway of T4L

4.1. Multiparameter Fluorescence spectroscopy reveals sub-millisecond intermediate transitions in protein unfolding

The manuscript is based on collaborative work in a joint project of the following persons: Katherina Hemmen, Dmitro Rodnin, Daniel Rohrbeck, Hugo Sanabria, Soheila Rezaei Adariani, Suren Felekyan, Ralf Kühnemuth and Claus A.M. Seidel.

The contributions to this manuscript are as follows: SRA conducted experiments under supervision of **KH**, DmR, and HS during her 10 weeks BioSoft Summer school. DaR did the majority of the CD and steady-state titrations. DmR measured a part of the MFD and fFCS data. SF developed fFCS. RK prepared the setup for measuring FCS. **KH** purified and labeled the protein and conducted eTCSPC and FCS measurements and missing MFD, fFCS, CD and steady-state titrations. **KH** analyzed the data. HS, TP and CS discussed data analysis. HS and CAMS designed the study. **KH**, HS and CAMS wrote the manuscript.

KH: 45 %

DmR: 12 %

DaR: 10 %

SRA: 5 %

SF: 6 %

RK: 6 %

HS: 8 %

CAMS: 8 %

4.2. Manuscript

Abstract

We used lysozyme from the phage T4 (T4L) as a simple model system to study folding and unfolding kinetics. Most proteins adopt already one or more conformations in their native state. Currently, no single biophysical technique exists to map protein kinetics covering all relevant timescales in a single experiment. Previous NMR and stopped-flow experiments showed that T4L unfolds via at least two intermediates with an intermediate transition in the millisecond regime. To fill the existing kinetic gap in the microsecond regime and to study conformational changes, we created a set of variants to span a network over the enzyme to probe distance fluctuations and distributions by Förster resonance energy transfer (FRET). The denaturant induced loss of secondary structure was monitored using CD-spectroscopy. Single-molecule high-precision FRET measurements and (filtered) fluorescence correlation spectroscopy (filteredFCS) allowed us to determine equilibrium transition rates between conformational states while equilibrium distributions were determined by time correlated single photon counting (eTCSPC). With our network, we show the stepwise, sequential unfolding of structural elements in T4L. *A priori* knowledge helped us to assign, formerly unknown, reaction rate times to intermediate transitions. Hence, our unique combination of methods delivered complementary information to standard methods and allowed us to deduce an unfolding scheme covering all potential kinetically relevant processes - spanning ten decades from ps to ms.

Significance Statement

Understanding the folding pathways of proteins and thus the underlying kinetics is required to understand and combat the occurrence of misfolded proteins, which are frequently linked to severe diseases such as Alzheimer's. Here, we present a unique combination of single-molecule high-precision FRET measurements and (filtered) fluorescence correlation spectroscopy (filteredFCS) with ensemble time correlated single photon counting (eTCSPC) which shed light on the fast time scales of protein unfolding-folding kinetics and, thus deliver complementary information to the slower processes measured by standard methods. In the presented study, we applied our combination of tools on the model enzyme phage T4 lysozyme (T4L) and we showed the sequential loss of structural elements completing the proposed unfolding scheme.

4.2.1. Introduction

Understanding how a linear strand of amino acids folds into its correct and functional three-dimensional protein structure is one of the remaining challenges in biochemistry. During folding the free energy of the formerly unfolded amino acid sequence decreases gradually and is minimal in the stable native structure. This process is often depicted by the so-called folding-funnels which give a cross-section through the multi-dimensional folding pathway (124). Experimentally it is possible to study folding pathways by destabilizing the native state using chemical denaturants. Typically, the stability and folding pathway of proteins is studied via equilibrium and non-equilibrium methods such as observing changes on the intrinsic fluorescence in denatured conditions (109), stopped-flow measurements (observing the tryptophan fluorescence or the circular dichroism (CD) of a protein), or hydrogen-exchange (125) and relaxation dispersion (2, 3) NMR experiments. Using the determined rate- and equilibrium constants the energy landscape of the linear folding/ unfolding pathway with its possible intermediates is reconstructed (126, 127) (Figure 4.2.1A-B).

The catalytic activity of many enzymes requires the existence of several conformational states under native conditions (Figure 4.2.1B) (native state equilibrium, *NSE*). As these conformations may differ significantly in stability and may populate thermally excited states, a complex folding pathway is anticipated (Figure 4.2.1A-C). Most experiments observe ensemble properties, which limits their resolution. Hence, the data interpretation is restricted to simple two state folding models (Figure 4.2.1A) with a single folded and an ensemble of denatured states (*DSE*). Here, we use single-molecule methods to increase our resolution by observing molecule by molecule. This allows for physical more reasonable descriptions by disentangling complex pathways of enzymes containing energetically different multiple native states, unfolding intermediates and the ensemble of fully denatured states (Figure 4.2.1A-C).

Lately, Förster resonance energy transfer (FRET) emerged as a complementary method within the field of protein folding (11, 128, 129). FRET is most sensitive in ~ 50 Å and can be measured under equilibrium and non-equilibrium conditions. By single-molecule FRET sub-states and the population distribution (25, 65) can be identified. Fluorescence detection with multiple parameters further increases the resolution (128, 130) and allows for the direct assessment of the sample's heterogeneity and dynamics (25, 65). Correlation based techniques permit to measure kinetics under equilibrium conditions in the ms to the ns time

regime without external synchronization. This is especially interesting for mild denaturing conditions where both folded and unfolded states coexist.

Here, T4L serves as an ideal model to study complex folding pathways. It has been the target of many studies concerning protein stability (27-31, 38) and protein folding (20, 21, 39-45, 64). It is a small 164 amino acid protein with a C-terminal (Figure 4.2.1D) and an N-terminal subdomain (21, 33, 42, 43). The less stable N-terminal subdomain is connected by the long helix C to the more stable C-terminal subdomain. Helix A, although located sequence-wise in the N-terminal subdomain, belongs structurally, and energetically to the C-terminus (42). T4L was considered a typical small two-state folding protein (39), but early experiments conducted under neutral conditions, suggested the existence of a more complex unfolding pathway (40). Nowadays, it is known that T4L unfolds via at least two intermediates, an unfolding intermediate J and a folding intermediate I (20) (Figure 4.2.1B). The transition from J to I is in the ms to sec timescale (20). Additionally, Kato *et al.* determined in 2007 (43), by hydrogen exchange NMR, a partially folded intermediate structure (20, 21, 41-43), where the stable C-terminus is still folded and the N-terminus is unstructured.

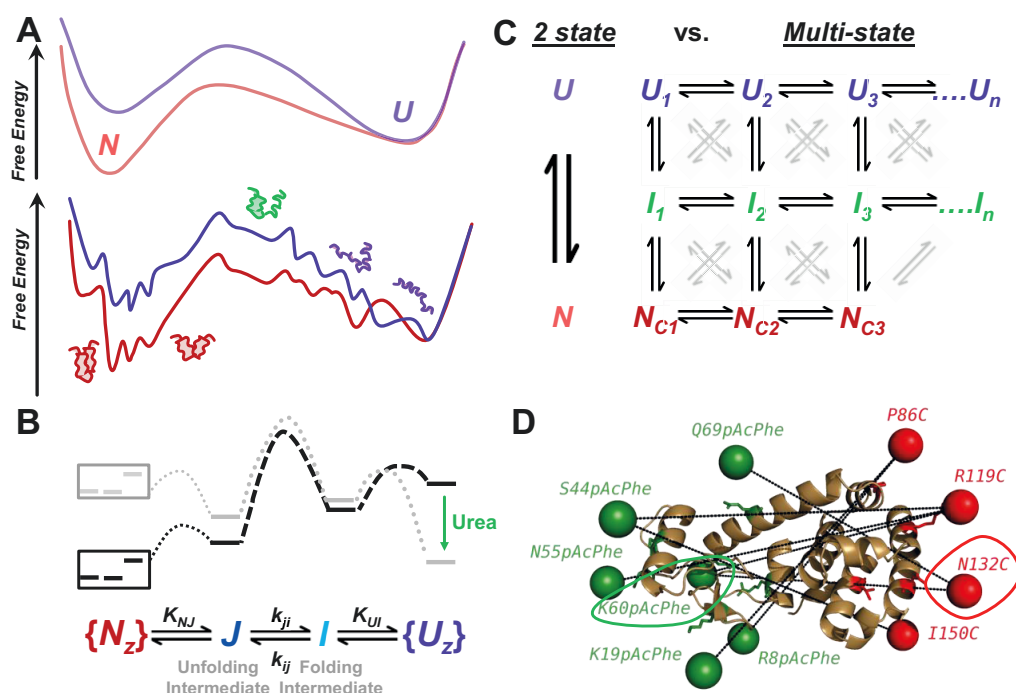


Figure 4.2.1 Complex unfolding network of Phage T4 Lysozyme. (A, B) Two-state folders go directly from the folded into the unfolded state, whereas the pathway for proteins possessing multiple native state (and intermediates) is much more complex. While folding, into native structures via possible intermediate structures the free energy of protein conformations is gradually decreased. (C) Unfolding of T4L occurs via two intermediates J and I (20). Energy landscapes for T4L wt* were calculated according to the values in table 2 (20). (D) Labeling positions and network to probe the unfolding of T4L. The unnatural amino acid para-Acetylphenylalanine (pAcF) was used to couple the donor Alexa488-hydroxylamine (green) to the protein; Cysteine residues were used to couple the acceptor Alexa647-maleimide (red). All experiments were performed for the variant 60-132, from other only selected experiments have been performed.

Furthermore, we found that three energetically different conformations coexist under native conditions in the lysozyme of phage T4 (T4L). Two fast interconverting major populated conformations N_{C1} and N_{C2} exchange with the low populated N_{C3} in the sub-ms timescale (Chapter 3). This conformational dynamics is required for its catalytic activity and results in a displacement of the two subdomains towards each other.

Here, we aim to investigate the unfolding processes in the complex unfolding pathway of T4L happening on timescales faster than milliseconds down to the nanoseconds timescale under equilibrium conditions. Considering the structure of the partially folded intermediate (43), we wondered whether the conformational dynamics of the two subdomains and the existence of a thermally excited state under native conditions might be the seeds, which facilitate the unfolding of T4L. Thus, N_{C3} might serve as “exit-state”. To address these questions we used a holistic approach and designed a labeling network of eight different variants (Figure 4.2.1D). The selected positions were designed to cover the whole protein, and thus report on changes in different parts in the proteins, which will allow us to trace the stepwise unfolding. The variants were labeled site-specifically using orthogonal chemistry (94, 95) to reduce uncertainties on the location of the dyes. For each variant, we conducted ensemble and single-molecule fluorescence spectroscopy experiments and confirmed the proper folding of all variants by CD spectroscopy. Next, we assessed the unfolding process by placing all variants in denaturing conditions (increasing concentration of urea) and monitored the ensemble properties with CD spectroscopy and time-resolved fluorescence intensity measurements. The second part of this study focuses on the unfolding kinetics of T4L using single-molecule fluorescence and fluorescence correlation methods of singly diffusing molecules. The derived results and time constants complete the unfolding scheme proposed by (20).

To our knowledge this is the first time in which such a combination of tools was used to fill the gaps at the fast timescales in the unfolding pathway of a protein and in general, it can be used to shed light on the fast unfolding reactions of proteins.

4.2.2. Results

4.2.2.1. Tertiary structure transition depends on label position

The proper folding of all T4L variants was assessed by comparing their CD spectra with the spectra of the cysteine-free pseudo-wild type T4L wt* (Figure 4.2.2A, Figure 4.3.4A). No difference between unlabeled and labeled protein could be seen. Subsequently, the loss of

secondary structure of the predominantly α -helical protein was recorded by CD. The normalized change in CD signal at 222 nm displayed the sigmoidal trend of a two-state folding protein (7) with a midpoint of transition (MoT), at which by definition the molecules are 50 % folded and 50 % unfolded, of 5.1 M for the selected variant 60-132 compared to a MoT of T4L wt* of 5.3 M (Figure 4.2.2C, Figure 4.3.4B). Table 4.2.1 summarizes the MoT of all variants. Insertion of mutations in T4L wt* only decreases slightly the stability of each variant (20, 21).

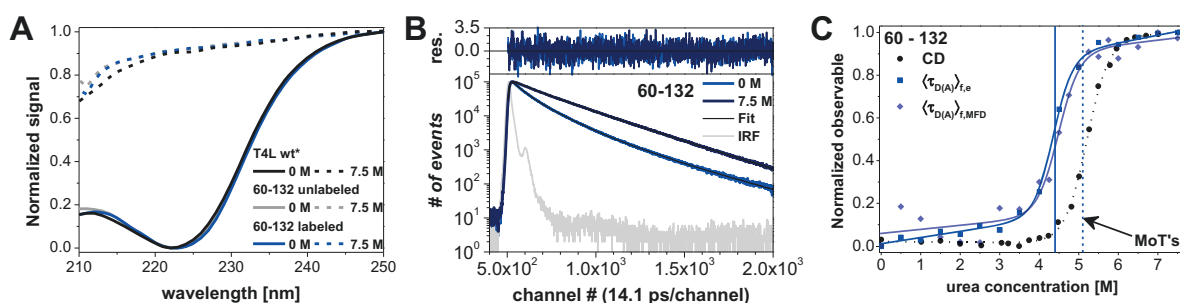


Figure 4.2.2 Loss and transitions in secondary and tertiary structure. **(A)** Normalized CD signal from labeled and unlabeled T4L wt* and 60-132 variant in presence and absence of urea. **(B)** Time-resolved fluorescence decay of variant 60-132 in presence and absence of urea, fitted with a multi-exponential model. Residuals are shown on top. **(C)** Normalized CD signal at 222 nm and fluorescence weighted averaged lifetime $\langle \tau_{D(A)f} \rangle$ of variant 60-132 in increasing concentrations of urea. The two-state fit is shown as solid line; midpoints of transitions (MoT) are shown with dashed lines (Table 4.2.1).

Next, the loss of tertiary structure was probed by ensemble time-correlated photon counting (eTCSPC). Figure 4.2.2B shows the time-resolved fluorescence intensity of double labeled (DA) protein in absence and presence of 7.5 M urea. The slower decay in the denatured protein is a clear indication of an in average larger interdye distance. The decay curves were fitted with a multi-exponential model and the fluorescence-weighted average lifetime $\langle \tau_{D(A)f} \rangle$ was calculated (SI 4.3.1.3, eq. 4.3.3). The evolution of $\langle \tau_{D(A)f} \rangle$ resembles the CD signal, but with an MoT of 4.4 M (Figure 4.2.2C, Table 4.2.1), which corresponds to more than half a mole of urea difference in comparison with the behavior of the secondary structure according to CD. Close inspection of all eight T4L variants divides them into two classes: (I) Change in FRET occurs prior to change in CD and (II) FRET and CD signal change synchronously (Table 4.2.1, Figure 4.3.4B, Figure 4.3.5A). To the first class belong all variants whose donor fluorophore is placed close to (position 55) or inside the two subdomains connecting helix C (position 60 and 69). For donor fluorophore positions 8, 19 and 44 the signals change synchronously.

Additionally, the time-resolved fluorescence intensities were fitted assuming a Gaussian distributed mean distance $\langle R_{DA} \rangle$ with a width w_{DA} or by a worm-like chain polymer model

(WLC) (SI 4.3.2.2, Table 4.3.2 and Figure 4.3.5). The first model did not disclose significant new information: As expected from increasing $\langle \tau_{D(A)f} \rangle$, the $\langle R_{DA} \rangle$ increased. Also, the distribution width w_{DA} broadened. The second model cannot be biologically justified especially for low urea concentrations where proteins definitively do not show polymer-like behavior, even if the model might fit the data. Nevertheless, it can be stated that the total chain length L increased in concord with either decreasing or equivalent stiffness κ . This resulted in an enlarged persistence length l_p for all except the variant 55-119. In this case, l_p decreases slightly.

Table 4.2.1: Summary of MoT for Circular Dichroism and ensemble time-resolved fluorescence-intensity experiments (A) and for the variant K60pAcF N132C for PDA and filteredFCS experiments (B).

A	Variant	MoT urea [M]		B	K60pAcF N132C	MoT urea [M]
		CD	$\langle \tau_{D(A)f} \rangle$			
	R8pAcF P86C	3.7	4.1		$\langle R_{DA} \rangle_E$	5.6
	K19pAcF P86C	4.6	4.7		hw_{DA}	4.2
	S44pAcF R119C	5.0	4.9		$\langle R_{DA} \rangle_{E,1}$	4.2
	S44pAcF I150C	4.8	4.7		$\langle R_{DA} \rangle_{E,2}$	5.6
	N55pAcF R119C	4.8	4.1		Amplitude LF species	5.4
	K60pAcF R119C	4.9	4.2		Amplitude HF species	4.7
	K60pAcF N132C	5.1	4.4		Amplitude t_1	5.3
	Q69pAcF N132C	5.0	3.8		Amplitude t_3	5.1
					t_{diff}	5.5

Further evidence that T4L lost its tertiary structure before losing its secondary structure stems from the observed changes in the steady-state fluorescence anisotropy r_{ss} , which was determined independently for each single labeled variant (Figure 4.3.6). We found that r_{ss} decreased strongly for the fluorophore positions in the N-terminal subdomain indicative for an increased local mobility in denatured conditions.

4.2.2.2. Complex unfolding behavior

We subsequently studied the dynamics of T4L unfolding using single-molecule fluorescence spectroscopic methods. Figure 4.2.3A shows exemplary the unfolding of variant 60-132 in a two dimensional multiparameter fluorescence detection plot (2D MFD), where the two FRET indicators donor to acceptor fluorescence ratio F_D/F_A and fluorescence-weighted average fluorescence lifetime of donor in the presence of acceptor $\langle \tau_{D(A)f} \rangle$ are displayed (SI 4.3.1.5, 4.3.1.6 and 4.3.2.4, Figure 4.3.7). This representation allows directly assessing the heterogeneity and dynamics of the sample (25).

Under native conditions, the 2D MFD plot of T4L shows only a single population slightly off the orange static FRET line. This shift reflects the dynamic averaging of the three native conformations (Chapter 3). Upon addition of the denaturant urea, the native conformations were destabilized. This leads to a tailing of the 2D population towards a higher F_D/F_A ratio and longer fluorescence lifetime $\langle\tau_{D(A)}\rangle_f$, indicative of the existence of molecules in an extended conformation. A possible unfolding path is sketched by the green line. A minor population clearly shifted off the static FRET line appeared with further increasing urea concentrations (4.5 M). The shift reflected the dynamic averaging with a state which exhibits a longer donor fluorescence lifetime $\langle\tau_{D(A)}\rangle_f$ and higher F_D/F_A ratio than the native state. This might be explained by a collapse of the elongated intermediate state along the sketched cyan line. At 4.75 M this minor population becomes the major population indicating that the MoT between the folded \rightarrow unfolded state is between 4.5 -4.75 M urea. This agrees well with the MoT determined by ensemble fluorescence measurements (4.2.2.1, Figure 4.2.2C, Table 4.2.1A). Under strong denaturing conditions, only one broad, inherently dynamic population is found. Although all populations show the same progression towards higher F_D/F_A ratios, longer $\langle\tau_{D(A)}\rangle_f$ and lower E , the detailed development is strongly position-dependent (Figure 4.2.3, Figure 4.3.7).

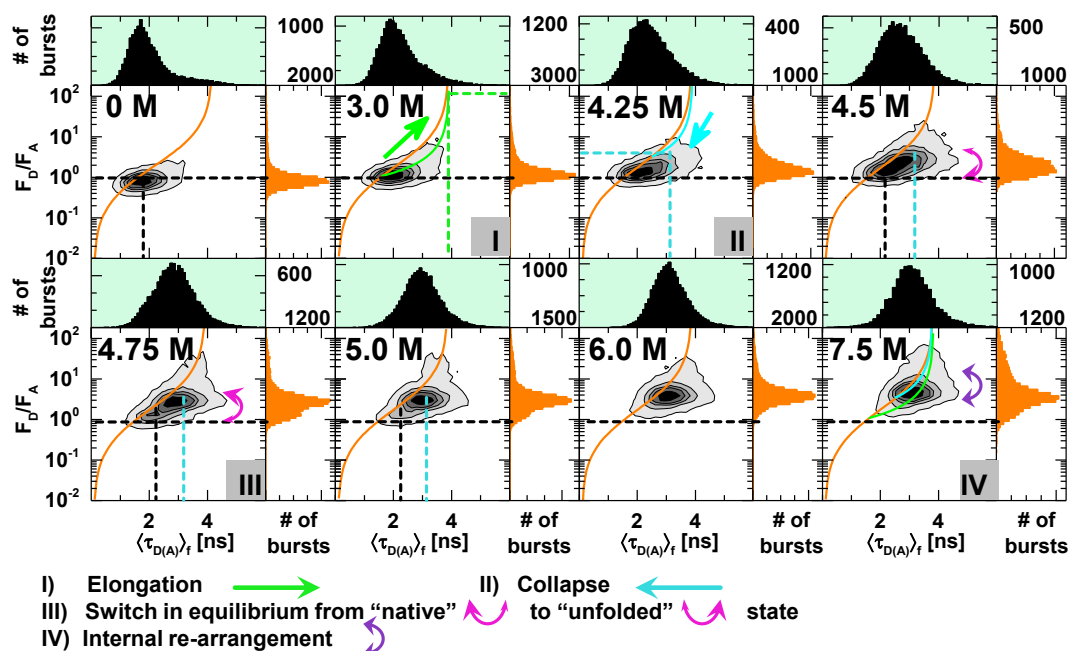


Figure 4.2.3 Variant 60-132 shows a complex unfolding behavior. Two dimensional histograms of green to red signal ratio F_D/F_A vs. lifetime of donor in the presence of acceptor $\langle\tau_{D(A)}\rangle_f$ in increasing urea concentrations. One dimensional projections for F_D/F_A and $\langle\tau_{D(A)}\rangle_f$ are also shown. Static FRET lines are shown in orange. Dynamic FRET lines in bright blue and green were generated empirically to guide the eye. Correction factors for background in green and red, detection efficiency ratio and parameters used to generate the static FRET lines are summarized in SI 4.3.2.4, Table 4.3.2

4.2.2.3. *Dynamics in the ms timescale*

Next, we quantified the dynamics and their timescales outlined in Figure 4.2.3 with photon distribution analysis (PDA), full and filtered fluorescence correlation spectroscopy. The dynamics already highlighted by the shifts in MFD plots (Figure 4.2.3, Figure 4.3.7), can be quantified with PDA.

Kinetic information can be extracted by binning the single-molecule photon trace in different time-windows (TWs) and comparing the width of the distribution by considering photon shot-noise due to differences in the integration time. Dynamics in the ms-timescale, i.e. during burst duration, leads to broadening and narrowing, respectively, of the data histogram (48, 63-65). Previously PDA was used to decipher single-molecule kinetics and limiting states using specific model functions (48, 65). Given the expected complexity of the system, we first use a model free approach to identify slow kinetics at the cost of being unable to fully untangle the limiting states and the kinetic scheme. We find changes of the apparent distance distributions in dependence of the TWs (0.5 to 12 ms), which cannot be explained by shot-noise alone. In Figure 4.2.4A, this is shown for the variant 60-132 at 4.75 M. This is a clear indication for dynamics in the sub-millisecond to millisecond range.

As fullFCS and filteredFCS offer a higher dynamic range compared to PDA we decided to describe the urea dependency empirically by a given fixed time-window of 1 ms, in which we fitted the interdyne distance with a single Gaussian distribution with mean interdyne distance $\langle R_{DA} \rangle_E$ and half-width h_{WDA} (SI 4.3.1.7, 4.3.2.5). Figure 4.2.4B shows that both increased upon protein unfolding for the labeling site 60-132 in a two state manner. First h_{WDA} doubles suddenly with a MoT of 4.2 M urea, this is followed by a steep increase of $\langle R_{DA} \rangle_E$ with a MoT of 4.6 M. Generally, $\langle R_{DA} \rangle_E$ slowly increases before and after the transition point, whereas h_{WDA} stays more or less constant.

In summary, all eight variants showed an increase in fluorophore distance: in some variants gradual shifts and in some steep jumps were observed. However, as PDA relies on intensity based histograms of single-molecule data, its ability to detect and describe complex dynamics is limited (48, 65).

4.2.2.4. *fullFCS reveals early intermediate transition*

To shed light on fast processes, we conducted fluorescence correlation spectroscopy experiments under continuous wave excitation (fullFCS). FullFCS has a wide dynamic range from ps to ms and can extract time constants with high precision. Figure 4.2.4C shows the

autocorrelation functions (*ACF*) of the signal of the donor fluorophore in absence (D0) and presence (DA) of acceptor. All *ACF* curves have in common a sharp rise in the ns time – the photon antibunching time t_{AB} –, occurring at the fluorescence lifetime of the used fluorophore (here: ~ 4 ns), a polarization dependent term in the ~ 10 ns timescale reflecting the overall rotational correlation of the protein (t_{RC}) and finally the correlation is lost at timescales larger than the diffusion time (t_{diff}) where the correlation decays to a flat line. Relaxation time constants found on timescales between t_{RC} and t_{diff} sense further motions of the protein (\sim backbone). As can be seen in Figure 4.2.4D, the single labeled protein showed two additional terms. A global fit of the *ACFs* of D0 and DA samples revealed two FRET-induced terms under native urea concentrations ($\sim 0.2 \mu s$ and $\sim 300 \mu s$) in the DA sample – the latter time constant has already been identified previously (Chapter 3). Additionally, two time constants were found in both D0 and DA sample ($\sim 20 \mu s \rightarrow 2 \mu s$, $\sim 100 \mu s$), which might be due to the dye photophysics of triplet blinking or the quenching by nearby amino acids (131, 132). The relaxation time constants due to FRET quenching only present in the DA sample display a peak at ~ 3 M urea. The shape resembles an upside down chevron plot typically obtained in stopped flow experiments (20). This peaking shape can also be found partially in the donor-acceptor cross-correlation curves (D-A *CCF*) (Figure 4.3.9C, D). Here, under native conditions three relaxation time constants faster than diffusion are found (~ 2 ns, ~ 41 ns, $\sim 3 \mu s$), which diminishes to only one with increasing urea concentration at ~ 70 ns. Also, a slower relaxation time constant of $\sim 400 \mu s$ – similar to the $\sim 300 \mu s$ found in the donor-autocorrelations from the DA-sample in the absence of urea – is found throughout all urea concentrations.

From fullFCS analysis, we conclude that T4L excerpts dynamics on a wide range of timescales, in both single and double labeled samples over all urea concentrations, originating, partially, from fluorophore photophysics and from protein motion induced (FRET-) quenching processes. At higher urea concentrations, the probed fluorescence fluctuations seem to be dominated by non-FRET related processes as here the DA-sample did not display additional relaxation time constants. However, the population in the MFD plots (Figure 4.2.3) is centered at $\langle \tau_{D(A)} \rangle_f$ about 3 ns. This is significantly shorter than the fluorescence lifetime of a D0 sample, thus quenching due to FRET still occurs.

4.2.2.5. *filteredFCS detects late intermediate transition*

To investigate the FRET-induced processes present under highly denaturing conditions, we performed filteredFCS experiments, which allowed us to “zoom-in” onto all FRET-induced processes (47, 70, 72). A great advantage of this method is that fluctuations due to photophysical “artifacts” are suppressed.

The necessary filters for correlation were generated by selecting a “low FRET” (LF) and a “high FRET” (HF) population from TW-based analysis (SI 4.3.1.9, 4.3.2.7 and Figure 4.3.11G-I) of the MFD data, with which we can amplify the transition from the folded to the fully unfolded state of T4L thereby passing through all intermediate states. The top panel of Figure 4.2.4E shows exemplary the species-selective cross-correlation curves (*sCCF*) for the variant 60-132 in absence and presence of 6.5 M urea. The increase in diffusion time (Figure 4.3.11E) and the difference in dynamics are immediately evident.

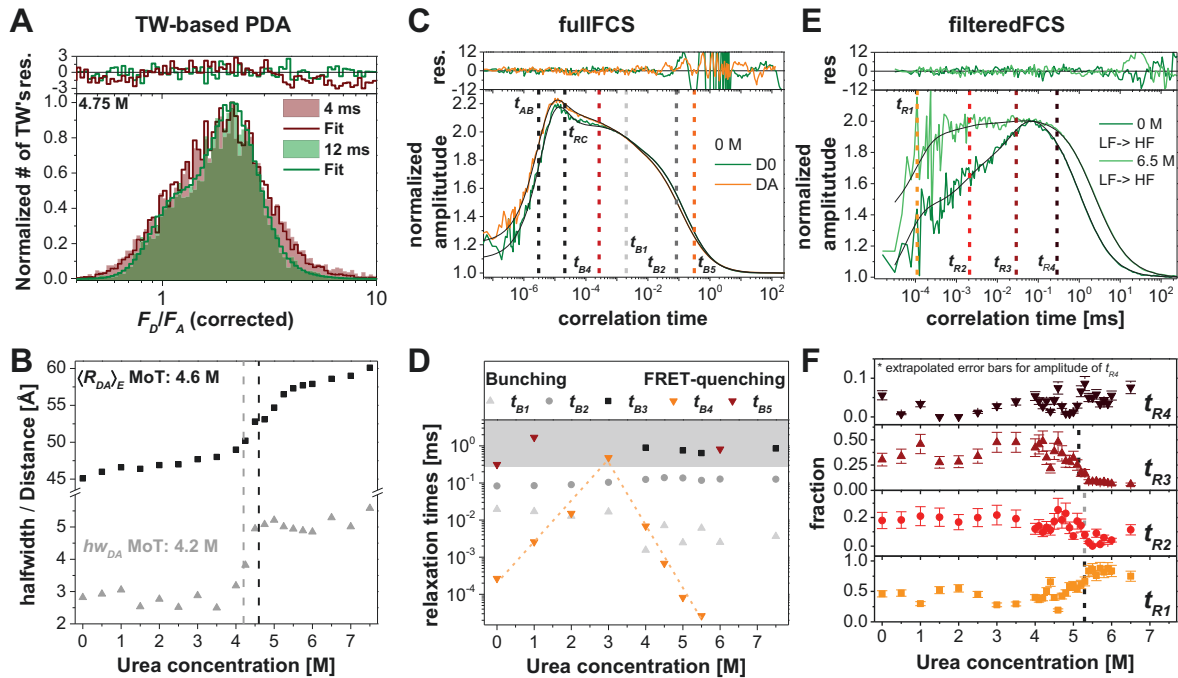


Figure 4.2.4 T4L displays dynamics on a wide range of timescales during unfolding (A) PDA of variant 60-132 in 4.75 M urea with 2 different time-windows, residuals for fitting both with the same model are shown on top. (B) In increasing urea concentration, mean interdyde distance $\langle R_{DA} \rangle_E$ and half-width hw_{DA} increase in a two state manner. (C) fullFCS curve of donor-ACF of single (D0) and double labeled (DA) protein in absence of urea. Vertical lines indicate the found bunching (t_{B1} , t_{B2} , t_{B4} , t_{B5} , t_{RC}) and antibunching terms (t_{AB}). Global time constants are shown in black-gray, time constants found only in DA samples are shown in orange and red. (D) Relaxation times obtained from fitting green ACF of D0- and DA-labeled samples globally. (E) filteredFCS sCCF curves of low FRET to high FRET correlation function of DA sample in the absence and presence of 6.5 M urea. Vertical lines indicate the average global t_{rel} (\pm standard deviation): $0.11 \mu s$ ($\pm 0.06 \mu s$), $2.1 \mu s$ ($\pm 0.7 \mu s$), $29 \mu s$ ($\pm 8 \mu s$) and $287 \mu s$ ($\pm 54 \mu s$). (F) Amplitudes obtained in sCCF curves depending upon the urea concentration. Error bars indicate average standard error of the mean and gray shaded area times slower than the diffusion time. Error bars for the fraction x_4 of t_4 is the mean standard error of those for x_1 , x_2 and x_3 as in the fitting routine $x_4 = 1 - x_1 - x_2 - x_3$.

Interestingly, four relaxation time constants t_R were found independent of the urea concentration (vertical lines in Figure 4.2.4E, Figure 4.3.11A). The amplitude of the slowest t_{R4} ($287 \pm 54 \mu\text{s}$) first diminishes to zero but begins to rise again after 2 M and reaches its initial value again at high concentrations of urea (Figure 4.2.4F). Amplitude of t_{R3} ($29 \pm 8 \mu\text{s}$) and t_{R2} ($2.1 \pm 0.7 \mu\text{s}$) simultaneously decrease (sharply) with increasing urea concentration whereas the amplitude of the fastest t_{R1} ($0.11 \pm 0.06 \mu\text{s}$) increases up to $\sim 80 \%$. The trend of the amplitudes for t_{R1} and t_{R3} could be fitted with the two-state model with MoT's of 5.3 M and 5.1 M urea (Table 4.2.1B, Table 4.3.7).

Out of the four t_R found, two have already been assigned in previous experiments (Chapter 3) to transitions in the native state. There, the functional cycle of T4L was studied and revealed a native state equilibrium of three states (Nc_1 , Nc_2 , Nc_3). t_2 was assigned to the fast exchange of $Nc_1 \rightleftharpoons Nc_2$ and t_4 to the slow exchange of $Nc_2 \rightleftharpoons Nc_3$ (Chapter 3).

Additionally the species-selective autocorrelation functions (*sACF*'s) were globally fitted and the species fractions of LF, HF and the dynamic molecules were determined (Figure 4.3.11B-D). The amount of HF species (MoT 4.7 M) drops faster than the amount of LF species rises (MoT 5.4 M) (Table 4.2.1B). This shift in MoT's is compensated by an increase of dynamic molecules (Figure 4.3.11D).

In summary, even under highly denaturing conditions processes that change the FRET-level are still present and here the filteredFCS analysis shows its full advantage of suppressing protein-motion unrelated fluorescence fluctuation. Due to our previous study (Chapter 3) we can assign two out of the four t_R 's found to definite processes.

4.2.3. Discussion

4.2.3.1. Network of variants maps the sequential loss of structural elements

For all but one variant, 8-86, eTCSPC and CD displayed the characteristic trends of two-state folders, i.e. direct progression from the folded state towards the ensemble of denatured states (*DSE*) (7) (Figure 4.3.4B, Figure 4.3.5A, Table 4.2.1A). However, differences among the variants allow assessing the stability of tertiary structure, and thus mapping the stepwise, sequential loss of structure in T4L. We evaluate the secondary structure elements and tertiary arrangements by CD and eTCSPC, respectively (Figure 4.2.2). As it is necessary to evaluate the influence of mutations and labeling on the protein stability (Table 4.2.1A) (20, 21), the CD-measurements serve as references to compensate for variant specific effects.

The MoT's in CD of the variants and T4L wt* are distributed between 3.7 M (1st transition in 8-86, see below) and 5.3 M (T4L wt*) urea (Table 4.2.1A). This difference in protein stability presumably stems from the inserted mutations, an effect observed (and benefitted from) in earlier studies as well (20, 21). By comparing the eTCSPC- to the CD-MoT's, the variants can be grouped in two classes: Synchronous change in CD and eTCSPC (class I, donor fluorophore at position 55-69) and the FRET-MoT precedes the CD-MoT (class II, donor fluorophore at positions 8-44) (Table 4.2.1A, Figure 4.2.5A). We correct for mutational influences by dividing the MoT of eTCSPC by the MoT of CD, and determine a relative stability of tertiary vs. secondary structure.

Figure 4.2.5A shows that the loss of tertiary structure started from the middle of helix C, label position 69, at ~ 75 % of the urea concentration necessary to dissolve half of the secondary structure. This was subsequently followed by position 60 and 55 at ~ 83 % of the CD-MoT. Thus, the interactions stabilizing the positions 55 to 69 relative to the N-terminal subdomain were lost at 90 % of the original CD-signal. As the CD-signal originates mainly from helices, this indicated that the helices are still intact, even though the overall protein architecture is not. Next, when half the tertiary interactions to positions 44 and 19 are diminished, the CD signal is nearly halved. Finally, the interaction between the C-terminal subdomain and helix A, label position 8, is lost. It has to be noted that the variant 8-86 shows a three-state unfolding behavior in CD. The label position 8 is furthermore interesting in the way that, although sequence-wise belonging to the beginning of the N-terminus, structurally-wise helix A attaches to the C-terminus (33, 41, 42) and is held in position by hydrogen bonds and other intra-protein acting forces (41).

Generally, only the holistic approach of a network of FRET pairs, with which we could selectively probe the changes in different regions of the protein, allowed us to show the sequential unfolding of the structural elements of T4L with increasing concentration of urea.

4.2.3.2. FCS fills kinetic gaps in T4L unfolding scheme

Experimentally we found relaxation times in the sub- μ s, μ s and ms- range. It is known, that ns dynamics is related to peptide-chain diffusion and sub- μ s dynamics to the formation of transient secondary structures. The concerted motion of smaller and larger partly or completely folded domains takes place in the μ s and ms timescale (19, 128), respectively. The measured relaxation time constants are now related to the individual transitions (*NSE* \rightarrow

$J, J \rightarrow I, I \rightarrow DSE$) to obtain a kinetic scheme of the unfolding process. Hereby also the three native conformations (N_{C1}, N_{C2}, N_{C3}) are considered.

FullFCS has the largest dynamic range of the methods used and is able to probe dynamics down to the ns range. D0 and DA ACFs share two to three time constants ($t_{B1} \sim 2\text{--}20\ \mu\text{s}$, $t_{B2} \sim 100\ \mu\text{s}$, $t_{B3} \sim 1\ \text{ms}$). The shortest component might reflect the triplet state blinking of the used fluorophore Alexa488 (133). The two slower relaxation time constants, however, are most likely induced either by local changes in the fluorophore environment or by the quenching of the fluorophore due to tryptophan (three in T4L), tyrosine (six), methionine (five) or a histidine (one) (131). Both effects lead to brightness fluctuation in the recorded fluorescence signal. In many studies this “self-quenching” of the fluorophore by the amino acid residues of single labeled protein is used (134-138).

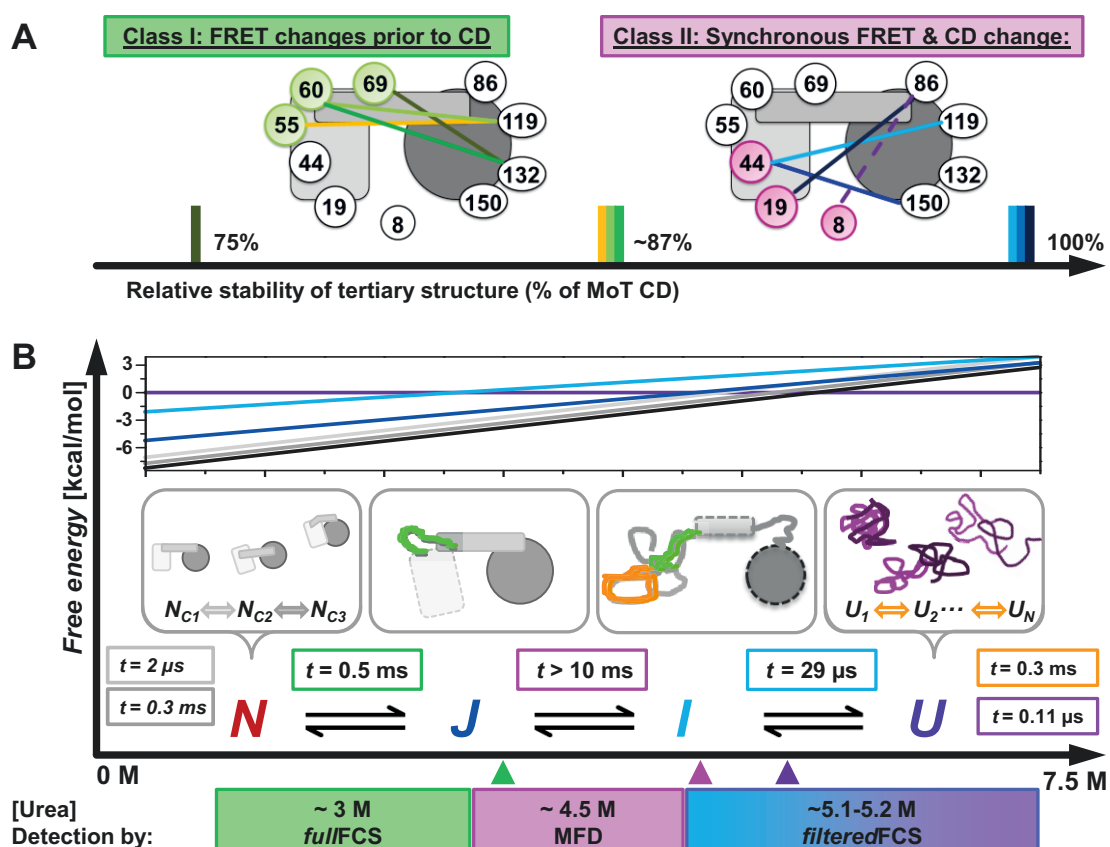


Figure 4.2.5 Our combined spectroscopic fluorescence approach resolves intermediate transitions towards unfolded states. (A) Depending upon label position, the MoT of fluorescence and CD either coincide or the MoT of fluorescence occurs at lower denaturant concentration than in CD spectroscopy. Note: Variant 8-86 displays a 3-state behavior in CD, MoT of transition with higher amplitude is shown. (B) Using the preknowledge that (i) we have an equilibrium of at least three native states and (ii) that the transition between the unfolding intermediate J to the folding intermediate I occurs on the ms timescales, we are able to assign our relaxation time constants found to the distinct unfolding steps. Free energy of states relative to U were calculated according to the values in table 2 (20).

Here, motions of the (partially) unfolded peptide chain bring the quenching residues close ($< 5 \text{ \AA}$) to the fluorophore and thus induce brightness fluctuations in the fluorescence signal. The $100 \text{ }\mu\text{s}$ component falls well into the time regime identified already by others for concerted chain motions (4, 19, 128, 134, 136-138). The $300 \text{ }\mu\text{s}$ component is involved in the subdomain motions occurring while substrate cleavage (Chapter 3). The slowest component, however, is slower than diffusion, where FCS loses its sensitivity.

Additionally, we found relaxation time constants present only in DA-samples, which are related to FRET- (quenching) induced process(es) and form a triangle with a peak in the range $\sim 3 \text{ M}$ urea (Figure 4.2.4D), and are thus probing comparably large distance changes between the two labeling sites in the protein ($\sim 20\text{-}100 \text{ \AA}$). The distribution of those time constants obtained gave rise to an upside down chevron-plot, well-known from stopped flow folding studies in the ms to s range (20). The reason it is “inverted” is that we show relaxation time constants instead of reaction rate constants. Note that the probed timescales range is from $\sim 10 \text{ ns}$ up to 0.5 ms , and that these rates were determined in equilibrium conditions. We assigned this transition occurring under moderately destabilizing conditions, to the formation of an early unfolding intermediate (Figure 4.2.5B), thus the transition from the native state ensemble *NSE* to *J* (20). The urea-concentration depending t_{rel} lie between $0.2 - 500 \text{ }\mu\text{s}$. Processes in the sub- μs to μs range generally involve the formation/loss of transient secondary structure and smaller (concerted) domain motions (4, 19). A similar peak is seen in the donor-acceptor cross-correlation functions *D-A CCFs* (Figure 4.3.9C). In the *D-A CCFs*, a fast 100 ns process is dominating, which reflects the chain dynamics of the unfolded peptide chain also found in other studies (130, 135) and is on the same timescale, on which we expect to find our linker dynamics present in well-folded proteins.

Next, it is known that the transition between the two intermediates $J \rightarrow I$ (Figure 4.2.5B) is the rate limiting step in the (un)folding of T4L and occurs on the ms to sec timescale (20, 43), thus is too slow to be directly probed by our applied fluorescence methods. However, we found in the 2D MFD plots (Figure 4.2.3), between 4.5 and 4.75 M urea a visible shift in two populations. As both populations are separated in the MFD plot, the transition of molecules to the respective other state has to be slower than burst duration (a few ms). Furthermore, the PDA-MoT (Figure 4.2.4B) and the eTCSPC-MoT were located in this urea concentration (Table 4.2.1A). As the loss of secondary structure happens at higher urea concentration, the observed change here is related to a rearrangement of the tertiary structure. Taken together, these methods report indirectly on the slow transition occurring here.

Subsequently, the final transition from *I* to the *DSE* follows (Figure 4.2.5B). Here, it was advantageous that we used double labeled samples to probe the fluorescence fluctuations and not only single labeled samples. Hence, we could generate species-selective filters, which accentuate the transition between two species, and generate species-selective correlation curves (47). Here, we selected the end states, high FRET and low FRET (Figure 4.3.11), to probe the full transition from the folded to the completely unfolded state passing through all intermediate transitions. In total, the global fit of both *sCCFs* and both *sACFs* showed complex dynamics and required four time constants (Figure 4.2.4E-F), out of which two ($t_{R2} \sim 2 \mu\text{s}$ and $t_{R4} \sim 300 \mu\text{s}$) have already been identified in a previous study to describe the conformational fluctuations between the three native states (Chapter 3). As would be expected for a time constant involved in enzymatic activity, the amplitude of t_{R2} diminishes under highly denaturing conditions, in concord with t_{R3} . However, this is not valid for t_{R4} . Here, the native and unfolded amplitude is nearly identical, although at both 2 M and 2.5 M urea, this time constant is absent in the *sCCFs*. Its re-appearance can be explained by increasing unfolding dynamics, where slow processes in the sub-ms range reflect the unwrapping of larger folded subunits, e.g. α -helices (4, 19). The complex dynamics and existence of slow(er) processes is also supported by the shape of the data in the 2D MFD plots: wide spread along the static FRET-line in both F_D/F_A and $\langle\tau_{D(A)}\rangle_f$ dimension (Figure 4.2.3, Figure 4.3.7). The amplitudes of t_{R1} and t_{R3} display a two state behavior with MoT's > 5.1 M urea (Figure 4.2.4B). A time constant similar to $t_{R3} = 29 \mu\text{s}$ was also found by other protein unfolding studies (134, 136, 138) and classified as concerted chain motions of partially folded structures. Werner *et al.* e.g. found in their study combining FCS and TCSPC that the species fraction of an intermediate of cytochrome c and the amplitude of a $30 \mu\text{s}$ process diminished synchronously (136).

The fastest time constant $t_{R1} = 0.11 \mu\text{s}$, which is identical to the one found in *D-A CCFs* of fullFCS (Figure 4.2.4C-D), with a final amplitude of $\sim 80 \%$ falls well into the timescale attributed to chain dynamics and overall global reconfiguration of a protein and was found previously by other studies as well (128, 130, 135). At first pass, it seemed counterintuitive that this relaxation time constant has an amplitude of $\sim 40 \%$ under native conditions (Figure 4.2.4F); however, the time scale of chain dynamics is identical to those of the motion of the flexible linkers used to connect the fluorophore to our protein. Thus, the increase in amplitude reflects the gain of flexibility within the amino acid chain of T4L with increasing urea concentrations, whereas in absence of urea only the linker flexibility is probed. This is

also evident from the broadening of the distance distribution obtained in ensemble (w_{DA} , Figure 4.3.5C) and MFD measurements (hw_{DA} , Figure 4.2.5B). Using nanosecond-FCS, Nettels *et al.* e.g. found for the FRET-labeled cold-shock protein Csp a ~ 40 - 70 ns relaxation time constant, which was dependent on the amino acid separation between the fluorophores (130). In single-labeled Protein L, a process in the 200 – 300 ns range was found (135).

Interestingly, in the majority of studies conducted so far only a single “unfolding” related time constant is reported (130, 134-136) (apart from occasional triplet blinking), although processes on many timescales are predicted to be involved (4, 19, 128), exactly as we found in our study. The use of single- and double-labeled samples allowed us to differentiate between “self-quenching” and FRET-induced fluorescence fluctuations, and thus increase the information content when compared to single-label studies. Additionally, the application of fullFCS helps to detect dynamics down to the ns range, whereas filteredFCS increases the sensitivity of the *D-A CCFs* by shielding the signals from photophysical processes by using species-selective filters to generate the *sCCFs* and *sACFs*. In fact, taken the results from MFD, PDA, fullFCS and filteredFCS together, we can deduce a kinetical folding scheme encompassing three native states, intermediates and the unfolded state ensemble. Indeed, we might have rather a quasicontinuum distribution of relaxation time constants instead of fixed time constants.

4.2.3.3. Bridging stability and kinetics

Assembling the information from several previous studies (20, 21, 41-44) and the results obtained here, we can infer some key features of T4L’s linear unfolding pathway and its intermediate structures (Figure 4.2.5). In the view of native state stability, we already determined a distribution under native conditions (N_{C1} , N_{C2} and N_{C3}) (Chapter 3).

The first intermediate *J* is populated in mildly denaturing conditions. At 3 M urea (Figure 4.2.4D) the chevron-like distribution has its peak in fullFCS; here the protein shows native-like CD-signal (Figure 4.2.2, Figure 4.3.4B) and merely a slight change in donor fluorescence lifetimes (Figure 4.3.5A), the most likely structure of an intermediate under these conditions might resemble marginally rearranged subdomains with mostly native-like structure (Figure 4.2.5B). As N_{C1} and N_{C2} are almost equal in energy, and N_{C3} being higher in energy (Chapter 3), this brings N_{C3} closest to *J* (Figure 4.2.5). Thus, *J* could be structurally similar to the excited native conformation N_{C3} , which could hence serve as “exit state” through which T4L enters its unfolding pathway. However, so far the structure of N_{C3} is unknown (Chapter 3)

and preliminary FRET-restrained molecular dynamics simulations show a reordering of the subdomains' orientation towards each other accompanied with a minor loss of secondary structure (Mykola Dimura (HHU), personal communication).

As the second unfolding step between 4.5 and 4.75 M urea, $J \rightarrow I$, has been shown to be the slowest (20, 43), we can safely propose a large loss in I of structure compared to J . This is supported by the eTCSPC data. For FRET, the MoT of $\langle \tau_{D(A)} \rangle_f$ is 4.4 M, thus, the loss of tertiary structure has reached more than 50 % (Figure 4.2.2C). Yet, the CD signal has decreased by only ~ 20 % of its initial amplitude (Figure 4.2.2C), indicating that just some of the predominantly α -helical structure of T4L is lost. This is in line with the partially folded intermediate found by others (20, 21, 42, 44): The less stable N-terminal subdomain, which consists mainly of loops and β -sheets is unfolded, but the C-terminal subdomain is still folded. We agree with previous studies that this intermediate is located on the unfolded side of the rate-limiting step, i.e. it resembles the folding intermediate I (Figure 4.2.1, Figure 4.2.5B). The high loss in FRET-signal is due to the fact, that we have always one of our labels placed in the now unfolded the C-terminal subdomain, and thus it is also clear that I consists of an ensemble of states (Figure 4.2.5B). In the completely unfolded state of the DSE, however, all secondary structure elements of T4L, and their respective 3-dimensional arrangement, are lost and might only be transiently present. In the view of the folding funnel (9), the states are considered not as single conformations but as an ensemble of conformations of comparable stability. The previous stopped-flow experiments conducted by Cellitti *et al.* showed that the transition from I to the DSE is fast (i.e. only small energy barrier exist between them) and that the equilibrium between I and U is strongly shifted towards U (Table 2 in (20)), making I difficult to detect (Figure 4.3.3). Hence, under strongly denaturing conditions, T4L quickly changes from the basin of I to U and back. Thus, the linear unfolding scheme proposed is in fact a slice through the folding funnel of T4L and its projection in one dimension. For I , the ensemble of states is characterized by the flexibility in structure, thus FRET-distances, in the unfolded part of T4L. The slowest relaxation time probed ($t_{R4} \sim 200 \mu\text{s}$) (Figure 4.2.4F) might resemble the exchange between the two basins U (DSE) and partially unfolded states I (4, 19). The detailed characteristics of the DSE will be investigated in a further study (Chapter 5). To follow the unfolding of the C-terminal subdomain in higher detail, it would be necessary to probe the C-terminal subdomain by FRET pairs with a shorter Förster radius.

4.2.3.4. Conclusion

In summary, the generated distance network enabled us to describe the stepwise unfolding of structural elements in T4L. We were able to trace and disentangle the fast folding–unfolding processes, which are typically inaccessible by conventional methods, by advanced single-molecule methods (MFD, PDA, fullFCS and filteredFCS). This allowed us to map the dynamics of T4L over a broad range of timescales and to derive a kinetical folding scheme ranging from ns to ms, which captured all potential relevant dynamics. Given the structural similarities of *J* to the *NSE* it is reasonable to assume that N_{C3} and *J* are energetically close, as the previously found state N_{C3} is the least populated. A unified folding model which combines the *NSE* (Chapter 3), folding intermediates (20, 21, 41-44), the unfolded state and experimentally derived energetic considerations (4.3.1.11) is depicted in Figure 4.3.3. This unique combination of tools can be applied to other proteins in order to understand their unfolding schemes or to proteins that are difficult to structurally characterize due to their intrinsic dynamic behavior (i.e. intrinsically disordered proteins).

4.2.4. Materials & Methods

Sample preparation. Mutagenesis of the pseudo-wildtype (wt*) T4L gene in the pET11a vector was done according to standard procedures. Protein production, purification and labeling were performed as described elsewhere (94, 95) (Chapter 3). Detailed measurement and data analysis protocols for all experiments can be found in the supplementary information (4.3.1), a tabled summary of fitting results in Section 4.3.1.11.

Circular dichroism spectroscopy (CD). CD-spectra of unlabeled and labeled T4L were measured on a Jasco J-815 spectrometer equipped with a Pelletier temperature controlled sample holder. Spectra were measured at protein concentrations of $\sim 1 \mu\text{M}$ at 25°C . Unfolding experiments were carried out as a function of denaturant (50 mM Na-phosphate, 0 to 7.5 M urea). The samples were allowed to equilibrate to the various urea concentrations for at least 24 hr before measurements were conducted. Simultaneously to CD-spectra, absorption spectra from 200 to 260 nm were recorded to determine the exact protein concentration. Each spectrum was recorded three times, averaged, and corrected for buffer background.

Time-resolved fluorescence (eTCSPC). Ensemble time-resolved measurements were performed either using an IBH-500U (IBH, Scotland) or a Fluotime 200 (Picoquant, Germany). Single and double labeled samples were both measured up to a peak count of

100'000 photons. By convolution with the instrument response function, the obtained fluorescence decay curves were fitted to several models.

Single-molecule fluorescence. Single-molecule fluorescence measurements (MFD) were performed at room temperature in aqueous solution (50 mM Na-phosphate, 150 mM NaCl, 0 to 7.5 M urea, 1 μ M unlabeled T4L) as described elsewhere (Chapter 3). For Photon Distribution Analysis (PDA) the MFD data was used (48, 63-65). We applied PDA in two different modes, static (sPDA, 1 ms time windows (TW)) and dynamic (dPDA, global fit of 2 ms, 3 ms and 4 ms TW's). In sPDA, the data was fitted to a Gaussian distributed distance $\langle R_{DA} \rangle_E$, whereas dPDA assumes two dynamically exchanging states $\langle R_{DA} \rangle_{E,1}$ and $\langle R_{DA} \rangle_{E,2}$ and an additional static fraction. Both models include also a fraction of FRET inactive molecules.

fullFCS. All fullFCS measurements were conducted on a confocal fluorescence microscope (Olympus IX81, Hamburg, Germany). As excitation sources the parked beam of an Ar-ion laser (488 nm) or a laser diode (635 nm) (both Olympus) were used, both lasers were operated in continuous wave mode. The laser light was directed into a 60x water immersion objective (NA < 1.2) by a dichroic beam splitter and focused in the sample, illuminating a diffraction limited excitation volume. The emitted light was collected by the same objective and separated into the two polarizations (parallel and perpendicular, PBS 101, Thorlabs) relative to the exciting laser beam. The fluorescence signal is further divided into two spectral ranges (630 DCXR, Chroma), equipped with bandpass filters (HC 520/35 (Semrock) and HQ 715/120 (Chroma)) for donor (Alexa488) and acceptor (Alexa647) fluorescence, respectively. The signal from single photon sensitive detectors (MicroPhotonDevices PD5CTC for the green channels and PerkinElmer AQR-14 for the red channels) was recorded photon-by-photon in time-tagged-time-resolved (ttr) mode with picosecond accuracy (HydraHarp400, Picoquant, Berlin, Germany). Samples were diluted in NUNC chambers (Lab-Tek, Thermo Scientific, Germany) in a total of 500 μ L volume in buffer (50 mM Na-phosphate, 150 mM NaCl, 0 M to 7.5 M urea, 1 μ M unlabeled T4L). The sample concentration of single and double labeled protein was adjusted in the low nM range. Single labeled variants were measured for 2.5 hrs, double-labeled variants for 10 hrs.

The data was correlated in a time-window of 32 ps to 2 s and fitted with home written software to standard FCS models containing a diffusion term, a photon antibunching term, a term describing the rotational correlation times and additional bunching (autocorrelation functions, *ACF*) or antibunching (cross correlation functions, *CCF*) terms.

filteredFCS. FilteredFCS measurements (47, 70, 72) were performed at room temperature in aqueous solution (50 mM Na-phosphate, 150 mM NaCl, 0 to 7.5 M urea, 1 μ M unlabeled T4L) as described elsewhere (Chapter 3). The data was collected for 3 hr, correlated by applying species-specific filters in a time-window of 16 ns up to 4 s and fitted to a similar model as described for FCS consisting of a diffusion term and several correlation (species cross correlation functions, *sCCF*) and anti-correlation terms (species auto correlation functions, *sACF*).

Two-state model. All experimental observables displaying the sigmoidal behavior of a two state process (folded \rightarrow unfolded protein) were fitted using the two state model described by Santoro and Bolen in 1988 (7).

4.3. Supplementary Information

4.3.1. Methods

4.3.1.1. Experimental design

We use our toolkit of (fluorescence) spectroscopy methods to probe the unfolding of the model protein Phage T4 lysozyme (T4L) by addition of the denaturant urea. Here, we focus mainly on the variant 60-132 and support our findings with additional measurement from few other variants. Firstly, we determined the correct folding of all our used variants with CD-spectroscopy and then followed the dissolving of the proteins' secondary structure by increasing the concentration of denaturant.

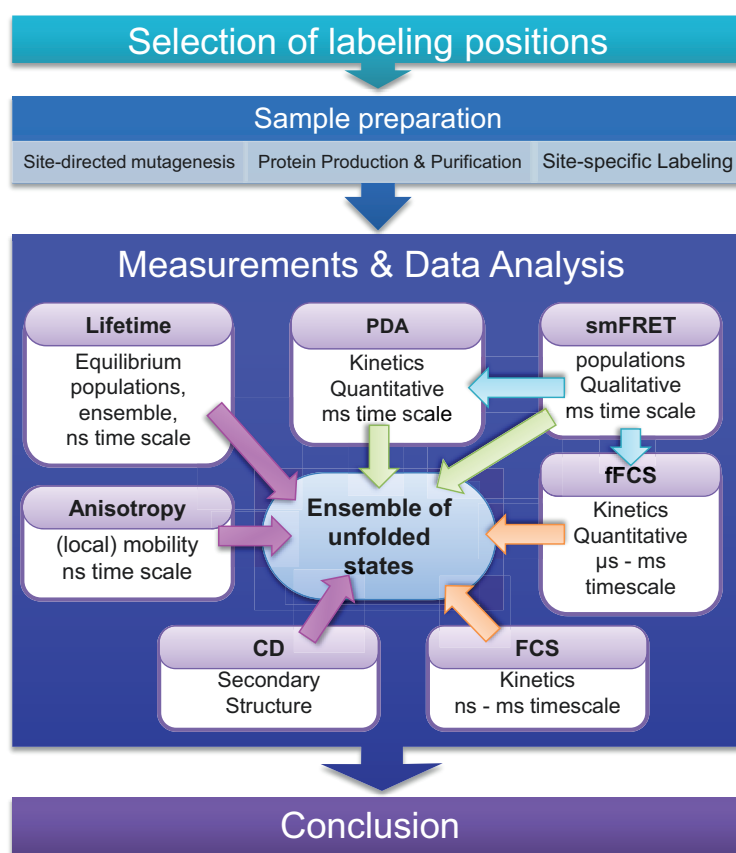


Figure 4.3.1 Flowchart of data acquisition and data analysis. The T4L variants used within this work are subjected to the following experiments while increasing the concentration of the denaturant urea: Circular Dichroism Spectroscopy (CD), steady-state fluorescence measurements, time-resolved fluorescence lifetime and anisotropy analysis using ensemble Time correlated Photon Counting (eTCSPC) and multiparameter fluorescence detection (smFRET). Photon distribution analysis (PDA) was used to analyze kinetics in the ms time range. Additionally, of the variant K60pAcF N132C fluorescence correlation spectroscopy (FCS) and *filtered* FCS (fFCS) measurements were conducted to resolve kinetics in the ns – ms (μ s – ms) time range.

All variants – single and double labeled – are measured under ensemble, equilibrium conditions using time-resolved ensemble Time Correlated Single Photon Counting (eTCSPC) to characterize the samples' ensemble behavior and the local mobility of the individual segments on the ns time scale. Single-molecule measurements using multiparameter

fluorescence detection (MFD) reveal the sample heterogeneity and Photon distribution analysis (PDA) of single-molecule bursts using multiple time windows allows the identification of dynamics in the ms time range. Finally, full fluorescence correlation spectroscopy (fullFCS) was used to resolve the dynamics in the ensemble of unfolded state (DSE) in the ns to low ms time scale. A “zoom-in” on dynamic processes influencing FRET was achieved by the use of *filtered* FCS (filteredFCS). A schematic overview for a better illustration is given in Figure 4.3.1.

4.3.1.2. *Protein production, purification and labeling*

The T4L variants were produced in *E. coli* and purified as described previously (Chapter 3). Labeling was done site-specifically using Alexa488-hydroxylamine as donor and Alexa647-maleimide as acceptor dye according to the protocol given in (94). 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.).

4.3.1.3. *Ensemble Time Correlated Single Photon Counting with high precision*

Data acquisition

Ensemble Time Correlated Single Photon Counting (eTCSPC) measurements were performed using either an IBH-500U (IBH, Scotland) or a Fluotime200 (Picoquant, Germany) system.

The excitation source of the IBH machine were a 470 nm diode laser (LDH-P-C470, Picoquant, Germany) operating at 10 MHz for donor excitation and at 635 nm (LDH-(-1-126, Picoquant, Germany) for acceptor excitation. The excitation and emission slits were set to 2 nm and 16 nm, respectively. The excitation source of the Fluotime200 system was a white light laser (SuperK extreme, NKT Photonics, Denmark) operating at 20 MHz for both donor (485 nm) and acceptor (635 nm) excitation with excitation and emission slits set to 2 nm and 5 nm, respectively. Additionally, in both systems, cut-off filters were used to reduce the amount of scattered light (>500 nm for donor and >640 nm for acceptor emission).

For green detection, the monochromator was set to 520 nm and for red detection to 665 nm. All measurements were conducted under magic angle conditions (excitation polarizer 0°, emission polarizer 54.7°), except for anisotropy where the position of the emission polarizer was alternately set to 0° (VV) or 90° (VH).

In the IBH system, the TAC-histograms were recorded with a bin width of 14.1 ps within a time window of 57.8 ns, while the Fluotime200 was set to a bin width of 8 ps within a time

window of 51.3 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* (~ 230 ps FWHM for the IBH, ~ 150 ps for the Fluotime200) 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.

For the IBH system, it was needed to perform before each measurement session a reference measurement with a continuous light signal 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.

Fitting procedure

Given different model functions described below the experimental fluorescence intensity decays F_{exp} 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, if necessary, 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 *Lin* (139). Given these corrections the experimental time-resolved fluorescence intensities of the samples F_{sample} are proportional to:

$$F_{sample}(t) \propto (F_{exp}(t) \otimes IRF + c) \cdot Lin \quad 4.3.1$$

Fluorescence lifetime and quantum yields of single labeled variants

The fluorescence decay of the single donor or acceptor labeled variants was multi-exponential, most likely due to local quenching and steric effects. To account for these effects, the single labeled variants were fitted by a multi-exponential relaxation model:

$$F_{D(0)/A(0)}(t) = \sum_i x^{(i)} \exp(-t / \tau^{(i)}) \quad 4.3.2$$

Here, $x^{(i)}$ are pre-exponential factors and $\tau^{(i)}$ the fluorescence lifetimes. The species-averaged lifetime $\langle \tau \rangle_x$ and fluorescence averaged lifetime $\langle \tau \rangle_f$ of the sample are calculated as:

$$\langle \tau \rangle_x = \sum_i x^{(i)} \cdot \tau^{(i)} \quad \text{and} \quad \langle \tau \rangle_f = \frac{1}{\langle \tau \rangle_x} \sum_i x^{(i)} \cdot (\tau^{(i)})^2 \quad 4.3.3$$

Assuming only dynamic quenching, the fluorescence quantum yield of dyes Φ_{Dye} can be determined using equation 4.3.4:

$$\phi_{Dye} = \frac{\langle \tau \rangle_{x,Dye} \cdot \phi_{ref}}{\langle \tau \rangle_{x,ref}} \quad 4.3.4$$

As reference, single labeled DNA with either Alexa488 ($\Phi_{D0} = 0.8$, $\langle \tau \rangle_{x,D0} = 4$ ns) or Cy5 ($\Phi_{A0} = 0.32$, $\langle \tau \rangle_{x,A0} = 1.17$ ns) were used.

Data analysis of double-labeled variants

Determination of the Förster radius R_0

Depending upon the distance R_{DA} between the donor and the acceptor fluorophore in the biomolecule, the donor fluorescence might be quenched due to Förster Resonance Energy Transfer (FRET) which results in a decreased species weighted average fluorescence lifetime of the DA-sample when compared to the DOnly labeled sample and the appearance of measurable acceptor fluorescence. The amount of quenching, i.e. the FRET-efficiency E , is highly distance dependent as given by the Förster formula (eq. 4.3.5):

$$E = (1 + R_{DA}^6 / R_0^6)^{-1} \quad 4.3.5$$

with R_0 being the Förster radius, which depends on the refractive index of the surrounding medium n , the spectral overlap J between the donor emission and the acceptor absorption spectrum, the fluorescence quantum yield of the donor Φ_{D0} and the orientation factor κ^2 . The fluorophores are coupled to the biomolecule by a long, flexible linker. Therefore, we assume isotropic averaging of donor-acceptor dipole orientation and we use the approximation $\kappa^2 \approx 2/3$. This is supported by the anisotropy studies shown in (Chapter 3, herein Table 3.3.3). We found that n and J depend on the urea concentration, whereas Φ_{D0} is sample-dependent. The rate constant of the FRET-process is independent of the donor fluorescence quantum yield. Therefore, we define a reduced Förster-radius R_{0r} solely dependent of the spectral overlap, the orientation factor and the refractive index (eq. 4.3.6):

$$R_{0r} = \left[\frac{9(\ln 10)}{128\pi^5 \cdot N_A} \cdot \frac{J \cdot \kappa^2}{n^4} \right]^{\frac{1}{6}} = 0.2108 \cdot \text{\AA} \cdot \left[\frac{\kappa^2}{n^4} \cdot \left(\frac{J(\lambda)}{\text{mol}^{-1} \cdot \text{dm}^3 \cdot \text{cm}^{-1} \cdot \text{nm}^4} \right) \right]^{\frac{1}{6}} \quad 4.3.6$$

with N_A being Avogadro's constant. R_{0r} is 56 Å in the absence of urea and 54 Å in buffer containing 7.5 M urea. To precisely analyze the contribution of quenching due to FRET and distinguish from the possibly present local quenching of the donor, we analyze the DA-sample globally with its respective DOnly-sample. The FRET-induced donor quenching can be described by different models adapted to the particular question. Here we use two different models: a formal description by Gaussian distributions and a Worm-like chain model.

We account for donor labeled molecules lacking an acceptor by including the fraction of DOnly- molecules (x_{DOnly}) in the analysis model using a modified version of eq. 3.2.1:

$$\begin{aligned} F_{\text{FRET}}(t) &\propto \left[(1 - x_{\text{DOnly}}) F_{\text{D(A)}}(t) + x_{\text{DOnly}} F_{\text{D(0)}}(t) \right] \otimes \text{IRF} + c) \cdot \text{Lin} \\ F_{\text{Ref}}(t) &\propto (F_{\text{D(0)}}(t) \otimes \text{IRF} + c) \cdot \text{Lin} \end{aligned} \quad 4.3.7$$

Gaussian distributed distances-fitting (RDA)

The dyes are coupled to the biomolecules of interest via long and flexible linkers (~20 Å). This has to be accounted for in the analysis. The flexible linking assures that the dyes can move freely. The inter-dye correlation time is approximately (100-200 ns) (140). This is significantly faster than the integration time of the experiment but slower than the time-scale of fluorescence (4 ns). Therefore, the broadening of the distance distributions due to the linkers have to be considered. The distance R_{DA} between the FRET pair in one conformational state can then not be described by one distance but is instead modeled as a probability function $p(R_{DA})$ showing Gaussian distribution characteristics with a mean distance $\langle R_{DA} \rangle$ and width w_{DA} :

$$p(R_{DA}) = \frac{1}{w_{DA} \sqrt{\pi/2}} \exp \left(-2 \left[\frac{R_{DA} - \langle R_{DA} \rangle}{w_{DA}} \right]^2 \right) \quad 4.3.8$$

$$F_{\text{D(A)}}(t) = F_{\text{D(0)}} \cdot \int_{R_{DA}} p(R_{DA}) \cdot \exp \left(-t \cdot k_0 \left[1 + (R_0 / R_{DA})^6 \right] \right) dR_{DA} \quad 4.3.9$$

k_0 is the radiative rate constant of fluorescence ($k_0 = 0.224 \text{ ns}^{-1}$). The linkers used in the course of this work have a length of ~ 20 Å and it was estimated using Accessible Volume Simulations (53) that a width of ~ 6 Å is to be expected in the folded protein.

Wormlike chain model (WLC)

Another class of fitting models used to describe a special case of biomolecules are polymer models. Derived from polymer theory, they are applied when it is likely to assume an unfolded biomolecule, i.e. under highly denaturing conditions, e.g. in denaturants as urea and guanidine hydrochloride. Under those conditions, the majority, but by no means all, of the

tertiary and even secondary interactions are lost and the protein behaves as a polymer. Given the probability distribution function of the donor-acceptor distances $p(R_{DA})$ the mean transfer efficiency $\langle E \rangle$ is obtained by:

$$\langle E \rangle = \frac{\int_0^\infty E(R_{DA}) p(R_{DA}) dR_{DA}}{\int_0^\infty p(R_{DA}) dR_{DA}} = \frac{\int_0^\infty (1 + R_{DA}^6 / R_0^6)^{-1} p(R_{DA}) dR_{DA}}{\int_0^\infty p(R_{DA}) dR_{DA}} \quad 4.3.10$$

Here, $E(R_{DA})$ is the distance-dependent FRET which can be described by the Förster formula and $p(R_{DA})$ is the probability distribution of the polymer chain of having a population with the distance R_{DA} between the two dyes. The exact form of $p(R_{DA})$ depends on the assumed polymer model. In the simplest case, the peptide chain shows random coil or freely joint chain characteristics where the fluorescence decay can be approximated by the Gaussian Chain Model (14). This model assumes that each bond has a fixed length and that adjacent bond angles are uncorrelated. This results in a Gaussian distribution:

$$p(R_{DA}) = \frac{4\pi^2}{(2/3 \cdot \langle R_{DA}^2 \rangle)^{3/2}} \exp\left(-\frac{3R_{DA}^2}{2\langle R_{DA}^2 \rangle}\right) \quad 4.3.11$$

With $\sqrt{\langle R_{DA}^2 \rangle} = l \cdot n^{1/2}$ being the root-mean-square, end-to-end length of the chain (16). l is the bond length (3.6 Å per amino acid residue) and n the number of bonds between the two dyes. Thus, n is the only free fitting parameter.

However, for the model protein T4L used here, fitting with the above described model could not be achieved with a satisfactory result, i.e. flat distribution of weighted residuals and a $\chi^2_r \sim 1$. Thus, it was necessary to use the more complex model Worm-like chain. Here, the biopolymer is assumed to be a semi-flexible rod of length L and stiffness κ and is characterized by a so-called persistence length l_p , which is defined as the product of L and κ . Instead of a fixed segment length based on the monomers of the polymer, the chain is thought of being comprised of longer, uncorrelated links (12) and the persistence length is defined as this length over which the correlations between the bond angles persevere. The bending energy E associated with a chain configuration $r(l)$ is given by:

$$\beta \cdot E = \frac{\kappa}{2} \int_0^l \|\partial_l t\|^2 dl \quad 4.3.12$$

where $\beta = 1/k_B T$ is the inverse temperature, l is the arclength, and $t = \partial_l r$ is the unit tangent vector ($\|t\| = 1$) (141). Unfortunately, it lacks an exact numerical solution and the

approximations summarized and derived in (141) were thus implemented in the fitting routine. Inserting the given constants a , b , c , $(c_{ij})_{i,j}$ and d into equation 21 of the paper from Becker *et al* (141), leads to equation 4.3.13:

$$p(R_{DA}) = \frac{1-c \cdot R_{DA}^2}{(1-R_{DA}^2)^{5/2}} \cdot e^{\left(\frac{d \cdot \kappa \cdot a \cdot b \cdot (1+b)}{1-(bR_{DA})^2} \cdot R_{DA}^2 \right)} \cdot e^{\left[\left(\frac{3}{4} \frac{1}{\kappa} - \frac{1}{2} \right) \cdot R_{DA}^2 + \left(\frac{23}{64} \frac{1}{\kappa} + \frac{17}{16} \right) \cdot R_{DA}^4 + \left(\frac{7}{64} \frac{1}{\kappa} - \frac{9}{16} \right) \cdot R_{DA}^6 \right] \cdot \frac{1}{1-R_{DA}^2}} \cdot I_0 \left(\frac{-d \cdot k \cdot a \cdot (1+b) \cdot R_{DA}}{1-(bR_{DA})^2} \right) \quad 4.3.13$$

$a = 14.054$
 $b = 0.473$
with: $c = 1 - (1 + (0.38 \cdot \kappa^{-0.95})^{-5})^{-1/5}$
 $d = \begin{cases} 1 & \kappa < 1/8 \\ 1 - \frac{1}{0.177 / (\kappa - 0.111) + 6.40(\kappa - 0.111)^{0.783}} & \text{otherwise} \end{cases}$
 I_0 : Bessel - function of Order Zero

4.3.1.4. **Residual Fluorescence Anisotropy**

To determine the residual anisotropy r_{ss} , the excitation and emission polarization dependent fluorescence (I_{VV} , I_{VH} , I_{HV} , I_{HH}) is measured. r_{ss} can then be calculated according to

$$r_{ss} = \frac{I_{VV} - gI_{VH}}{I_{VV} + 2gI_{VH}} \quad \text{with} \quad g = \frac{I_{HV}}{I_{HH}} \quad 4.3.14$$

where I_{VV} and I_{VH} , denote the fluorescence signal measured either vertically (0°) or horizontally (90°) polarized after vertically polarized excitation, and I_{HV} and I_{HH} denote the fluorescence signal measured either vertically (0°) or horizontally (90°) polarized after horizontally polarized excitation.

4.3.1.5. **Multiparameter Fluorescence Detection**

Data acquisition and analysis for MFD was done as described elsewhere (Chapter 3)(25). As measurement buffer PBS, pH 7.5, containing additionally different concentrations of urea were used. Additionally, 1 μ M of unlabeled T4L wt* was added to coat the surface of the measurement chamber and thus, minimize absorption of the labeled protein to it.

4.3.1.6. **Guidelines for reading MFD histograms**

To be able to properly read and thus understand MFD histograms correctly, one has to take care of some guidelines. A short list is presented here and for more detailed information the reader is referred to (25).

I) Donor only population is shown at high F_D/F_A ratio with lifetime ~ 4 ns (donor-only for Alexa488).

II) High FRET appears at shorter lifetimes when the fluorescence of acceptor is high ($F_D/F_A \rightarrow 0$).

III) Static FRET states distributions have elliptical 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 (63).

IV) Static FRET states follow a theoretical line that accounts for dye linker mobility called "static FRET line" Eq. 4.3.11 (63).

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 averaged lifetime, $\langle \tau_{D(A)} \rangle_f$. Therefore, dynamics in sub-ms range is expected when the shape is not elliptical. 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.

4.3.1.7. Photon Distribution Analysis

Photon or Probability distribution analysis (PDA) uses the photon traces of single-molecule measurements to deduce various FRET or fluorescence anisotropy related parameters (64) and displays them in occurrence histograms. The build histograms are then compared with theoretical distributions of the respective parameters (probability histograms). However, critical care has to be taken for used values of crosstalk, background, fluorescence quantum yields (48). Thus, it is possible to distinguish an extra broadening of histograms due to measurement artefacts from actual broadening due to sample heterogeneities. It is possible also to detect and describe simple dynamics in a sample by time-window analysis (TWA). Here, the photon trace is binned in different time-windows (TW's), i.e. 0.5 ms, 1 ms or even 5 ms (65). If a system undergoes dynamic exchange between two or more states in the time range of the diffusion time (\sim low ms range), which also limits the usable minimal/maximal time window for a sample, this can be easily detected by eye with TWA. The dynamics of this time range will induce a time-window dependent narrowing/broadening of the obtained e.g. FRET-efficiency histograms. A detailed theoretical deduction can be found in (48, 64, 65).

In the course of this work, static PDA was used to estimate the mean and the half-width of averaged on ms time scale distances distribution between the dyes attached to the sample.

The mean distance was modeled as being Gaussian distributed with a limited half-width, i.e. the half-width was given in % of the distance and thus rising with an increasing distance between the FRET pair. Besides, a fraction of DOnly-labeled molecules was included.

4.3.1.8. *Full Fluorescence Correlation Spectroscopy (fullFCS)*

Data acquisition

All fullFCS (from ps to s) measurements were conducted on a confocal fluorescence microscope (Olympus IX81, Hamburg, Germany). As excitation sources the parked beam of an Ar-ion laser (488 nm) or a laser diode (635 nm) (both Olympus) were used. Both lasers were operated in continuous wave mode. The laser light was directed into a 60x water immersion objective (NA = 1.2) by a dichroic beam splitter and focused in the sample, illuminating a diffraction limited excitation volume. The emitted light was collected by the same objective and separated into the two polarizations (parallel and perpendicular, PBS 101, Thorlabs) relative to the polarization of the exciting laser beam. The fluorescence signal is further divided into two spectral ranges (630 DCXR, Chroma), equipped with bandpass filters (HC 520/35 (Semrock) and HQ 715/120 (Chroma)) for donor (Alexa488) and acceptor (Alexa647) fluorescence, respectively. The signal from single photon counting detectors (MicroPhotonDevices PD5CTC for the green channels and PerkinElmer AQR-14 for the red channels) was recorded photon-by-photon in time-tagged-time-resolved (tttr) mode with picosecond accuracy (HydraHarp400, Picoquant, Berlin, Germany).

For fullFCS measurements, prior to each sample measurement, a diluted nM concentration of Rh110 and Atto647N, respectively, was measured for 1 min in buffer containing no urea and in buffer containing the respective urea concentration, and the collar ring was adjusted such that the counts per molecule got maximal, i. e. the detection volume minimal. These reference measurements were necessary to correct for the influence of urea on the index of refraction, changing the size and shape of the observation volume and on the viscosity. Samples were diluted in NUNC chambers (Lab-Tek, Thermo Scientific, Germany) in a total of 500 μ L volume in the respective measurement buffer containing 50 mM Na-phosphate, 150 mM NaCl and between 0 M and 7.5 M urea. Additionally, 1 μ M of unlabeled T4L was added to coat the chamber surface. The sample concentration was adjusted in the low nM range, yielding in average ~ three molecules in the focus at a time.

For single labeled samples (DOnly and AOnly), photons were collected for 2.5 hrs resulting in average about $3 \cdot 10^9$ collected photons, while for double labeled samples (DA) the measurement time was increased to 10 hrs resulting in average about $8 \cdot 10^9$ collected

photons. The power at objective for green excitation was 20 μW (485 nm) and for red excitation 10 μW (635 nm).

Data analysis

Prior to any data analysis, the saved data files ranging from 10 GB for the measurements on single labeled variants up to 50 GB for those on double labeled variants had to be split in smaller files for being able to perform the subsequent correlations on a standard personal computer. Files were split in fractions of three million photons, resulting into split file size of ~ 11 MB. Splitting and correlation was performed using a Labview-based program. Photons were correlated within a time window of 32 ps up to 2 s. The correlation function is calculated as follows:

$${}^{A,B}G(t_c) = 1 + \frac{\langle \delta^A S(t) \cdot \delta^B S(t + t_c) \rangle}{\langle {}^A S(t) \rangle \cdot \langle {}^B S(t) \rangle} \quad 4.3.15$$

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

Keeping in mind our setup with four channels (green parallel (Gp), green perpendicular (Gs), red parallel (Rp) and red perpendicular (Rs)), the correlation of the double labeled samples yielded in total twelve curves: four color- autocorrelation curves (GpGs, GsGp, RpRs, RsRp) and eight color-cross correlation curves (GpRp, RpGp, GsRs, RsGs, GpRs, RsGp, GsRp, RpGs). Single labeled samples, of course, yielded only the two respective color-autocorrelation curves.

The general correlation function $G(t_c)$ of a solution with N fluorescent molecules in the focus, which displays no further fluctuation contributions due to e.g. enzyme kinetics, follows the analytical form (69):

$$G(t_c) = 1 + \frac{1}{N} \cdot G_{diff}(t_c) = 1 + \frac{1}{N} \cdot \left(1 + \frac{t_c}{t_{diff}} \right)^{-1} \cdot \left(1 + \left(\frac{\omega_0}{z_0} \right)^2 \cdot \frac{t_c}{t_{diff}} \right)^{-\frac{1}{2}}. \quad 4.3.16$$

As detection volume a 3-dimensional Gaussian shaped element with spatial distribution of the detection probabilities was assumed: $W(x, y, z) = \exp(-2(x^2 + y^2)/\omega_0^2) \cdot \exp(-2z^2/z_0^2)$. The $1/e^2$ radii in x and y or in z direction are denoted by ω_0 and z_0 , respectively. The diffusion

time of particles, t_{diff} is related to the diffusion constant D as $t_{diff} = \omega_0^2 / 4D$. The amplitude of the correlation function at $t \rightarrow 0$ is scaled with the reciprocal of the average number of fluorescent particles N in the confocal volume.

The presence of additional processes apart from diffusion which lead to fluctuations in the fluorescence signal e.g. quenching kinetics, triplet formation and photon antibunching are inducing additional characteristic decay or rise terms in the correlation curve (69).

Thus, with assumption that all characteristic correlation times are well separated, the shape of the color- autocorrelation curves of our single and double labeled samples can be described by the following analytical formula:

$$G(t_c) = b + \frac{1}{N} \cdot G_{diff}(t_c) \cdot G_{abun}(t_c) \cdot G_{rc}(t_c) \cdot G_{bun}(t_c) \quad 4.3.17$$

with b being a constant offset, $G_{abun}(t_c)$ describing the photon antibunching term, $G_{rc}(t_c)$ a term associated with the rotational correlation of the molecule and $G_{bun}(t_c)$ describes all further (protein-motion induced) processes which lead to additional decay terms:

$$\begin{aligned} G_{abun}(t_c) &= 1 - x_{ab} \cdot e^{-t_c/t_{ab}} \\ G_{rc}(t_c) &= 1 + x_{rc} \left(\frac{1}{1 + x_{rc2}} \cdot e^{-t_c/t_{rc1}} + \frac{x_{rc2}}{1 + x_{rc2}} \cdot e^{-t_c/t_{rc1} \cdot t_{rc}} \right) \\ G_{bun}(t_c) &= 1 + x_{b1} - x_{b1} \cdot e^{-t_c/t_{b1}} + x_{b2} - x_{b2} \cdot e^{-t_c/t_{b2}} + x_{b3} - x_{b3} \cdot e^{-t_c/t_{b3}} + x_{b4} - x_{b4} \cdot e^{-t_c/t_{b4}} \end{aligned} \quad \begin{array}{l} 4.3.18 \\ \text{a-c} \end{array}$$

The photon antibunching term is described by its amplitude x_{ab} (usually ~ 1) and time constant t_{ab} which is the reciprocal sum of the rate constants for excitation and fluorescence decay (~ 4 ns for the donor, ~ 2 ns for the acceptor). The rotational correlation term, whose amplitude x_{rc} is polarization dependent, contains up to two time constants t_{rc1} and $t_{rc2} = t_{rc} \cdot t_{rc1}$, their respective amplitudes described by the exponential prefactors. To be able to describe the correlation function to a sufficient level (flat residuals), we needed additionally up to four bunching times t_{b1} to t_{b4} – depending on the urea concentration. The bunching amplitudes are defined by the respective prefactors (or fractions) x_{b1} to x_{b4} and normalized such that $x_{b1} + x_{b2} + x_{b3} + x_{b4} = 1$.

Corresponding color- autocorrelation curves of single and double labeled samples were fitted globally, i.e. the antibunching, the rotational correlation and all bunching times were identical for all curves while amplitudes of antibunching and bunching terms were kept identical for each pair of curves, e.g. for DOnly-GsGp and DOnly-GpGs as well as for DA-GsGp and DA-GpGs. In contrast, the amplitudes of the rotational correlation terms were left free for

individual fitting due to their polarization dependency. In general, especially at low concentration where additional enzyme kinetic based fluorescence fluctuations were expected, for the DA samples more bunching times than for the single labeled ones were needed. Respective amplitudes were set to zero in the single labeled curves.

In the color- cross correlation curves, processes which are related to the FRET are detected as anticorrelation terms. Additionally, the photon antibunching term $G_{abun}(t_c)$ and a polarization-dependent bunching term $G_{rc,c}(t_c)$, related to the rotational correlation, are needed to fully describe the form of the color- cross correlation function:

$$G(t_c) = b + \frac{1}{N} \cdot G_{diff}(t_c) \cdot G_{abun}(t_c) \cdot G_{rc,c}(t_c) \cdot G_{cab}(t_c) \quad 4.3.19$$

where

$$G_{rc,c}(t_c) = 1 - x_{rc} + x_{rc} \cdot e^{-t_c/t_{rc}} \quad 4.3.20$$

$$G_{cab}(t_c) = 1 - f \cdot (x_{cab1} \cdot e^{-t_c/t_{cab1}} + x_{cab2} \cdot e^{-t_c/t_{cab2}} + x_{cab3} \cdot e^{-t_c/t_{cab3}} + x_{cab4} \cdot e^{-t_c/t_{cab4}} + x_{cab5} \cdot e^{-t_c/t_{cab5}})$$

$G_{diff}(t_c)$ and $G_{abun}(t_c)$ are identical to the respective terms of the color- autocorrelation function. The bunching term $G_{rc,c}(t_c)$ is described by its amplitude x_{rc} and time constant t_{rc} . The antibunching terms are described by a form factor f , the time constants t_{cab1} to t_{cab5} and their respective amplitudes (fractions) x_{cab1} to x_{cab5} which are normalized such that $x_{cab1} + x_{cab2} + x_{cab3} + x_{cab4} + x_{cab5} = 1$.

For fitting, all parameters except N , f and x_{rc} were set globally over all eight color-cross correlation curves. In low urea concentration, additionally another antibunching term could not be set globally as it also displayed a polarization dependent amplitude.

4.3.1.9. Filtered Fluorescence Correlation Spectroscopy (filteredFCS)

Data acquisition and analysis was performed as described previously (Chapter 3)(47).

Filtered FCS requires prior knowledge of the time-resolved fluorescence and polarization decays for each species (47, 70). In contrast to fullFCS here pulsed excitation is required. For a mixture of more than two species, we generated two “patterns” (filter) corresponding to two “pseudo-species” termed low FRET (LF) and high FRET (HF). For this, single-molecule measurements of the same molecules and conditions were used. The bursts were analyzed using 1 ms time- windows and plotted in a two- dimensional histogram of number of detected photons versus proximity ratio. Based on our experience, the low FRET species was defined as all TW’s displaying a proximity ratio between 0 and 0.2, while the filter for the high FRET species was generated from bursts showing a proximity ratio of ~ 0.6 up to 1. Here, the exact lower border was dependent upon variant and urea concentration.

The yielded correlated curves (two $sCCFs$ (LF-> HF, HF-> LF) and two $sACFs$ (LF->LF, HF-> HF) were fitted globally using similar equations as described in 4.3.1.8. Again, species cross-correlations display anticorrelation while species auto-correlations show bunching terms. Unlike fullFCS, for $sCCFs$ and $sACFs$ no photon antibunching or rotational correlation bunching terms are observed. Thus, the needed equations simplify to:

$$G_{sCCF}(t_c) = b + \frac{1}{N} \cdot G_{diff}(t_c) \cdot G_{ab}(t_c) \cdot G_{slow}(t_c) \quad 4.3.21$$

$$G_{sACF}(t_c) = b + \frac{1}{N} \cdot G_{diff}(t_c) \cdot G_b(t_c) \cdot G_{slow}(t_c) \quad 4.3.22$$

with $G_{ab}(t_c)$, $G_b(t_c)$ and $G_{slow}(t_c)$:

$$G_{ab}(t_c) = 1 - f \cdot (x_{ab1} \cdot e^{-t_c/t_{ab1}} + x_{ab2} \cdot e^{-t_c/t_{ab2}} + x_{ab3} \cdot e^{-t_c/t_{ab3}} + x_{ab4} \cdot e^{-t_c/t_{ab4}}) \quad 4.3.23$$

$$G_b(t_c) = 1 - x_{ab1} + x_{ab1} \cdot e^{-t_c/t_{ab1}} - x_{ab2} + x_{ab2} \cdot e^{-t_c/t_{ab2}} - x_{ab3} + x_{ab3} \cdot e^{-t_c/t_{ab3}} - x_{ab4} + x_{ab4} \cdot e^{-t_c/t_{ab4}} \quad a,b$$

$G_{slow}(t_c)$ is modeled as an additional correlation ($sACF$) and anti-correlation ($sCCF$) term, respectively. All time constants are kept global, while amplitudes are only linked for both $sCCFs$. The sum of the exponential prefactors, i.e. amplitudes of time constants, is normalized to 1.

4.3.1.10. Two-state model

All experimentally observables whose values displayed the typical development of a two-state system, i.e. the folded protein unfolds directly into its denatured state without any intermediate, were fitted with the model developed by Santoro and Bolen in 1988 (7):

$$y = \frac{y_N + m_N \cdot c + (y_D + m_D \cdot c) \cdot \exp(-\Delta G^0 - m \cdot c / R \cdot T)}{1 + \exp(-\Delta G^0 - m \cdot c / R \cdot T)} \quad 4.3.24$$

with y_N and y_D being the native and denatured state baseline, m_N and m_D the native and denatured state slope, m is the slope of rise part and ΔG^0 the free energy of unfolding.

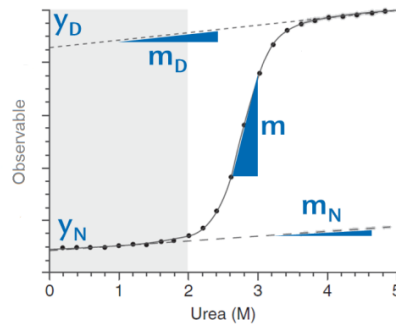
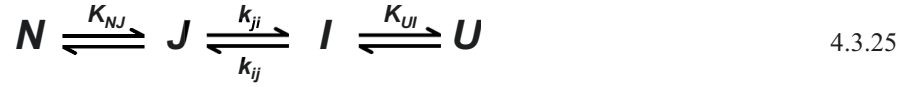


Figure 4.3.2 Development of an experimentally observable in a two state system.

4.3.1.11. Calculation of the energy landscape

Cellitti *et al.* (20) determined the stability of the native and unfolded state and the two intermediates as a function of the denaturant urea. We used their values (Table 2 in (20)) and calculated the energy landscapes using the four-state model:



$$K_{UI} = K_{UI}^{H_2O} \cdot \exp\left(\frac{m_{UI}}{RT} \cdot c\right) \quad K_{NJ} = K_{NJ}^{H_2O} \cdot \exp\left(\frac{m_{NJ}}{RT} \cdot c\right) \quad k_{IJ} = k_{IJ}^{H_2O} \cdot \exp\left(\frac{m_{IJ}}{RT} \cdot c\right) \quad k_{JI} = k_{JI}^{H_2O} \cdot \exp\left(\frac{m_{JI}}{RT} \cdot c\right)$$

c (in mol/L) is the urea concentration, R the universal gas constant ($1.99 \cdot 10^{-3}$ Kcal/(K*mol)) and T the temperature (295 K). However, as we noticed a typing error in their equation 12-13 (44) and that they divided their final values by 2, we provide here the corrected equations (eq. 4.3.25-4.3.26), but keep the division. Assuming both a fast pre-equilibrium from $N \rightarrow J$ and $U \rightarrow I$, the stabilities of the individual states can be calculated according to:

$$\text{state energy:} \quad \Delta G_{UI} = -RT \ln K_{UI} + m_{UI} \cdot c \quad \Delta G_{NJ} = -RT \ln K_{NJ} + m_{NJ} \cdot c \quad 4.3.26$$

$$\text{activation energy:} \quad \Delta G_{IJ}^\ddagger = -RT \ln \left(\frac{k_{IJ}}{k_0} \right) + m_{IJ} \cdot c \quad \Delta G_{JI}^\ddagger = -RT \ln \left(\frac{k_{JI}}{k_0} \right) + m_{JI} \cdot c \quad 4.3.27$$

with $k_{IJ} = 15.8 \text{ s}^{-1}$, $k_{JI} = 5 \cdot 10^{-4} \text{ s}^{-1}$, $K_{UI} = 1926 \text{ s}^{-1}$, $K_{NJ} = 1.3 \cdot 10^{-4} \text{ s}^{-1}$, $m_{IJ} = 0.04 \text{ cal}^* \text{L/mol}^2$, $m_{JI} = -0.3 \text{ cal}^* \text{L/mol}^2$, $m_{UI} = 1.6 \text{ cal}^* \text{L/mol}^2$ and $m_{NJ} = -0.7 \text{ cal}^* \text{L/mol}^2$. k_0 was set to 10^9 s^{-1} . For the ensemble of native states, we added the stabilities obtained from our own measurements (Chapter 3). Here, we used the species fractions x_i for the free enzyme to adjust the average stability of N_{C1} ($x_1 = 0.54$), N_{C2} ($x_2 = 0.30$), and N_{C3} ($x_3 = 0.16$) such that it resembles the value measured from Cellitti *et al.* for N (compare Chapter 3). Relative to N_{C1} , N_{C2} is 0.58 kcal/mol and N_{C3} is 1.21 kcal/mol more stable than N_{C1} . Thus, the stabilities relative to the measured ensemble of average N from (20) were calculated as:

$$\begin{aligned} N_{C1}: \quad \Delta G_{N_{C1}} &= \Delta G_{NJ} - x_2 \cdot \Delta G_{N_{C1}N_{C2}} - x_3 \cdot \Delta G_{N_{C1}N_{C3}} \\ N_{C2}: \quad \Delta G_{N_{C2}} &= \Delta G_{N_{C1}} + \Delta G_{N_{C1}N_{C2}} \\ N_{C3}: \quad \Delta G_{N_{C3}} &= \Delta G_{N_{C1}} + \Delta G_{N_{C1}N_{C3}} \end{aligned} \quad 4.3.28$$

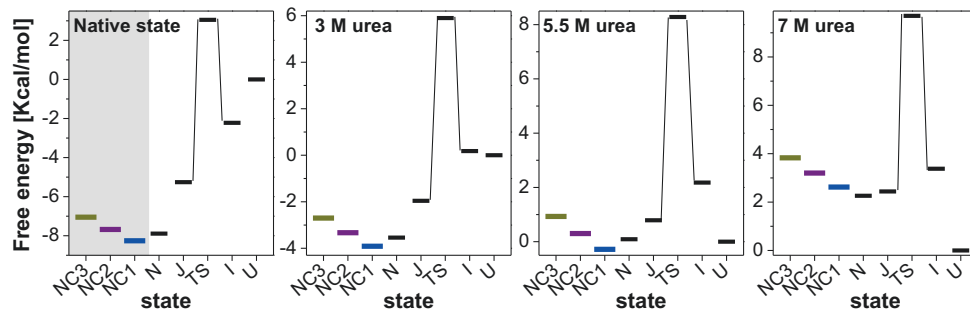


Figure 4.3.3 Energy landscape calculated according to the values obtained in (44) and Chapter 3 using equations 4.3.25-4.3.28. We assumed that the energy difference between the native states stays constant.

Figure 4.3.3 shows how the energy landscape tilts upon the addition of urea. In the absence of urea, the native state is the most stable conformation, while with increasing concentration of urea the native state is destabilized and the intermediates and the unfolded state, respectively, are stabilized. Note that N_{C3} is an exited state and very close in energy to J .

4.3.2. Supporting results

4.3.2.1. CD-spectroscopy of further variants

Circular dichroism spectroscopy under native conditions in the absence of urea proved that all variants were correctly folded (Figure 4.3.4A). Upon addition of urea, the secondary structure elements are dissolved and the CD- signal measured at 222 nm is lost. Figure 4.3.4B shows that all except one variant unfold apparently in a two state manner. Only one variant, R8pAcF P86C, displays a clear three state behavior which is most probably caused by the position of the donor dye, R8pAcF. This label was placed in the first α -helix, which structurally belongs to the C-terminal domain but sequentially to the N-terminal domain. It is only weakly attached to rest of the C-terminal domain, thus it might be that by introducing the label there, the connection is broken easier and the helix is being dissolved.

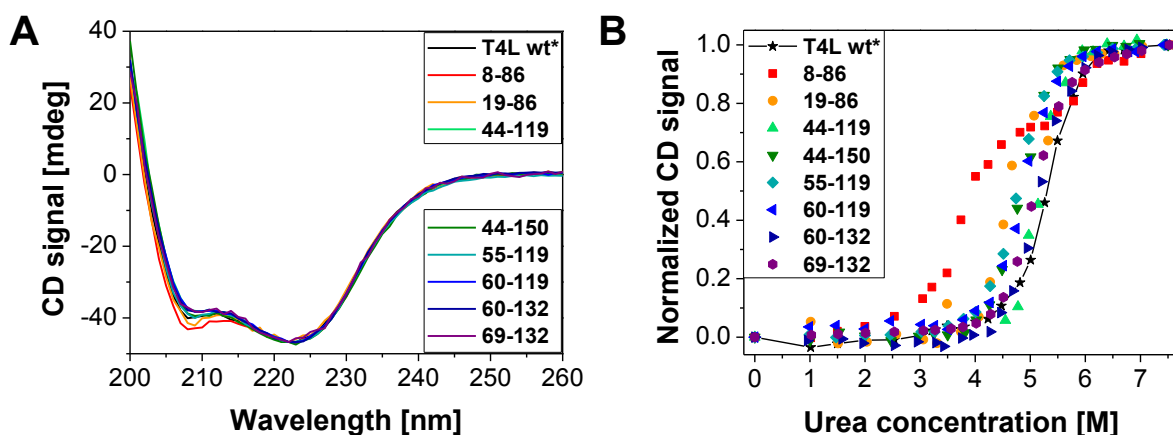


Figure 4.3.4 Summary of CD spectroscopy results. (A) Overlay of CD spectra from all variants and the cysteine-free wild type T4L wt* in absence of urea to check for proper folding. To correct for slight differences in protein concentration, the spectra have been normalized to the concentration of T4L wt*. (B) Normalized CD signal at 222 nm measured in increasing concentrations of urea. All variants show the behavior of a typical two-state folder except 8-86.

4.3.2.2. Fluorescence lifetime analysis of double-labeled samples

The time-resolved fluorescence decay of single and double labeled samples was analyzed with a several models as described in 4.3.1.3. The results of fitting the Donor-only labeled samples with a multiexponential model are summarized in Table 4.3.2A, additionally the species-averaged lifetime of direct acceptor excited molecules and the residual fluorescence anisotropies are given.

UNFOLDING PATHWAY OF T4L

Table 4.3.1 Fitting results of fluorescence lifetime of donor only labeled variants. $\langle\tau\rangle_x$ is the species weighted fluorescence lifetime, the fluorescence quantum yield Φ_x is calculated using a reference with known fluorescence quantum yield (Donor: $\langle\tau\rangle_{A488} = 4.0$ ns, $\Phi_{A488} = 0.8$; Acceptor: $\langle\tau\rangle_{Cy5} = 1.17$ ns, $\Phi_{Cy5} = 0.32$)

[M urea]	τ_1 [ns]	x_1	τ_2 [ns]	x_2	τ_3 [ns]	x_3	τ_x [ns]	τ_f [ns]	Φ_D	$r_{ss,D}$	χ^2_r	τ_x [ns]	Φ_A	$r_{ss,A}$
R8pAcF P86C														
0	4.16	0.92	1.88	0.08			3.98	4.07	0.8	0.18	1.13	1.25	0.34	0.24
2	4.24	0.65	3.11	0.30	0.55	0.06	3.70	3.93	0.74	0.17	1.03	1.32	0.36	0.24
3	4.28	0.52	3.37	0.42	0.81	0.06	3.70	3.89	0.74	0.16	0.96	1.33	0.36	0.24
4	4.14	0.59	3.03	0.33	0.69	0.08	3.51	3.77	0.70	0.15	1.05	1.36	0.37	0.25
4.5	4.08	0.60	2.93	0.31	0.69	0.09	3.42	3.72	0.68	0.13	1.03	1.43	0.39	0.24
5	3.88	0.80	1.92	0.11	0.26	0.09	3.33	3.73	0.67	0.12	1.15	1.52	0.42	0.24
5.5	3.88	0.73	2.45	0.19	0.54	0.08	3.32	3.63	0.66	0.11	1.00	1.42	0.39	0.24
6	3.87	0.85	2.44	0.10	0.60	0.05	3.56	3.74	0.71	0.10	0.93	1.51	0.41	0.23
6.5	3.97	0.59	3.04	0.33	0.73	0.08	3.40	3.64	0.68	0.10	1.01	1.43	0.39	0.23
7	3.84	0.78	2.39	0.16	0.47	0.06	3.40	3.64	0.68	0.10	1.02	1.41	0.38	0.23
7.5	3.75	0.82	1.65	0.09	0.21	0.09	3.25	3.64	0.63	0.10	1.05	1.50	0.41	0.24
K19pAcF P86C														
0	3.81	0.92	1.35	0.08			3.61	3.74	0.72	0.14		1.39	0.38	0.25
2	4.02	0.63	2.87	0.30	0.46	0.07	3.41	3.69	0.68	0.18	1.09	1.41	0.38	0.24
3	4.09	0.52	3.08	0.41	0.51	0.07	3.42	3.68	0.68	0.17	1.08	1.41	0.39	0.24
4	3.96	0.64	2.80	0.28	0.45	0.07	3.38	3.66	0.68	0.16	1.10	1.47	0.40	0.24
4.5	3.93	0.66	2.70	0.26	0.42	0.08	3.32	3.63	0.66	0.16	1.17	1.50	0.41	0.24
5	3.79	0.77	1.91	0.15	0.44	0.11	3.15	3.57	0.63	0.15	1.08	1.49	0.41	0.24
5.5	3.85	0.68	2.44	0.22	0.41	0.10	3.19	3.56	0.64	0.11	1.14	1.54	0.42	0.24
6	3.80	0.78	2.20	0.12	0.49	0.10	3.27	3.61	0.65	0.12	1.08	1.46	0.40	0.24
6.5	3.78	0.71	2.26	0.19	0.37	0.10	3.15	3.53	0.63	0.10	1.18	1.54	0.42	0.24
7	3.75	0.72	2.16	0.17	0.35	0.11	3.11	3.52	0.62	0.10	1.19	1.53	0.42	0.24
7.5	3.75	0.81	2.14	0.10	0.50	0.09	3.29	3.60	0.66	0.12	1.11	1.47	0.40	0.24
S44pAcF R119C														
0	4.29	0.96	1.23	0.04			4.16	4.25	0.83	0.13	1.16	1.20	0.33	0.26
2	4.18	0.96	1.33	0.04			4.06	4.14	0.81	0.13	1.21	1.42	0.39	0.23
3	4.19	0.88	3.08	0.10	0.76	0.02	4.00	4.09	0.80	0.12	1.26	1.42	0.39	0.23
4	4.14	0.87	3.03	0.10	0.57	0.03	3.92	4.04	0.78	0.11	1.13	1.44	0.39	0.23
4.5	4.11	0.88	2.83	0.09	0.58	0.03	3.90	4.02	0.78	0.10	1.10	1.45	0.40	0.23
5	4.01	0.95	1.28	0.05			3.87	3.65	0.77	0.10	1.14	1.45	0.40	0.23
5.5	3.99	0.94	1.78	0.06			3.86	3.93	0.77	0.09	1.16	1.52	0.41	0.22
6	3.99	0.93	2.21	0.04	0.31	0.03	3.81	3.94	0.76	0.09	1.03	1.49	0.41	0.22
6.5	3.94	0.94	1.52	0.06			3.79	3.87	0.76	0.09	1.28	1.53	0.42	0.23
7	3.88	0.93	1.48	0.07			3.71	3.81	0.74	0.08	1.26	1.54	0.42	0.23
7.5	3.90	0.97	0.56	0.03			3.78	3.89	0.76	0.09	1.10	1.50	0.41	0.23
S44pAcF I150C														
0	4.29	0.96	1.23	0.04			4.16	4.25	0.83	0.13	1.16	1.31	0.36	0.25
2	4.18	0.96	1.33	0.04			4.06	4.14	0.81	0.13	1.21	1.32	0.36	0.25
3	4.19	0.88	3.08	0.10	0.76	0.02	4.00	4.09	0.80	0.12	1.26	1.21	0.33	
4	4.14	0.87	3.03	0.10	0.57	0.03	3.92	4.04	0.78	0.11	1.13	1.33	0.36	0.24
4.5	4.11	0.88	2.83	0.09	0.58	0.03	3.90	4.02	0.78	0.10	1.10	1.31	0.36	0.23
5	4.01	0.95	1.28	0.05			3.87	3.65	0.77	0.10	1.14	1.24	0.34	0.24
5.5	3.99	0.94	1.78	0.06			3.86	3.93	0.77	0.09	1.16	1.26	0.34	
6	3.99	0.93	2.21	0.04	0.31	0.03	3.81	3.94	0.76	0.09	1.03	1.43	0.39	0.24
6.5	3.94	0.94	1.52	0.06			3.79	3.87	0.76	0.09	1.28	1.28	0.35	0.20
7	3.88	0.93	1.48	0.07			3.71	3.81	0.74	0.08	1.26			
7.5	3.90	0.97	0.56	0.03			3.78	3.89	0.76	0.09	1.10	1.38	0.38	0.23
N55pAcF R119C														
0	4.23	0.87	1.85	0.13			3.92	4.08	0.78	0.14	1.31	1.34	0.37	0.25
2	4.11	0.89	1.88	0.11			3.87	3.99	0.77	0.13	1.34	1.34	0.37	0.25
3	3.92	0.86	2.38	0.11			3.83	3.97	0.77	0.12	1.31	1.35	0.37	0.24

UNFOLDING PATHWAY OF T4L

4	4.11	0.76	2.87	0.19	0.61	0.05	3.71	3.90	0.74	0.11	1.08	1.34	0.37	0.24
4.5	3.99	0.86	2.04	0.11	0.31	0.03	3.65	3.86	0.73	0.11	1.17	1.37	0.37	0.23
5	4.01	0.81	2.53	0.15	0.53	0.04	3.65	3.84	0.73	0.09	1.12	1.36	0.37	0.23
5.5	4.02	0.76	2.92	0.18	0.73	0.05	3.64	3.82	0.73	0.07	1.06	1.36	0.37	0.23
6	3.99	0.73	3.01	0.22	0.76	0.05	3.62	3.78	0.72	0.07	1.11	1.36	0.37	0.23
6.5	4.23	0.38	3.48	0.57	0.80	0.05	3.64	3.79	0.73	0.08	1.16	1.37	0.38	0.23
7	3.92	0.86	2.38	0.11	0.48	0.03	3.65	3.80	0.73	0.07	1.25	1.36	0.37	0.24
7.5	3.87	0.88	2.19	0.09	0.38	0.03	3.60	3.76	0.72	0.08	1.25	1.42	0.39	0.24
K60pAcF R119C														
0	4.26	0.91	1.76	0.09			4.03	4.16	0.81	0.18	1.18	1.34	0.37	0.25
2	4.13	0.92	1.57	0.08			3.94	4.06	0.79	0.17	1.19	1.36	0.37	0.23
3	4.09	0.92	1.66	0.08			3.90	4.01	0.78	0.17	1.22	1.37	0.37	0.22
4	4.03	0.93	1.49	0.07			3.85	3.96	0.77	0.16	1.16	1.36	0.37	0.23
4.5	4.00	0.93	1.57	0.07			3.81	3.92	0.76	0.13	1.15	1.39	0.38	0.23
5	3.97	0.93	1.27	0.07			3.79	3.91	0.76	0.10	1.19	1.39	0.38	0.23
5.5	3.95	0.94	1.33	0.06			3.80	3.90	0.76	0.10	1.14	1.38	0.38	0.19
6	3.94	0.95	1.32	0.05			3.81	3.89	0.76	0.10	1.10	1.39	0.38	0.23
6.5	3.92	0.95	1.24	0.05			3.79	3.88	0.76	0.10	1.18	1.39	0.38	0.23
7	3.91	0.95	1.21	0.05			3.78	3.87	0.76	0.10	1.15	1.39	0.38	0.23
7.5	3.89	0.95	1.06	0.05			3.76	3.85	0.75	0.10	1.11	1.40	0.38	0.23
K60pAcF N132C														
0	4.35	0.76	2.87	0.19	0.37	0.05	3.86	4.12	0.82	0.18	1.03	1.5	0.41	0.29
0.5	4.21	0.83	2.17	0.11	0.24	0.05	3.77	4.06	0.81	0.17	0.96	1.49	0.41	0.29
1	4.18	0.83	2.38	0.12	0.31	0.05	3.79	4.03	0.81	0.17	1.05	1.46	0.4	0.282
1.5	4.15	0.84	2.31	0.11	0.27	0.06	3.73	4.01	0.80	0.17	0.98	1.47	0.4	0.285
2	4.15	0.82	2.53	0.14	0.36	0.05	3.75	3.99	0.80	0.17	1.08	1.44	0.39	0.274
2.5	4.11	0.83	2.36	0.11	0.32	0.05	3.71	3.96	0.79	0.17	1.02	1.46	0.4	0.281
3	4.08	0.84	2.21	0.11	0.31	0.05	3.68	3.94	0.79	0.17	1.04	1.44	0.39	0.269
3.5	4.04	0.85	2.17	0.10	0.27	0.06	3.65	3.92	0.78	0.16	1.06	1.46	0.4	0.282
4	4.03	0.84	2.15	0.10	0.29	0.06	3.61	3.90	0.78	0.16	1.00	1.43	0.39	0.268
4.5	3.98	0.84	1.83	0.09	0.24	0.08	3.52	3.87	0.77	0.13	1.03	1.46	0.4	0.269
5	3.96	0.85	1.74	0.07	0.24	0.07	3.52	3.86	0.77	0.11	0.99	1.43	0.39	0.248
5.5	3.94	0.85	1.61	0.07	0.19	0.08	3.49	3.85	0.77	0.10	1.02	1.45	0.4	0.252
6	3.95	0.87	1.69	0.06	0.21	0.07	3.56	3.87	0.77	0.10	1.02	1.42	0.39	0.245
6.5	3.93	0.88	1.87	0.07	0.26	0.05	3.59	3.81	0.76	0.10	1.01	1.42	0.39	0.195
7	3.91	0.88	1.80	0.06	0.21	0.06	3.58	3.84	0.77	0.09	1.01	1.42	0.39	0.261
7.5	3.89	0.89	1.66	0.05	0.22	0.05	3.58	3.82	0.76	0.09	1.03	1.44	0.39	0.241
Q69pAcF N132C														
0	4.04	0.89	1.41	0.11			3.76	3.93	0.75	0.18	1.10	1.33	0.36	0.29
2	4.21	0.79	2.59	0.15	0.42	0.06	3.73	4.02	0.75	0.13	1.08	1.42	0.39	0.28
3	4.16	0.80	2.55	0.14	0.39	0.06	3.71	3.98	0.74	0.12	1.04	1.36	0.37	0.28
4	4.14	0.75	2.68	0.17	0.37	0.08	3.60	3.92	0.72	0.12	1.12	1.35	0.37	0.27
4.5	4.10	0.77	2.58	0.16	0.39	0.07	3.60	3.90	0.72	0.12	1.13	1.34	0.37	0.26
5	4.02	0.86	2.23	0.09	0.41	0.06	3.66	3.91	0.73	0.12	1.10	1.43	0.39	0.25
5.5	4.03	0.76	2.52	0.16	0.36	0.08	3.50	3.84	0.70	0.10	1.11	1.30	0.35	0.25
6	3.96	0.88	1.84	0.07	0.37	0.05	3.64	3.87	0.73	0.11	1.22	1.41	0.39	0.25
6.5	3.98	0.77	2.35	0.14	0.28	0.09	3.41	3.79	0.68	0.10	1.10	1.31	0.36	0.26
7	4.02	0.73	2.73	0.20	0.35	0.07	3.49	3.80	0.70	0.10	1.09	1.33	0.36	0.25
7.5	3.92	0.89	2.14	0.07	0.43	0.05	3.64	3.83	0.73	0.11	1.09	1.45	0.40	0.25

Double-labeled samples were analyzed using a simple multi exponential model, a Gaussian distribution and the from polymer science derived model Worm-like chain model. The results are summarized in Table 4.3.2B. Figure 4.3.5A shows that the fluorescence-weighted averaged lifetime for all variants rises with increasing urea concentration in a two-state

manner, indicative for a growing distance between the two labeling sites, thus an unfolding of the protein. The extent of increase, however, is dependent upon the variant. The mean distance between the two labeling positions has been determined with the Gaussian distribution model (Figure 4.3.5B-C) which shows again a typical two-state behavior; the only exception might be the variant N55pAcF R119C displaying only a gradual increase, but here the fluorescence- weighted averaged lifetime increases only by 0.2 ns. The fit results of the Worm-like chain model draw a more diverse picture of our variants (Figure 4.3.5D-F). They can be grouped in two classes: (I) decreasing stiffness (K19pAcF P86C, S44pAcF R119C, S44pAcF I150C, N55pAcF R119C and K60pAcF R119C) and (II) fluctuating-slightly increasing stiffness (R8pAcF P86C, K60pAcF N132C and Q69pAcF N132C). For class I, both labeling positions are “inside” of the protein, whereas for class II, one labeling positions is rather “outside” of the protein’s tertiary structure.

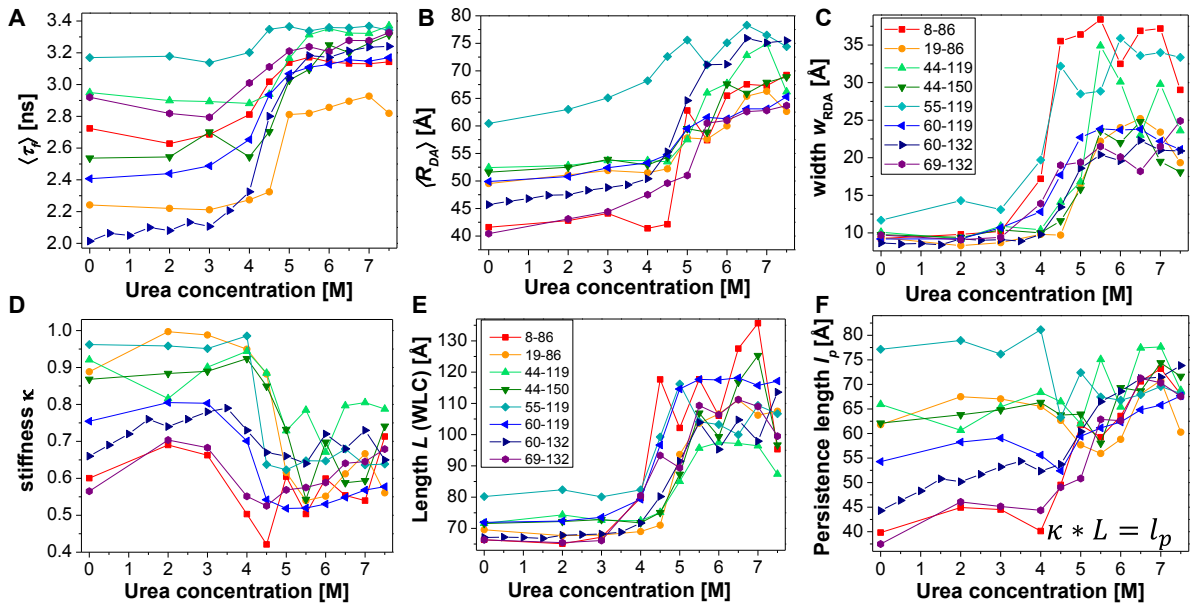


Figure 4.3.5 Summary of eTCSPC results from double-labeled samples. (A) Fluorescence-weighted average lifetime obtained from fitting with the multiexponential model. (B-C) Mean dye distance $\langle R_{DA} \rangle$ and width w_{DA} obtained from fitting with Gaussian distribution model. (D-F) Stiffness κ , length L and persistence length l_p obtained from fitting with the Worm-like chain model. Note: The color code is the same for all graphs, but the legend was only put into two plots, where space was available.

Finally, for the persistence length – product of stiffness and length – even three classes of variants can be distinguished: (A) decreasing l_p (N55pAcF R119C), (B) slightly increasing l_p (K19pAcF P86C, S44pAcF R119C, S44pAcF I150C and K60pAcF R119C) and (C) strongly increasing l_p (R8pAcF P86C, K60pAcF N132C and Q69pAcF N132C).

Table 4.3.2A Fit results of double labeled samples with the multi exponential model.

[M urea]	τ_1 [ns]	x_1	τ_2 [ns]	x_2	τ_3 [ns]	x_3	τ_4 [ns]	x_4	τ_f [ns]	χ^2_r
R8pAcF P86C										
0	4.05	0.22	1.62	0.39	0.53	0.39			2.73	1.06
2	3.95	0.21	1.55	0.44	0.44	0.35			2.63	1.09
3	3.96	0.22	1.65	0.44	0.46	0.34			2.69	1.15
4	3.77	0.29	1.71	0.38	0.43	0.32			2.81	1.21
4.5	3.65	0.43	1.76	0.32	0.41	0.25			3.02	1.13
5	3.56	0.53	1.57	0.25	0.31	0.22			3.14	1.17
5.5	3.7	0.49	2.15	0.32	0.51	0.19			3.18	1.09
6	3.55	0.55	1.68	0.25	0.32	0.2			3.14	1.26
6.5	3.6	0.53	1.96	0.29	0.46	0.18			3.13	1.09
7	3.56	0.55	1.88	0.27	0.44	0.18			3.13	1.03
7.5	3.51	0.58	1.65	0.24	0.3	0.18			3.14	1.21
K19pAcF P86C										
0	3.05	0.28	1.59	0.52	0.46	0.2			2.24	1.07
2	3.09	0.25	1.65	0.54	0.48	0.21			2.22	1.07
3	3.07	0.26	1.63	0.54	0.44	0.2			2.21	1.10
4	3.13	0.27	1.64	0.53	0.45	0.2			2.27	1.08
4.5	3.13	0.3	1.61	0.5	0.43	0.2			2.32	1.09
5	3.47	0.4	1.9	0.36	0.52	0.24			2.81	1.09
5.5	3.52	0.37	1.96	0.38	0.47	0.25			2.82	1.09
6	3.4	0.46	1.8	0.31	0.47	0.23			2.86	1.12
6.5	3.43	0.46	1.78	0.32	0.37	0.22			2.9	1.21
7	3.45	0.47	1.85	0.32	0.39	0.21			2.93	1.2
7.5	3.33	0.48	1.77	0.3	0.47	0.22			2.82	1.06
S44pAcF R119C										
0	4.03	0.34	2.13	0.47	0.6	0.19			2.13	1.09
2	3.93	0.28	2.14	0.52	0.68	0.2			2.9	1.05
3	3.94	0.27	2.2	0.52	0.69	0.21			2.89	1.02
4	3.84	0.3	2.15	0.51	0.64	0.19			2.88	1.07
4.5	3.85	0.31	2.22	0.49	0.68	0.2			2.94	1.03
5	3.7	0.51	1.87	0.33	0.45	0.16			3.17	1.14
5.5	3.67	0.55	1.74	0.2	0.23	0.25			3.31	1.26
6	3.73	0.61	2.04	0.24	0.56	0.15			3.35	1.09
6.5	3.68	0.55	1.8	0.2	0.23	0.25			3.32	1.16
7	3.65	0.56	1.75	0.19	0.22	0.25			3.32	1.18
7.5	3.7	0.66	2	0.23	0.54	0.12			3.37	1.13
S44pAcF I150C										
0	3.94	0.14	2.2	0.59	0.71	0.27			2.54	1.14
2	3.85	0.15	2.21	0.6	0.71	0.25			2.54	1.09
3	3.89	0.2	2.19	0.58	0.69	0.22			2.7	1.06
4	3.8	0.16	2.19	0.61	0.68	0.23			2.54	1.06
4.5	3.92	0.19	2.28	0.58	0.7	0.23			2.7	1.07
5	3.57	0.48	1.95	0.36	0.5	0.16			3.03	1.18
5.5	3.71	0.45	2.12	0.36	0.55	0.19			3.1	1.07
6	3.63	0.61	1.98	0.26	0.49	0.13			3.25	1.09
6.5	3.69	0.52	2.21	0.32	0.57	0.16			3.2	1.02
7	3.71	0.55	2.2	0.31	0.54	0.14			3.26	1.03
7.5	3.61	0.67	1.84	0.22	0.38	0.11			3.31	1.05
N55pAcF R119C										
0	3.94	0.39	2.32	0.49	0.72	0.12			3.17	1.02
2	3.91	0.39	2.43	0.5	0.75	0.11			3.18	1.15
3	3.64	0.51	2.07	0.37	0.35	0.12			3.14	1.16
4	3.67	0.55	2.14	0.34	0.61	0.11			3.2	1.09
4.5	3.71	0.63	1.93	0.24	0.54	0.13			3.35	1.06
5	3.7	0.65	1.98	0.22	0.56	0.13			3.37	1.15

UNFOLDING PATHWAY OF T4L

5.5	3.68	0.65	1.91	0.22	0.57	0.13			3.34	1.11
6	3.68	0.66	1.94	0.22	0.51	0.12			3.36	1.11
6.5	3.66	0.68	1.92	0.21	0.55	0.11			3.36	1.12
7	3.67	0.68	1.97	0.2	0.6	0.12			3.39	1.16
7.5	3.64	0.68	2.01	0.21	0.62	0.11			3.34	1.1
K60pAcF R119C										
0	3.68	0.17	1.92	0.56	0.6	0.27			2.41	1.12
2	3.63	0.18	1.98	0.57	61	0.24			2.44	1.05
3	3.59	0.21	1.99	0.56	0.6	0.23			2.49	1.08
4	3.56	0.29	1.96	0.49	0.56	0.22			2.65	1.04
4.5	3.6	0.42	1.94	0.38	0.52	0.2			2.94	1.1
5	3.59	0.5	1.92	0.32	0.5	0.18			3.07	1.1
5.5	3.59	0.53	1.94	0.3	0.52	0.17			3.11	1.06
6	3.57	0.55	1.91	0.29	0.5	0.16			3.13	1.15
6.5	3.6	0.56	2.03	0.29	0.55	0.16			3.16	1.13
7	3.57	0.57	2	0.28	0.56	0.15			3.15	1.16
7.5	3.57	0.59	2	0.27	0.54	0.14			3.17	1.1
K60pAcF N132C										
0	3.71	0.09	1.72	0.4	0.18	0.19	0.65	0.32	2.01	0.99
0.5	3.93	0.07	1.9	0.38	0.83	0.32	0.25	0.23	2.06	0.98
1	3.84	0.08	1.87	0.4	0.8	0.3	0.24	0.22	2.05	0.98
1.5	3.85	0.08	1.91	0.4	0.81	0.3	0.23	0.21	2.1	0.99
2	3.79	0.81	1.93	0.41	0.84	0.29	0.24	0.22	2.08	1.00
2.5	3.89	0.08	2	0.42	0.84	0.31	0.23	0.2	2.13	0.98
3	3.65	0.1	1.9	0.44	0.79	0.28	0.22	0.19	2.11	1.01
3.5	3.7	0.11	1.94	0.45	0.81	0.26	0.24	0.18	2.21	1.02
4	3.68	0.14	2.01	0.42	0.84	0.26	0.23	0.18	2.32	0.96
4.5	3.7	0.28	2.15	0.34	0.87	0.22	0.21	0.16	2.8	0.99
5	3.59	0.45	1.97	0.28	0.68	0.16	0.17	0.11	3.04	1.00
5.5	3.6	0.57	1.81	0.27	0.41	0.16			3.18	1.06
6	3.56	0.59	1.74	0.27	0.39	0.15			3.17	1.06
6.5	3.59	0.6	1.83	0.26	0.41	0.14			3.21	1.03
7	3.61	0.61	1.94	0.26	0.47	0.13			3.24	1.06
7.5	3.58	0.64	1.83	0.24	0.41	0.12			3.24	1.07
Q69pAcF N132C										
0	4.13	0.24	1.51	0.36	0.47	0.4			2.92	1.08
2	4.01	0.25	1.55	0.43	0.44	0.32			2.82	1.13
3	4.01	0.24	1.61	0.43	0.46	0.33			2.79	1.13
4	3.85	0.36	1.81	0.37	0.5	0.27			3.01	1.04
4.5	3.77	0.43	1.83	0.34	0.47	0.23			3.11	1.1
5	3.78	0.49	1.9	0.33	0.52	0.19			3.21	1.13
5.5	3.77	0.5	2.07	0.32	0.51	0.18			3.24	1.05
6	3.77	0.5	1.87	0.31	0.51	0.19			3.21	1.12
6.5	3.75	0.54	2.14	0.3	0.53	0.16			3.28	1.05
7	3.74	0.55	2.15	0.3	0.54	0.15			3.28	1.04
7.5	3.71	0.61	2.11	0.25	0.57	0.14			3.33	1.07

Table 4.3.2B Fitting results of double labeled samples for the Gaussian distribution and Worm-like chain model. Uncertainties given are standard errors of the mean.

[M urea]	$\langle R_{DA} \rangle$ [Å]	wDA [Å]	x_{DOnly}	χ^2_r	L [Å]	κ	x_{DOnly}	lp	χ^2_r
R8pAcF P86C									
0	41.6±0.3	9.3±0.2	0.18±0.004	1.11	66.3±0.5	0.60±0.01	0.13±0.003	39.8	1.14
2	42.8±0.1	9.8±0.2	0.18±0.002	2.08	65.1±0.5	0.69±0.02	0.15±0.005	44.9	1.03
3	44.1±0.2	10.2±0.2	0.19±0.003	1.12	67.2±0.5	0.66±0.02	0.16±0.01	44.5	1.04
4	41.4±0.5	17.2±0.7	0.18±0.006	1.10	79.8±1.0	0.50±0.01	0.16±0.004	40.1	1.09
4.5	42.1±2.1	35.5±4.3	0	1.05	117.7±9.9	0.42±0.03	0.11±0.02	49.5	1.06

UNFOLDING PATHWAY OF T4L

5	62.8±2.0	36.4±4.9	0	1.15	102.2±8.8	0.60±0.03	0.22±0.10	61.7	1.18
5.5	57.4±0.9	38.4±4.7	0	1.14	117.7±5.5	0.50±0.02	0.11±0.02	59.2	1.05
6	65.5±2.0	32.5±4.6	0	1.15	106.1±5.5	0.60±0.02	0.18±0.04	63.6	1.13
6.5	67.6±0.6	36.9±1.2	0	1.10	127.5±10.2	0.55±0.02	0.07±0.03	70.6	1.03
7	67.3±0.7	37.2±1.4	0	1.13	135.7±14.5	0.54±0.03	0.01±0.01	73.2	1.04
7.5	69.2±2.1	29.0±2.8	0.2±0.16	1.14	95.3±6.0	0.71±0.03	0.32±0.12	68.0	1.13
K19pAcF P86C									
0	49.5±0.05	9.4±0.2	0.05±0.003	1.40	69.6±0.4	0.89±0.01	0.06±0.003	61.8	1.19
2	51.1±0.05	8.3±0.1	0.06±0.003	1.24	67.7±0.3	1.00±0.02	0.07±0.002	67.5	1.05
3	51.9±0.05	8.7±0.2	0.05±0.003	1.26	67.9±0.3	0.99±0.02	0.07±0.002	67.0	1.08
4	51.5±0.05	9.8±0.2	0.07±0.004	1.24	69.0±0.4	0.95±0.02	0.08±0.003	65.5	1.01
4.5	52.2±0.05	9.7±0.2	0.08±0.004	1.26	71.0±0.4	0.88±0.02	0.09±0.003	62.7	1.06
5	57.8±0.5	16.5±1.1	0.03±0.01	1.25	93.7±2.1	0.62±0.02	0.13±0.01	57.7	1.09
5.5	57.6±0.6	22.2±1.7	0.07	1.10	103.7±2.7	0.54±0.02	0.08±0.01	55.9	1.01
6	60.0±0.8	24.0±2.7	0	1.20	106.6±4.6	0.55±0.02	0.04±0.01	58.8	1.08
6.5	65.4±0.7	25.2±1.3	0	1.15	111.2±4.2	0.61±0.02	0.03±0.01	68.0	1.06
7	66.3±1.9	23.4±2.4	0.04±0.03	1.20	106.2±3.4	0.67±0.02	0.08±0.01	70.8	1.11
7.5	62.6±0.3	19.3±0.4	0.1±0.01	1.20	107.5±5.0	0.57±0.02	0.02±0.01	60.3	1.04
S44pAcF R119C									
0	52.4±0.1	10.1±0.3	0.15±0.005	1.28	71.6±0.6	0.92±0.02	0.17±0.005	65.9	1.22
2	52.8±0.1	9.2±0.2	0.17±0.01	1.19	74.3±0.6	0.82±0.02	0.14±0.004	60.7	1.12
3	53.7±0.1	10.9±0.3	0.16±0.01	1.24	72.6±0.6	0.90±0.02	0.17±0.01	65.4	1.15
4	53.7±0.1	10.4±0.3	0.15±0.01	1.52	72.5±0.6	0.94±0.02	0.18±0.01	68.4	1.15
4.5	53.5±0.2	14.1±0.5	0.11±0.01	1.35	75.1±0.7	0.88±0.02	0.18±0.01	66.5	1.18
5	57.5±0.4	16.8±1.0	0.19±0.02	1.12	85.1±1.6	0.73±0.02	0.23±0.01	62.0	1.11
5.5	66.0±0.9	34.9±1.3	0	1.74	95.7±9.0	0.78±0.04	0.24±0.18	75.1	1.31
6	67.7±2.3	30.1±4.2	0.02±0.01	1.22	97.5±5.3	0.67±0.03	0.26±0.06	65.4	1.21
6.5	72.8±1.0	23.0±0.8	0.16±0.08	2.87	97.2±12.6	0.80±0.04	0.22±0.03	77.4	1.22
7	74.9±1.3	29.8±1.2	0	1.55	96.4±1.2	0.81±0.02	0.31±0.05	77.6	1.26
7.5	66.2±2.1	23.6±3.0	0.18±0.06	1.11	87.4±4.0	0.79±0.03	0.39±0.05	68.8	1.13
S44pAcF I150C									
0	51.6±0.05	9.8±0.2	0.03±0.003	1.16	71.6±0.4	0.87±0.01	0.05±0.002	62.1	1.09
2	52.5±0.05	9.4±0.2	0.04±0.01	1.13	72.2±0.4	0.88±0.01	0.05±0.002	63.8	1.05
3	53.9±0.1	10.4±0.2	0.08±0.006	1.1	72.8±0.4	0.89±0.01	0.10±0.004	64.8	1.04
4	52.9±0.05	10.0±0.2	0.04±0.005	1.18	71.8±0.4	0.92±0.01	0.06±0.003	66.3	1.05
4.5	54.3±0.1	11.6±0.3	0.06±0.01	1.11	75.1±0.5	0.85±0.01	0.09±0.004	63.8	1.02
5	59.5±0.4	15.8±0.7	0.07±0.01	1.13	87.3±1.3	0.73±0.02	0.12±0.01	63.9	1.12
5.5	58.9±0.4	23.6±1.1	0.2±0.07	1.03	106.9±3.4	0.54±0.02	0.05±0.01	58.0	1.03
6	67.6±0.4	22.0±0.5	0	1.13	99.3±4.0	0.70±0.02	0.13±0.05	69.3	1.12
6.5	65.9±0.2	24.8±0.5	0	1.02	116.8±5.6	0.59±0.02	0	68.7	1.03
7	67.9±0.2	19.5±0.5	0.15±0.06	1.01	125.3±10.8	0.59±0.02	0	74.4	0.96
7.5	68.9±0.1	18.1±0.4	0.10±0.01	1.11	96.7±4.4	0.74±0.03	0.23±0.08	71.6	1.22
N55pAcF R119C									
0	60.4±0.4	11.7±0.5	0.14±0.01	1.16	80.2±0.9	0.96±0.03	0.18±0.01	77.1	1.11
2	63.0±0.5	14.3±0.7	0.08±0.01	1.28	83.4±1.0	0.96±0.03	0.18±0.01	78.9	1.21
3	65.1±0.6	13.1±0.6	0.08±0.02	1.26	80.1±1.0	0.95±0.03	0.31±0.02	76.2	1.17
4	68.2±0.9	19.7±1.1	0	1.38	82.3±1.5	0.99±0.03	0.26±0.02	81.1	1.19
4.5	72.6±4.1	32.2±5.2	0	1.22	99.3±6.6	0.64±0.02	0.33±0.10	63.3	1.19
5	75.6±2.0	28.5±2.1	0	1.29	116.2±10.3	0.62±0.02	0.19±0.11	72.4	1.26
5.5	71.1±1.0	28.9±1.6	0.02±0.02	1.24	104.1±6.7	0.65±0.03	0.27±0.06	67.4	1.24
6	75.1±1.7	35.9±1.7	0.05±0.02	1.24	103.2±8.0	0.65±0.03	0.35±0.07	66.8	1.23
6.5	78.3±4.8	33.6±6.2	0	1.26	100±6.7	0.68±0.03	0.36±0.06	67.9	1.26
7	76.5±5.5	34.0±6.9	0.04±0.04	1.22	109.4±8.3	0.64±0.03	0.28±0.06	69.6	1.24
7.5	74.4±4.2	33.4±6.3	0.37±0.18	1.35	106.7±7.1	0.64±0.03	0.3±0.05	68.1	1.26
K60pAcF R119C									
0	49.9±0.05	9.2±0.2	0.05±0.003	1.22	71.9±0.4	0.75±0.01	0.05±0.002	54.3	1.13
2	50.8±0.05	9.2±0.2	0.05±0.003	1.25	72.4±0.4	0.81±0.02	0.05±0.002	58.3	1.13

UNFOLDING PATHWAY OF T4L

3	52.4±0.05	10.7±0.2	0.04±0.003	1.55	73.6±0.4	0.8±0.02	0.06±0.003	59.1	1.12
4	53.3±0.1	12.9±0.4	0.05±0.004	1.39	79.3±0.7	0.7±0.02	0.07±0.004	55.6	1.08
4.5	54.8±0.4	17.7±0.3	0.07±0.02	1.19	96.7±1.8	0.54±0.02	0.06±0.01	52.4	1.17
5	59.5±0.5	22.7±0.4	0	1.19	114.6±3.2	0.52±0.02	0	59.4	1.17
5.5	61.5±0.3	23.8±0.4	0.08±0.05	1.13	117.7±4.1	0.52±0.02	0	61.1	1.12
6	61.3±0.3	23.7±0.4	0	1.19	117.5±4.2	0.53±0.03	0	62.3	1.13
6.5	63.1±2.4	23.8±3.2	0	1.16	118.2±3.9	0.55±0.02	0	64.8	1.16
7	63.1±0.2	22.2±0.4	0.01±0.002	1.23	115.7±4.1	0.57±0.02	0	65.8	1.19
7.5	65.3±0.4	21.1±0.6	0.05±0.01	1.23	117.2±4.2	0.58±0.02	0	67.7	1.22
K60pAcF N132C									
0	45.7±1.4	8.7±0.1	0.14±0.004	1.01	67.1±0.7	0.66±0.02	0.05±0.002	44.3	0.99
0.5	46.3±1.4	8.5±0.1	0.15±0.003	1.06	67.2±0.7	0.69±0.01	0.05±0.001	46.4	1.00
1	46.8±1.4	8.6±0.1	0.14±0.004	1.10	67.1±0.7	0.72±0.01	0.05±0.002	48.3	0.96
1.5	47.4±1.4	8.4±0.1	0.16±0.005	1.09	66.8±0.7	0.76±0.02	0.06±0.001	50.8	0.96
2	47.5±1.4	9.2±0.3	0.12±0.008	1.07	67.8±0.7	0.74±0.01	0.05±0.002	50.2	0.98
2.5	48.3±1.4	9.0±0.2	0.13±0.007	1.06	68.0±0.7	0.76±0.02	0.06±0.002	51.7	0.97
3	48.8±1.5	9.1±0.2	0.04±0.003	1.10	68.2±0.7	0.78±0.02	0.05±0.002	53.2	1.01
3.5	49.3±1.5	8.9±0.2	0.17±0.005	1.13	68.8±0.7	0.79±0.02	0.07±0.002	54.4	1.02
4	50.4±1.5	9.8±0.2	0.18±0.007	1.01	71.7±0.7	0.73±0.01	0.08±0.002	52.3	0.97
4.5	55.3±1.7	13.4±0.4	0.33±0.017	1.00	80.2±0.8	0.67±0.01	0.17±0.005	53.7	1.01
5	64.6±1.3	18.6±0.9	0.19±0.055	1.01	91.5±1.8	0.66±0.02	0.18±0.014	60.4	1.02
5.5	71.1±2.1	20.4±1.4	0.16±0.086	1.08	103.9±6.2	0.64±0.03	0.13±0.053	66.5	1.10
6	71.2±1.4	19.6±1.2	0.09±0.084	1.03	95.3±1.9	0.72±0.02	0.17±0.022	68.6	1.04
6.5	75.9±2.3	22.3±0.4	0	1.02	104.7±4.2	0.68±0.02	0.12±0.048	71.2	1.02
7	75.1±2.3	21.0±1.5	0.09±0.11	1.05	97.9±2.9	0.73±0.02	0.2±0.030	71.5	1.06
7.5	75.5±3.0	20.9±1.7	0.10±0.12	1.06	113.6±5.7	0.65±0.02	0.05±0.062	73.8	1.13
Q69pAcF N132C									
0	40.4±0.3	9.7±0.5	0.21±0.01	1.21	66.3±1.1	0.57±0.03	0.15±0.01	37.5	1.31
2	43.1±0.1	9.1±0.2	0.22±0.003	1.16	65.5±0.5	0.70±0.02	0.19±0.01	46.1	1.10
3	44.4±0.1	9.4±0.2	0.22±0.003	1.12	66.1±0.4	0.68±0.01	0.18±0.003	45.1	1.05
4	47.5±0.2	13.9±0.5	0.26±0.01	1.04	80.5±0.2	0.55±0.003	0.19±0.006	44.3	1.11
4.5	49.6±0.3	19.0±0.7	0.22±0.01	1.06	93.3±1.7	0.53±0.01	0.18±0.01	49.1	1.00
5	51.0±0.4	19.4±1.3	0.25±0.02	1.11	89.4±2.2	0.57±0.02	0.16±0.01	50.8	1.14
5.5	60.5±0.3	21.5±0.8	0.21±0.004	1.07	109.3±2.1	0.58±0.02	0.16±0.01	62.9	1.03
6	61.0±2.0	20.1±3.0	0.07±0.07	1.08	106.4±6.0	0.59±0.03	0.20±0.04	62.6	1.08
6.5	62.6±1.4	18.2±1.7	0.29±0.06	1.11	111.2±3.2	0.64±0.02	0.15±0.01	71.2	1.08
7	62.8±1.5	21.5±2.2	0.22±0.06	1.07	109.0±4.0	0.65±0.03	0.07±0.01	70.3	1.06
7.5	63.7±2.6	24.9±3.6	0.20±0.14	1.13	99.5±5.2	0.68±0.03	0.28±0.04	67.6	1.13

4.3.2.3. Steady-state anisotropy

The residual anisotropy determining in general the mobility of dyes gives an indication how flexible the protein is at and around the labeling position, thus gives an indication whether the protein (backbone) is still folded. For the positions used as acceptor, the residual anisotropy is constant independent of the urea concentration, only for the position N132C a slight decrease is observed (Figure 4.3.6). For the donor positions labeled with Alexa488-hydroxylamine, the residual anisotropy is decreasing, i.e. the dye is “freed”. Here, some positions show the typical two- state behavior whereas for other positions a more gradual shift is observed.

Taken these results together, the residual fluorescence anisotropy indicates that the N-terminal domain is more unfolded, flexible than the C-terminal subdomain, which has also been shown to be more resistant against unfolding (44).

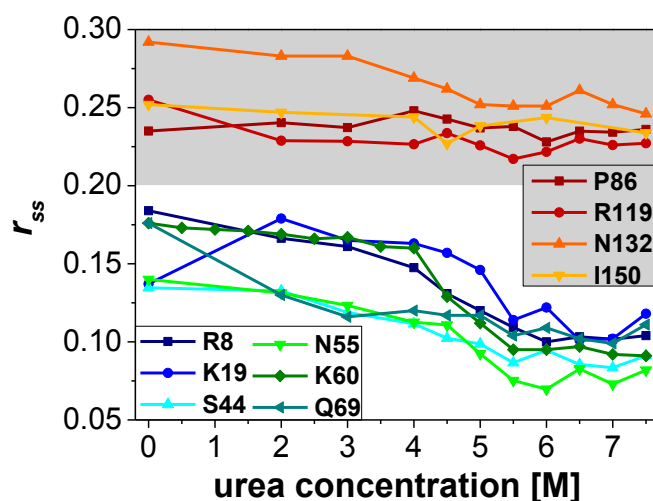
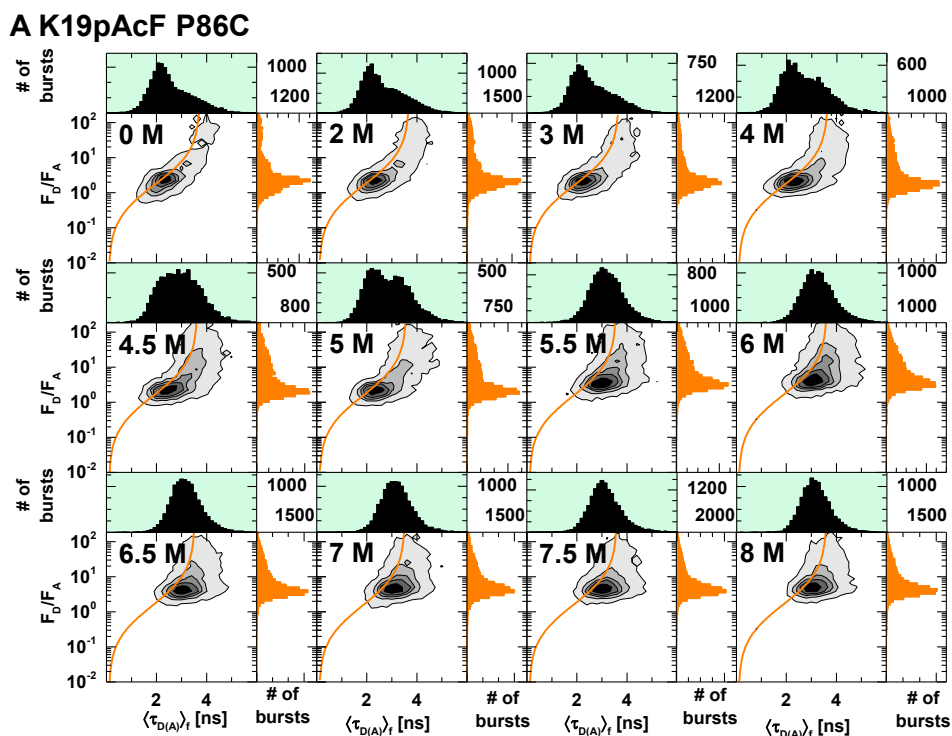


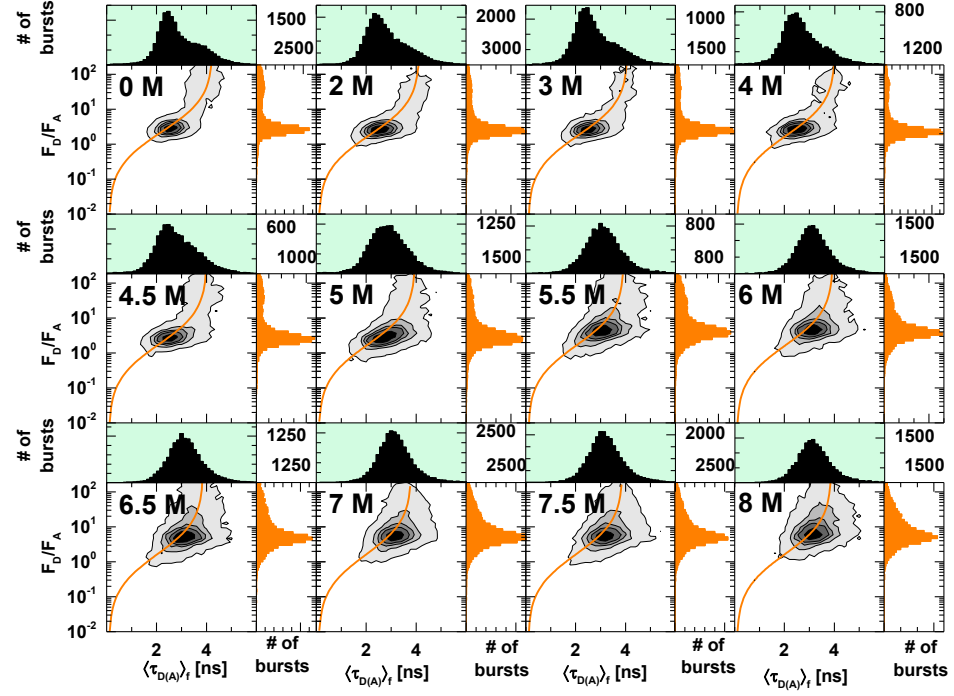
Figure 4.3.6 Residual Fluorescence Anisotropy of single labeled variants. All acceptor positions display a higher residual fluorescence anisotropy than the donor positions which is due to the different chemical structure and properties of the donor Alexa488 and the acceptor Alexa647.

4.3.2.4. Additional 2D MFD histograms

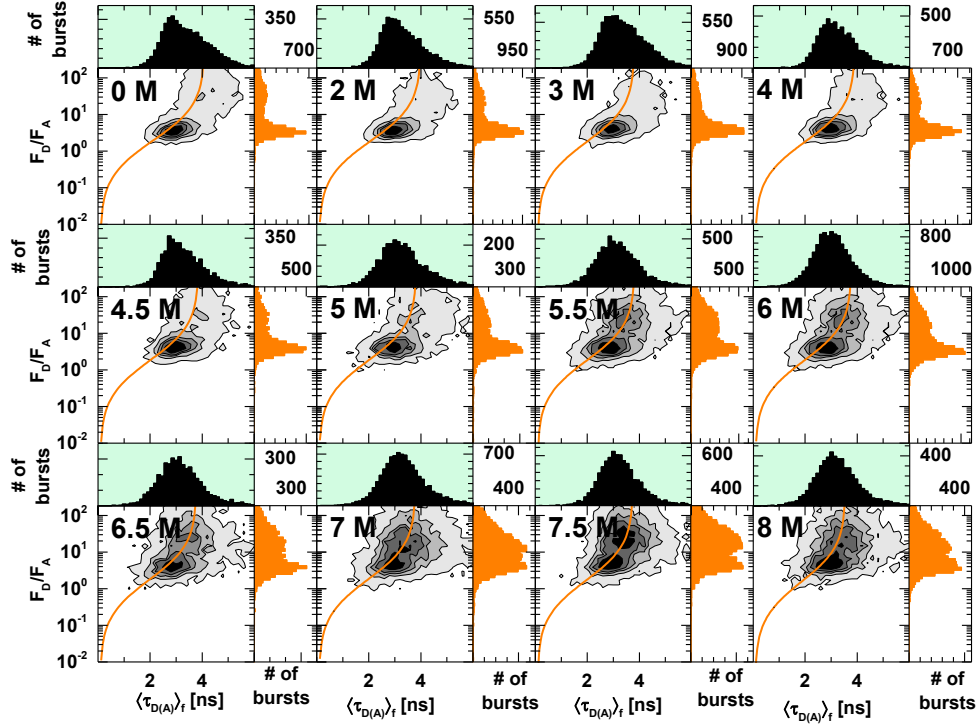
Figure 4.3.7 shows the unfolding of five variants in increasing urea concentrations monitored by single-molecule spectroscopy. The MFD plots of all variants show in general a similar behavior: the green to red signal ratio F_D/F_A shifts to higher value and the fluorescence lifetime of the donor in the presence of acceptor $\langle \tau_{D(A)} \rangle_f$ increases to values close to those of Donor-labeled molecules.



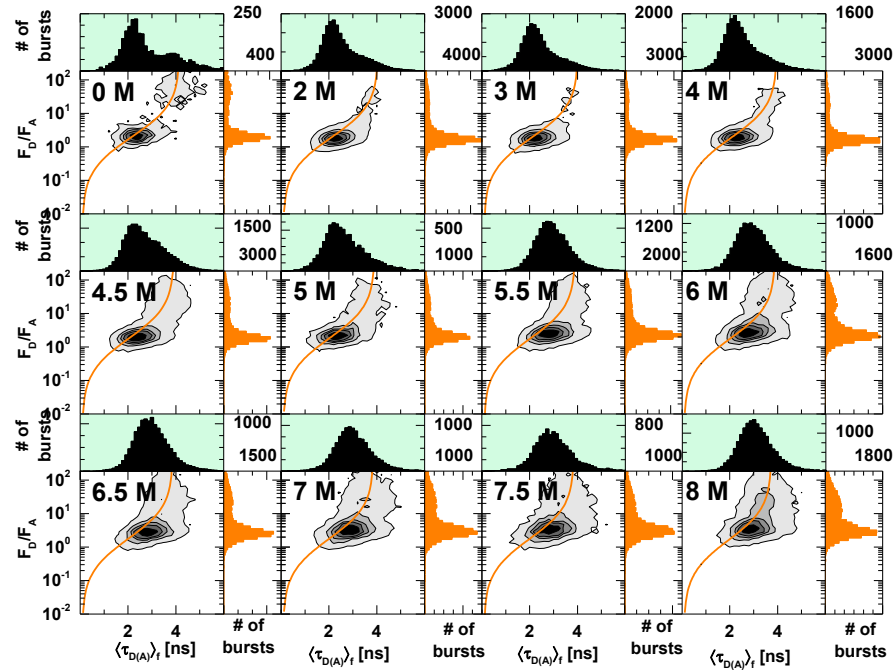
B S44pAcF R119C



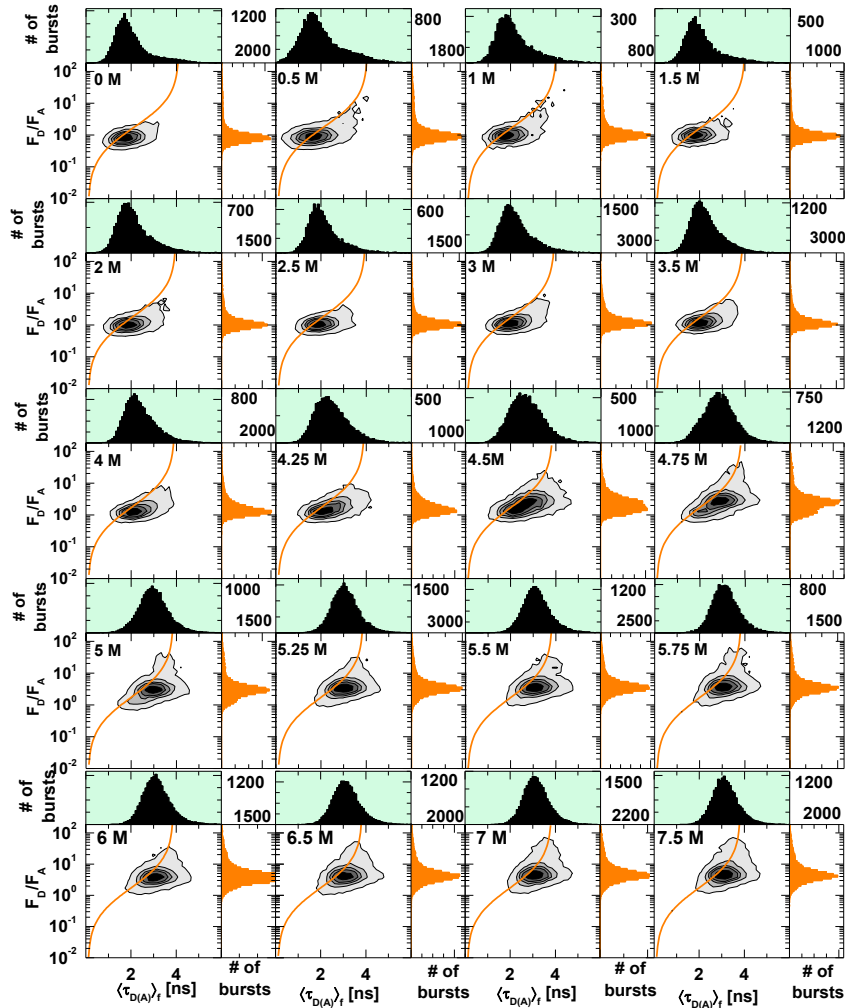
C N55pAcF R119C



D K60pAcF R119C



E K60pAcF N132C



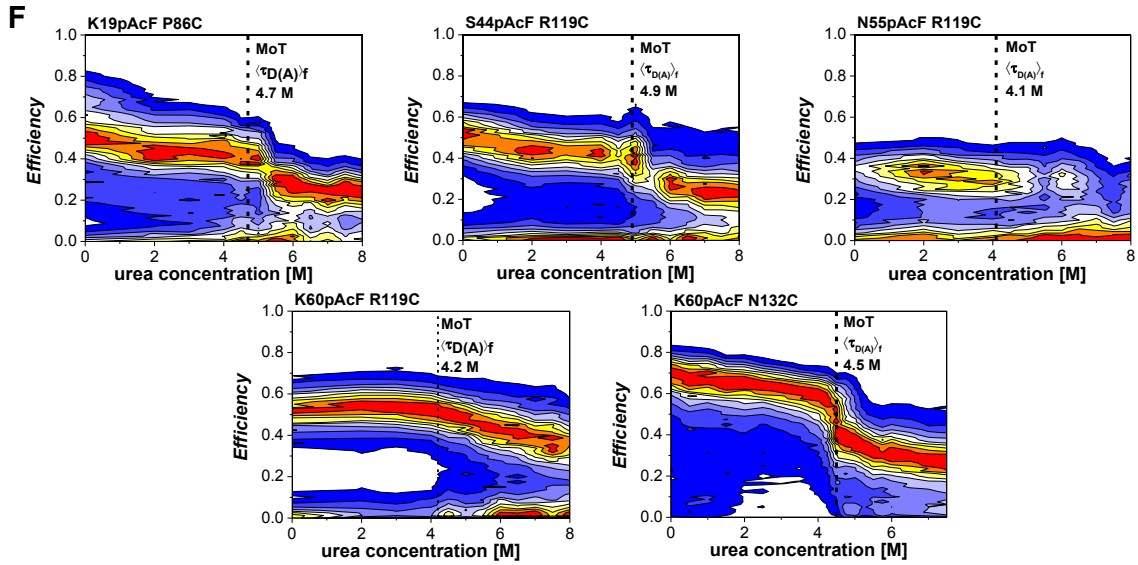


Figure 4.3.7 MFD Histograms of variants showing their unfolding. (A-E) Two dimensional histograms of donor to acceptor signal ratio F_D/F_A vs. lifetime of donor in the presence of acceptor $\langle \tau_{D(A)} \rangle_f$ of indicated variants in increasing urea concentrations. One dimensional projections for F_D/F_A and $\langle \tau_{D(A)} \rangle_f$ are also shown. Static FRET lines are shown in orange. Correction factors for background in green and red, detection efficiency ratio and parameters used to generate the static FRET lines are summarized in Table 4.3.3. (F) 2D projection for FRET *Efficiency* in increasing concentrations of urea (blue \rightarrow red).

Table 4.3.3 Correction factors and static FRET lines for MFD experiments. g_g is detection efficiency in green, by definition detection efficiency in red = 1. α is the crosstalk, i.e. the amount of green photons collected in the red channels. $\langle B_g \rangle$ and $\langle B_r \rangle$ are background count rates for buffer alone.

#	5				K19pAcPhe P86C			
0 M	$(0.7958/0.3400)/((4.0756/((-0.0377*x^3)+(0.2728*x^2)+0.5253*x+-0.0479))-1)$							
	g_g	0.75	α	1.7 %	$\langle B_g \rangle$	1.21 kHz	$\langle B_r \rangle$	0.95 kHz
2 M	$(0.6828/0.3800)/((3.6932/((-0.0489*x^3)+(0.2986*x^2)+0.5773*x+-0.0518))-1)$							
	g_g	0.70	α	2.2 %	$\langle B_g \rangle$	1.17 kHz	$\langle B_r \rangle$	0.74 kHz
3 M	$(0.6844/0.3900)/((3.6828/((-0.0491*x^3)+(0.3011*x^2)+0.5705*x+-0.0510))-1)$							
	g_g	0.70	α	2.2 %	$\langle B_g \rangle$	1.32 kHz	$\langle B_r \rangle$	0.78 kHz
4 M	$(0.6765/0.4000)/((3.6561/((-0.0500*x^3)+(0.3026*x^2)+0.5748*x+-0.0513))-1)$							
	g_g	0.70	α	2.2 %	$\langle B_g \rangle$	1.52 kHz	$\langle B_r \rangle$	0.85 kHz
4.5 M	$(0.6641/0.4100)/((3.6344/((-0.0508*x^3)+(0.3011*x^2)+0.5900*x+-0.0528))-1)$							
	g_g	0.70	α	2.2 %	$\langle B_g \rangle$	1.51 kHz	$\langle B_r \rangle$	0.81 kHz
5 M	$(0.6328/0.4100)/((3.5787/((-0.0526*x^3)+(0.2930*x^2)+0.6400*x+-0.0585))-1)$							
	g_g	0.70	α	2.2 %	$\langle B_g \rangle$	1.35 kHz	$\langle B_r \rangle$	0.74 kHz
5.5 M	$(0.6375/0.4200)/((3.5627/((-0.0535*x^3)+(0.3018*x^2)+0.6178*x+-0.0556))-1)$							
	g_g	0.70	α	2.2 %	$\langle B_g \rangle$	1.98 kHz	$\langle B_r \rangle$	0.94 kHz
6 M	$(0.6540/0.4000)/((3.6132/((-0.0510*x^3)+(0.2966*x^2)+0.6078*x+-0.0549))-1)$							
	g_g	0.70	α	2.2 %	$\langle B_g \rangle$	1.93 kHz	$\langle B_r \rangle$	0.88 kHz
6.5 M	$(0.6298/0.4200)/((3.5303/((-0.0544*x^3)+(0.3040*x^2)+0.6196*x+-0.0555))-1)$							
	g_g	0.70	α	2.2 %	$\langle B_g \rangle$	1.83 kHz	$\langle B_r \rangle$	0.85 kHz
7 M	$(0.6229/0.4200)/((3.5215/((-0.0546*x^3)+(0.3016*x^2)+0.6302*x+-0.0565))-1)$							
	g_g	0.70	α	2.2 %	$\langle B_g \rangle$	2.04 kHz	$\langle B_r \rangle$	0.91 kHz
7.5 M	$(0.6584/0.4000)/((3.6012/((-0.0510*x^3)+(0.3002*x^2)+0.5936*x+-0.0532))-1)$							
	g_g	0.70	α	2.2 %	$\langle B_g \rangle$	2.06 kHz	$\langle B_r \rangle$	0.95 kHz
#	9				S44pAcPhe R119C			
0 M	$(0.8302/0.3300)/((4.2502/((-0.0337*x^3)+(0.2562*x^2)+0.5307*x+-0.0490))-1)$							
	g_g	0.90	α	2.0 %	$\langle B_g \rangle$	1.67 kHz	$\langle B_r \rangle$	0.77 kHz
2 M	$(0.8126/0.3900)/((4.1397/((-0.0357*x^3)+(0.2660*x^2)+0.5212*x+-0.0474))-1)$							
	g_g	0.70	α	1.7 %	$\langle B_g \rangle$	2.42 kHz	$\langle B_r \rangle$	1.11 kHz
3 M	$(0.7999/0.3900)/((4.0886/((-0.0370*x^3)+(0.2708*x^2)+0.5218*x+-0.0472))-1)$							

UNFOLDING PATHWAY OF T4L

	g_g	0.70	α	1.7 %	$\langle B_g \rangle$	1.96 kHz	$\langle B_r \rangle$	1.01 kHz
4 M	$(0.7835/0.3900)/((4.0351/((-0.0382*x^3)+(0.2738*x^2)+0.5286*x+-0.0476))-1)$							
	g_g	0.70	α	1.7 %	$\langle B_g \rangle$	2.75 kHz	$\langle B_r \rangle$	1.23 kHz
4.5 M	$(0.7806/0.4000)/((4.0152/((-0.0387*x^3)+(0.2759*x^2)+0.5267*x+-0.0474))-1)$							
	g_g	0.70	α	1.7 %	$\langle B_g \rangle$	2.51 kHz	$\langle B_r \rangle$	1.17 kHz
5 M	$(0.7751/0.4000)/((3.9648/((-0.0395*x^3)+(0.2800*x^2)+0.5212*x+-0.0468))-1)$							
	g_g	0.75	α	2.6 %	$\langle B_g \rangle$	2.44 kHz	$\langle B_r \rangle$	1.00 kHz
5.5 M	$(0.7716/0.4100)/((3.9270/((-0.0405*x^3)+(0.2856*x^2)+0.5128*x+-0.0456))-1)$							
	g_g	0.70	α	1.7 %	$\langle B_g \rangle$	3.19 kHz	$\langle B_r \rangle$	1.30 kHz
6 M	$(0.7630/0.4100)/((3.9431/((-0.0401*x^3)+(0.2805*x^2)+0.5277*x+-0.0469))-1)$							
	g_g	0.75	α	2.6 %	$\langle B_g \rangle$	3.04 kHz	$\langle B_r \rangle$	1.14 kHz
6.5 M	$(0.7583/0.4200)/((3.8798/((-0.0417*x^3)+(0.2887*x^2)+0.5185*x+-0.0462))-1)$							
	g_g	0.70	α	1.7 %	$\langle B_g \rangle$	2.42 kHz	$\langle B_r \rangle$	1.09 kHz
7 M	$(0.7423/0.4200)/((3.8111/((-0.0436*x^3)+(0.2947*x^2)+0.5209*x+-0.0462))-1)$							
	g_g	0.70	α	1.7 %	$\langle B_g \rangle$	2.73 kHz	$\langle B_r \rangle$	1.19 kHz
7.5 M	$(0.7570/0.4100)/((3.8846/((-0.0408*x^3)+(0.2848*x^2)+0.5204*x+-0.0459))-1)$							
	g_g	0.75	α	2.6 %	$\langle B_g \rangle$	3.46 kHz	$\langle B_r \rangle$	1.35 kHz
#	13			N55pAcPhe R119C				
0 M	$(0.7836/0.3700)/((4.0828/((-0.0386*x^3)+(0.2710*x^2)+0.5482*x+-0.0511))-1)$							
	g_g	0.80	α	2.4 %	$\langle B_g \rangle$	0.89 kHz	$\langle B_r \rangle$	0.92 kHz
2 M	$(0.7733/0.3700)/((3.9926/((-0.0401*x^3)+(0.2797*x^2)+0.5331*x+-0.0487))-1)$							
	g_g	0.80	α	2.4 %	$\langle B_g \rangle$	1.10 kHz	$\langle B_r \rangle$	0.79 kHz
3 M	$(0.7493/0.3700)/((3.8086/((-0.0437*x^3)+(0.2990*x^2)+0.5057*x+-0.0443))-1)$							
	g_g	0.80	α	2.4 %	$\langle B_g \rangle$	1.30 kHz	$\langle B_r \rangle$	0.92 kHz
4 M	$(0.7408/0.3700)/((3.9035/((-0.0425*x^3)+(0.2833*x^2)+0.5532*x+-0.0502))-1)$							
	g_g	0.80	α	2.4 %	$\langle B_g \rangle$	1.40 kHz	$\langle B_r \rangle$	0.92 kHz
4.5 M	$(0.7307/0.3700)/((3.8635/((-0.0434*x^3)+(0.2861*x^2)+0.5544*x+-0.0501))-1)$							
	g_g	0.80	α	2.4 %	$\langle B_g \rangle$	1.41 kHz	$\langle B_r \rangle$	0.91 kHz
5 M	$(0.7292/0.3700)/((3.8355/((-0.0441*x^3)+(0.2901*x^2)+0.5485*x+-0.0494))-1)$							
	g_g	0.80	α	2.4 %	$\langle B_g \rangle$	1.44 kHz	$\langle B_r \rangle$	0.91 kHz
5.5 M	$(0.7278/0.3700)/((3.8205/((-0.0443*x^3)+(0.2909*x^2)+0.5471*x+-0.0492))-1)$							
	g_g	0.80	α	2.1 %	$\langle B_g \rangle$	1.70 kHz	$\langle B_r \rangle$	0.89 kHz
6 M	$(0.7238/0.3700)/((3.7806/((-0.0452*x^3)+(0.2961*x^2)+0.5392*x+-0.0481))-1)$							
	g_g	0.80	α	2.1 %	$\langle B_g \rangle$	1.77 kHz	$\langle B_r \rangle$	0.90 kHz
6.5 M	$(0.7275/0.3800)/((3.7871/((-0.0450*x^3)+(0.2965*x^2)+0.5345*x+-0.0475))-1)$							
	g_g	0.80	α	2.1 %	$\langle B_g \rangle$	1.76 kHz	$\langle B_r \rangle$	0.93 kHz
7 M	$(0.7310/0.3700)/((3.7964/((-0.0445*x^3)+(0.2954*x^2)+0.5312*x+-0.0471))-1)$							
	g_g	0.80	α	2.1 %	$\langle B_g \rangle$	1.82 kHz	$\langle B_r \rangle$	0.95 kHz
7.5 M	$(0.7211/0.3900)/((3.7602/((-0.0454*x^3)+(0.2978*x^2)+0.5343*x+-0.0472))-1)$							
	g_g	0.80	α	2.1 %	$\langle B_g \rangle$	1.88 kHz	$\langle B_r \rangle$	0.94 kHz
#	16			K60pAcPhe R119C				
0 M	$(0.8035/0.3700)/((4.1544/((-0.0365*x^3)+(0.2651*x^2)+0.5404*x+-0.0503))-1)$							
	g_g	0.75	α	1.6 %	$\langle B_g \rangle$	1.24 kHz	$\langle B_r \rangle$	0.78 kHz
2 M	$(0.7877/0.3700)/((4.0545/((-0.0381*x^3)+(0.2728*x^2)+0.5318*x+-0.0486))-1)$							
	g_g	0.90	α	1.7 %	$\langle B_g \rangle$	2.22 kHz	$\langle B_r \rangle$	1.09 kHz
3 M	$(0.7806/0.3700)/((4.0123/((-0.0391*x^3)+(0.2770*x^2)+0.5284*x+-0.0480))-1)$							
	g_g	0.90	α	1.7 %	$\langle B_g \rangle$	2.52 kHz	$\langle B_r \rangle$	1.17 kHz
4 M	$(0.7698/0.3700)/((3.9612/((-0.0401*x^3)+(0.2805*x^2)+0.5288*x+-0.0478))-1)$							
	g_g	0.90	α	1.7 %	$\langle B_g \rangle$	2.81 kHz	$\langle B_r \rangle$	1.23 kHz
4.5 M	$(0.7630/0.3800)/((3.9221/((-0.0410*x^3)+(0.2845*x^2)+0.5260*x+-0.0474))-1)$							
	g_g	0.90	α	1.7 %	$\langle B_g \rangle$	2.85 kHz	$\langle B_r \rangle$	1.23 kHz
5 M	$(0.7578/0.3700)/((3.9074/((-0.0412*x^3)+(0.2840*x^2)+0.5299*x+-0.0477))-1)$							
	g_g	0.80	α	1.7 %	$\langle B_g \rangle$	2.38 kHz	$\langle B_r \rangle$	1.22 kHz
5.5 M	$(0.7596/0.3800)/((3.8988/((-0.0412*x^3)+(0.2857*x^2)+0.5235*x+-0.0468))-1)$							
	g_g	0.90	α	1.7 %	$\langle B_g \rangle$	3.07 kHz	$\langle B_r \rangle$	1.26 kHz
6 M	$(0.7614/0.3800)/((3.8957/((-0.0411*x^3)+(0.2864*x^2)+0.5194*x+-0.0463))-1)$							
	g_g	0.90	α	1.7 %	$\langle B_g \rangle$	3.04 kHz	$\langle B_r \rangle$	1.24 kHz

UNFOLDING PATHWAY OF T4L

6.5 M	$(0.7575/0.3800)/((3.8760/((-0.0415*x^3)+(0.2878*x^2)+0.5187*x+-0.0461))-1)$							
	g_g	0.90	α	1.7 %	$\langle B_g \rangle$	3.40 kHz	$\langle B_r \rangle$	1.33 kHz
7 M	$(0.7553/0.3800)/((3.8645/((-0.0417*x^3)+(0.2888*x^2)+0.5183*x+-0.0460))-1)$							
	g_g	0.90	α	1.7 %	$\langle B_g \rangle$	3.73 kHz	$\langle B_r \rangle$	1.38 kHz
7.5 M	$(0.7512/0.3800)/((3.8493/((-0.0420*x^3)+(0.2894*x^2)+0.5197*x+-0.0460))-1)$							
	g_g	0.90	α	1.7 %	$\langle B_g \rangle$	3.54 kHz	$\langle B_r \rangle$	1.39 kHz
#	17				K60pAcPhe N132C			
0 M	$(0.7740/0.4100)/((4.1224/((-0.0379*x^3)+(0.2638*x^2)+0.5689*x+-0.0528))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.36 kHz	$\langle B_r \rangle$	0.43 kHz
0.5 M	$(0.7574/0.4100)/((4.0688/((-0.0389*x^3)+(0.2654*x^2)+0.5766*x+-0.0534))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.75 kHz	$\langle B_r \rangle$	0.73 kHz
1.0 M	$(0.7541/0.4000)/((4.0277/((-0.0397*x^3)+(0.2697*x^2)+0.5696*x+-0.0524))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.48 kHz	$\langle B_r \rangle$	0.38 kHz
1.5 M	$(0.7438/0.4000)/((4.0088/((-0.0401*x^3)+(0.2691*x^2)+0.5782*x+-0.0531))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.41 kHz	$\langle B_r \rangle$	0.48 kHz
2.0M	$(0.7476/0.3900)/((3.9799/((-0.0408*x^3)+(0.2746*x^2)+0.5650*x+-0.0517))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.46 kHz	$\langle B_r \rangle$	0.48 kHz
2.5 M	$(0.7448/0.4000)/((3.9703/((-0.0408*x^3)+(0.2745*x^2)+0.5655*x+-0.0516))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.48 kHz	$\langle B_r \rangle$	0.48 kHz
3.0 M	$(0.7372/0.3900)/((3.9408/((-0.0416*x^3)+(0.2765*x^2)+0.5687*x+-0.0519))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.57 kHz	$\langle B_r \rangle$	0.52 kHz
3.5 M	$(0.7260/0.4000)/((3.9114/((-0.0423*x^3)+(0.2769*x^2)+0.5762*x+-0.0524))-1)$							
	g_g	0.90	α	2.2%	$\langle B_g \rangle$	0.60 kHz	$\langle B_r \rangle$	0.52 kHz
4.0 M	$(0.7161/0.3900)/((3.9155/((-0.0421*x^3)+(0.2718*x^2)+0.5932*x+-0.0540))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.54 kHz	$\langle B_r \rangle$	0.46 kHz
4.25 M	$(0.7161/0.3900)/((3.9155/((-0.0421*x^3)+(0.2718*x^2)+0.5932*x+-0.0540))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.55 kHz	$\langle B_r \rangle$	0.47 kHz
4.5 M	$(0.7018/0.4000)/((3.8698/((-0.0433*x^3)+(0.2740*x^2)+0.6015*x+-0.0548))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.56 kHz	$\langle B_r \rangle$	0.47 kHz
4.75 M	$(0.7018/0.4000)/((3.8698/((-0.0433*x^3)+(0.2740*x^2)+0.6015*x+-0.0548))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.71 kHz	$\langle B_r \rangle$	0.50 kHz
5.0 M	$(0.7080/0.3900)/((3.8650/((-0.0432*x^3)+(0.2766*x^2)+0.5893*x+-0.0534))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.73 kHz	$\langle B_r \rangle$	0.50 kHz
5.25 M	$(0.7080/0.3900)/((3.8650/((-0.0432*x^3)+(0.2766*x^2)+0.5893*x+-0.0534))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.75 kHz	$\langle B_r \rangle$	0.51 kHz
5.5 M	$(0.6954/0.4000)/((3.8481/((-0.0438*x^3)+(0.2753*x^2)+0.6024*x+-0.0546))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.77 kHz	$\langle B_r \rangle$	0.51 kHz
5.75	$(0.6954/0.4000)/((3.8481/((-0.0438*x^3)+(0.2753*x^2)+0.6024*x+-0.0546))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.81 kHz	$\langle B_r \rangle$	0.52 kHz
6 M	$(0.7105/0.3900)/((3.8700/((-0.0430*x^3)+(0.2769*x^2)+0.5844*x+-0.0528))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.82 kHz	$\langle B_r \rangle$	0.51 kHz
6.5 M	$(0.7205/0.3900)/((3.8419/((-0.0435*x^3)+(0.2845*x^2)+0.5613*x+-0.0503))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.85 kHz	$\langle B_r \rangle$	0.52 kHz
7 M	$(0.7123/0.3900)/((3.8329/((-0.0437*x^3)+(0.2833*x^2)+0.5688*x+-0.0509))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.89 kHz	$\langle B_r \rangle$	0.53 kHz
7.5 M	$(0.7184/0.3900)/((3.8266/((-0.0436*x^3)+(0.2855*x^2)+0.5577*x+-0.0497))-1)$							
	g_g	0.90	α	2.2 %	$\langle B_g \rangle$	0.92 kHz	$\langle B_r \rangle$	0.53 kHz

4.3.2.5. Additional fullFCS results

Figure 4.3.8 shows the results of approximating the MFD data analyzed with 1 ms TW's. Here, the two variants N55pAcF R119C and K60pAcF R119C show a rather gradual increase in mean distance, whereas in the variants K19pAcF P86C and S44pAcF R119C a sudden jump from the native to the unfolded state seems to occur. The variant K60pAcF N132C behaves like a mixture of the two extreme described before: firstly, at low urea

concentrations the mean distance is gradually increasing, then at ~ 6 M the jump to the unfolded state occurs.

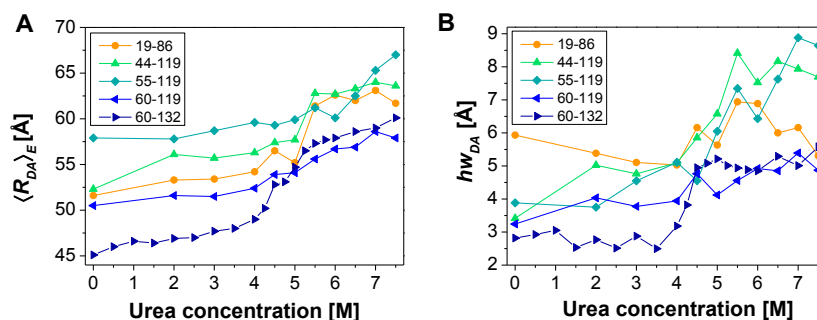


Figure 4.3.8 Summary of PDA results. (A-B) Mean distance and half-width obtained from fitting 1 ms TW MFD data with a single Gaussian distribution.

Table 4.3.4 Fitting results for Static PDA, 1 ms time-window (Section 4.3.1.7)

[M urea]	$\langle R_{DA} \rangle_E$ [Å]	(hw_{DA}) [Å]	x_{DOnly} [%]
K19pAcF P86C			
0	51.6±0.3	(5.9±0.04)	30.2±0.2
2	53.3±0.3	(5.4±0.03)	29.9±0.2
3	53.4±0.3	(5.1±0.03)	31.9±0.2
4	54.2±0.3	(5.0±0.03)	28.8±0.2
4.5	56.5±0.5	(6.2±0.05)	36.3±0.3
5	55.2±0.5	(5.6±0.05)	36.6±0.3
5.5	61.4±0.4	(6.9±0.05)	27.3±0.2
6	62.6±0.5	(6.9±0.06)	27.5±0.2
6.5	62.0±0.4	(6.0±0.04)	23.4±0.1
7	63.1±0.2	(6.2±0.02)	14.1±0.1
7.5	61.7±0.3	(5.3±0.03)	20.4±0.1
S44pAcF R119C			
0	52.3±0.2	(3.4±0.01)	51.7±0.2
2	56.1±0.1	(5.0±0.01)	24.1±0.1
3	55.7±0.2	(4.8±0.02)	35.5±0.1
4	56.3±0.2	(5.1±0.02)	33.4±0.1
4.5	57.4±0.3	(5.9±0.03)	37±0.2
5	57.7±0.3	(6.6±0.03)	27±0.1
5.5	62.8±0.4	(8.4±0.05)	28.7±0.2
6	62.7±0.4	(7.5±0.04)	24.0±0.1
6.5	63.3±0.4	(8.2±0.06)	26.4±0.2
7	64.0±0.3	(7.9±0.04)	18.4±0.1
7.5	63.6±0.3	(7.7±0.03)	17.1±0.1
N55pAcF R119C			
0	57.9±0.7	(3.9±0.05)	52.5±0.7
2	57.8±0.5	(3.8±0.03)	46.8±0.4
3	58.7±0.6	(4.5±0.04)	41.9±0.4
4	59.6±0.7	(5.1±0.06)	47.3±0.5
4.5	59.3±0.9	(4.6±0.07)	50.3±0.8
5	59.9±1.3	(6.0±0.1)	50.3±1.1
5.5	61.2±1.1	(7.3±0.1)	56.5±1.0
6	60.1±0.6	(6.4±0.07)	47.1±0.5
6.5	62.5±1.1	(7.6±0.1)	48.3±0.9
7	65.3±1.6	(8.9±0.2)	53.8±1.3
7.5	67.0±1.2	(8.6±0.2)	56.2±1.0
K60pAcF R119C			
0	50.5±0.5	(3.2±0.03)	52.2±0.50
2	51.6±0.1	(4.0±0.01)	24.4±0.06
3	51.5±0.2	(3.8±0.01)	24.6±0.07
4	52.4±0.2	(3.9±0.01)	23.1±0.07
4.5	53.9±0.2	(4.8±0.01)	27.8±0.09
5	54.1±0.3	(4.1±0.03)	30.6±0.20
5.5	55.6±0.2	(4.6±0.02)	27.4±0.10
6	56.7±0.3	(4.9±0.02)	38.0±0.17
6.5	56.9±0.3	(4.9±0.02)	36.1±0.17
7	58.6±0.3	(5.4±0.03)	38.8±0.21
7.5	57.9±0.4	(4.9±0.04)	40.5±0.31
K60pAcF N132C			
0	45.1±0.2	(2.8±0.01)	27.6±0.10
0.5	40.6±0.2	(2.9±0.01)	28.8±0.10
1	46.6±0.4	(3.1±0.03)	30.7±0.30
1.5	46.4±0.3	(2.5±0.01)	25.9±0.20
2	46.9±0.4	(2.8±0.02)	25.6±0.10
2.5	47.0±0.30	(2.5±0.01)	21.9±0.10
3	47.7±0.2	(2.9±0.01)	16.7±0.10
3.5	48.0±0.1	(2.5±0.01)	17.8±0.04
4	49.0±0.3	(3.2±0.02)	17.2±0.10
4.25	50.2±0.70	(3.8±0.05)	20.2±0.30
4.5	52.8±0.69	(4.9±0.06)	14.1±0.20
4.75	53.1±0.36	(5.1±0.03)	20.4±0.10
5	54.7±0.37	(5.2±0.03)	20.5±0.10
5.25	56.5±0.31	(5.0±0.03)	19.1±0.01
5.5	57.3±0.37	(4.9±0.03)	21.4±0.10
5.75	57.7±0.40	(4.9±0.03)	22.4±0.20
6	57.9±0.39	(4.9±0.03)	21.8±0.10
6.5	58.6±0.40	(5.3±0.04)	24.1±0.20
7	59.0±0.36	(5.0±0.03)	22.5±0.10
7.5	60.1±0.43	(5.6±0.04)	21.4±0.20

4.3.2.6. Additional fullFCS results

In Figure 4.3.9D exemplary correlation curves for FCS analysis are shown. The collected photon stream was correlated from 32 ps up to 10 sec. All curves show a right shift of the diffusion time, indicative of an increase in size of the unfolded protein compared to its native state – which is expected. Additionally, data measured in high concentrations of denaturants are noisier than the data measured under native conditions. Only in the red-autocorrelation curves, some of the different relaxation times can be differentiated “by eye” due to their characteristic correlation terms being visible.

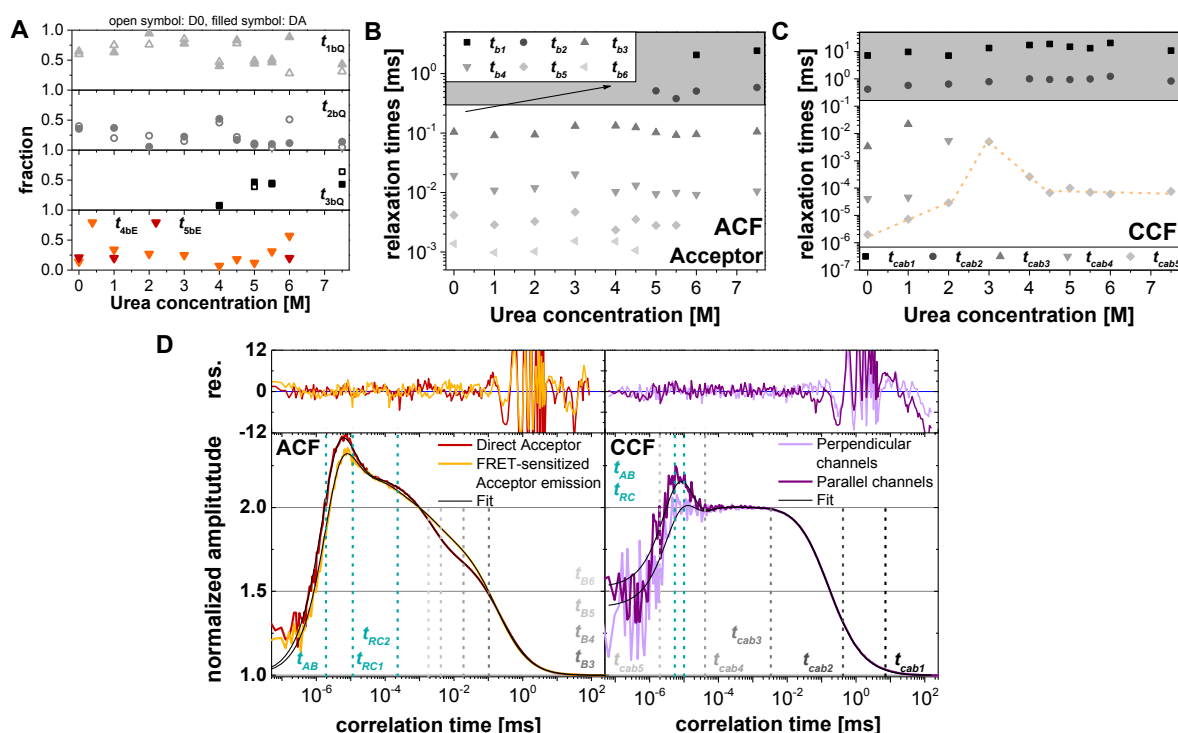
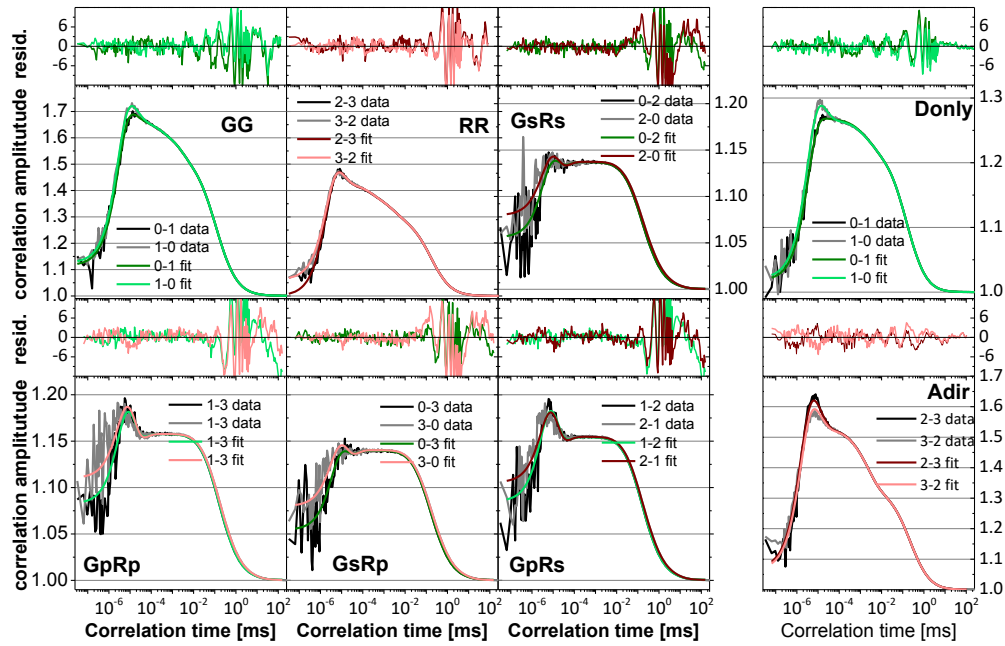


Figure 4.3.9 Exemplary fullFCS curves of the variant K60pAcF N132C and the relaxation times obtained. (A) Amplitudes of relaxation time constants obtained from global fitting of D0- and DA-labeled samples. (B, C) Relaxation times obtained from fitting acceptor color auto- correlation curves (top) or the eight color-cross correlation (bottom) curves globally. Gray shaded area indicates times slower than the diffusion time (Table 4.3.5). (D) Example of full FCS curves for the variant K60pAcF N132C-(DA) in the absence of urea. Shown are on the left the ACF's in the red channels, direct acceptor excited acceptor emission and FRET sensitized acceptor emission (i.e. donor excitation), and on the right exemplary two cross correlations curve of donor and acceptor channels. Residuals are shown on top. Vertical lines show times found (Table 4.3.5).

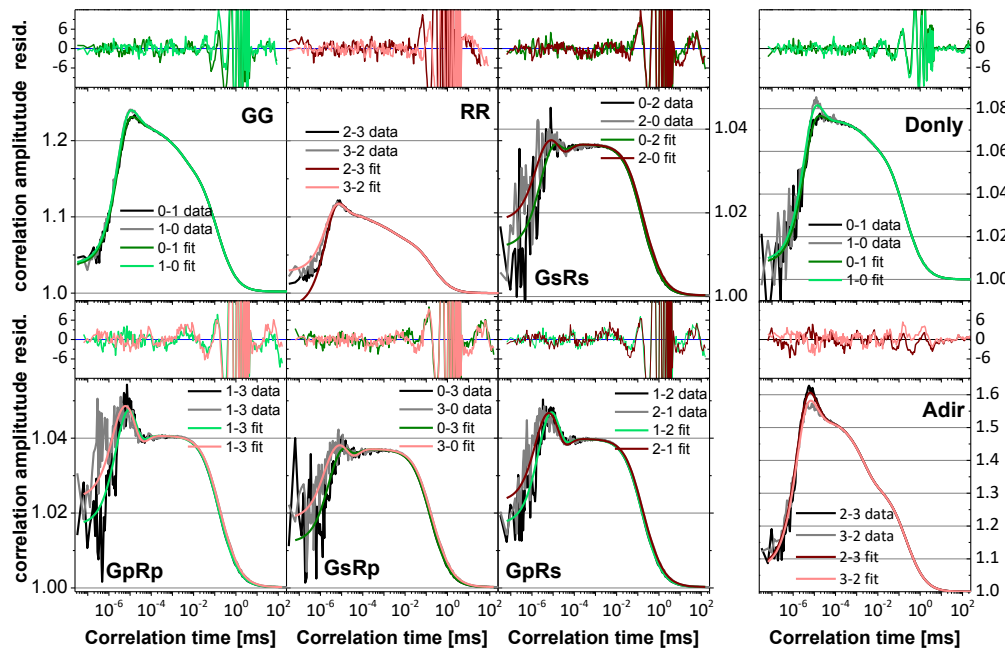
From Figure 4.3.9B it is evident that for both the acceptor-only and the FRET-sensitized acceptor signal, its dynamics slows down with increasing urea concentrations. The faster relaxation times in the μ s-time range disappear whereas slower times in the ms-range appear. However, all relaxation times slower than the diffusion time (gray shaded area) have to be handled with care as (1) the fit is not really reliable here and (2) they may have been introduced artificially because the detection area may no longer be a 3D Gaussian shaped

area but distorted due the changed index of refraction in the higher urea concentrations (142). For green-red color-crosscorrelations, up to 2 M urea two to three additional times faster than diffusion term have been detected. If connecting the fastest times found by line, partially a turned “V” can be recognized with a peak a 3 M urea – similar to that already found more clearly in the green color-autocorrelation from the double-labeled samples (main text, Figure 4.2.4C). The fastest time found in higher urea concentrations at ~ 50 -100 ns was also found having the highest amplitude (up to 80 %) in filteredFCS measurements.

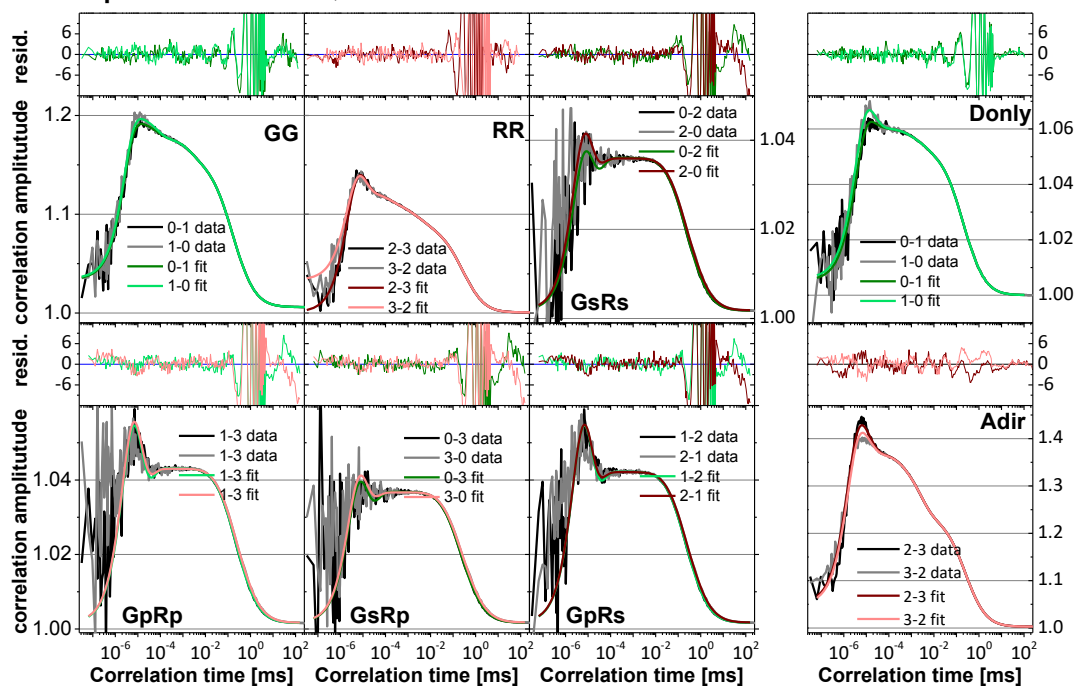
A K60pAcF N132C, 0 M urea



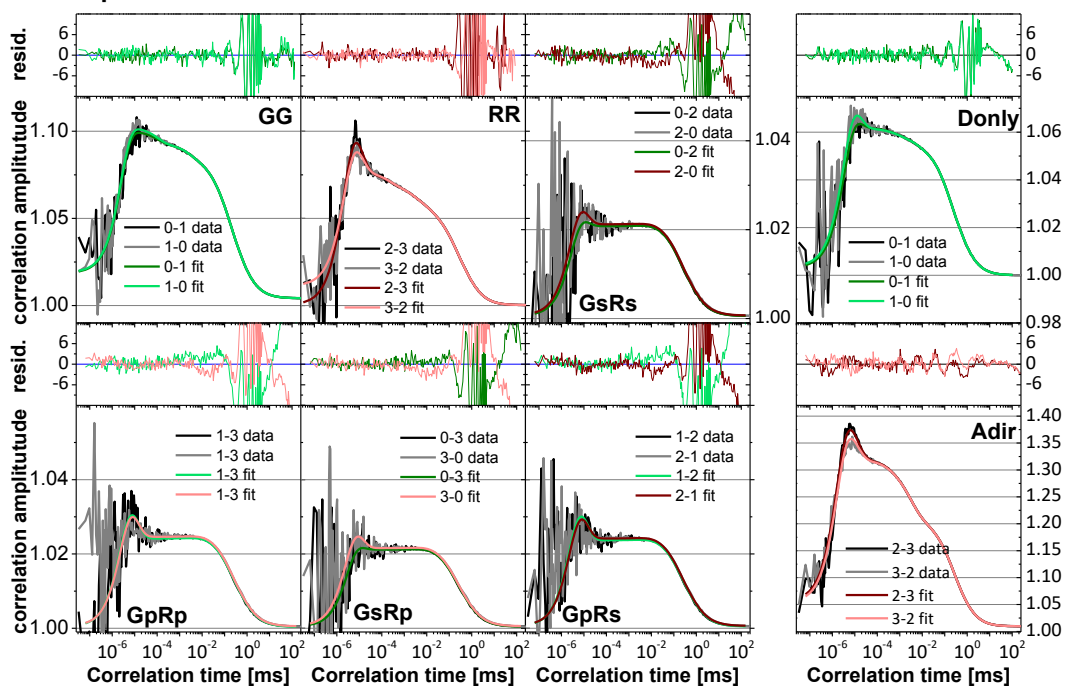
B K60pAcF N132C, 1 M urea



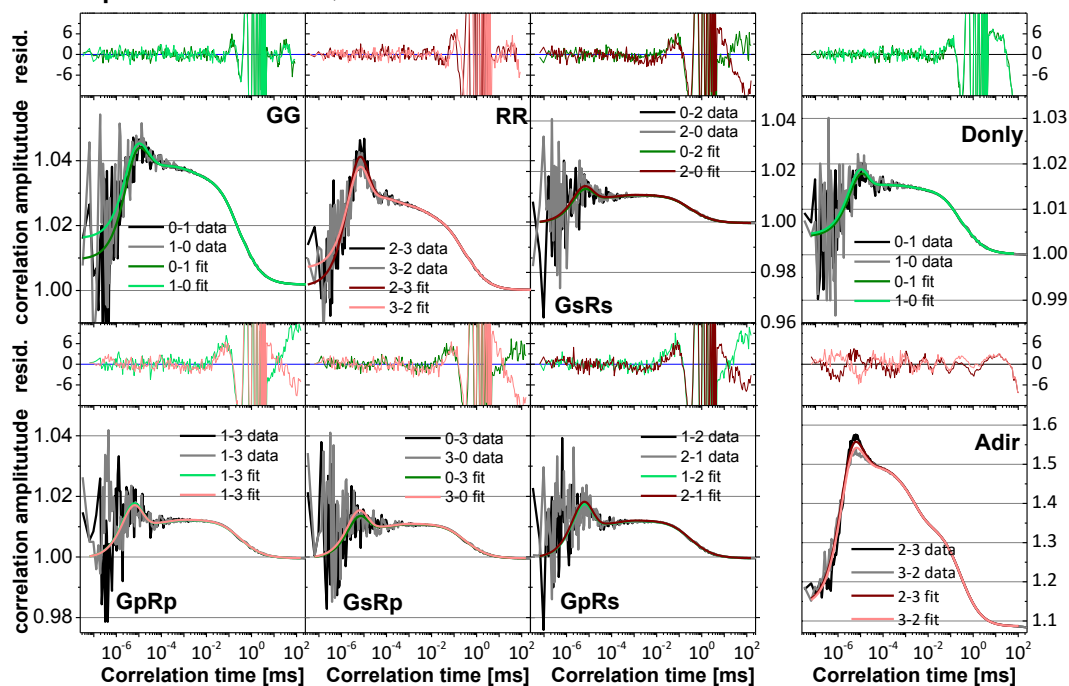
C K60pAcF N132C, 2 M urea



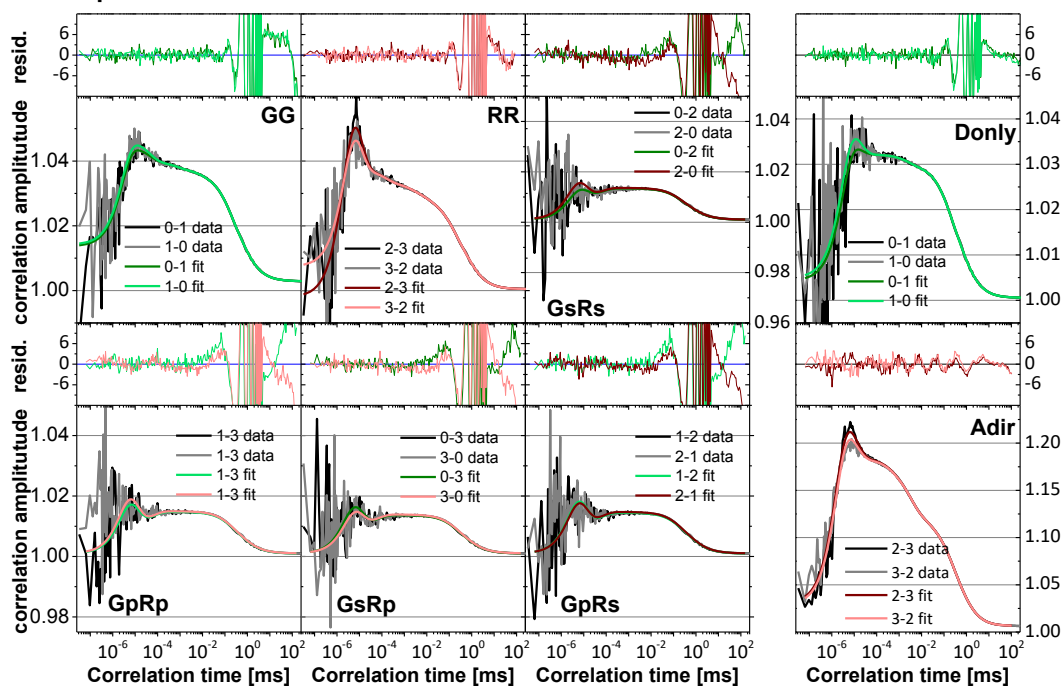
D K60pAcF N132C, 3 M urea



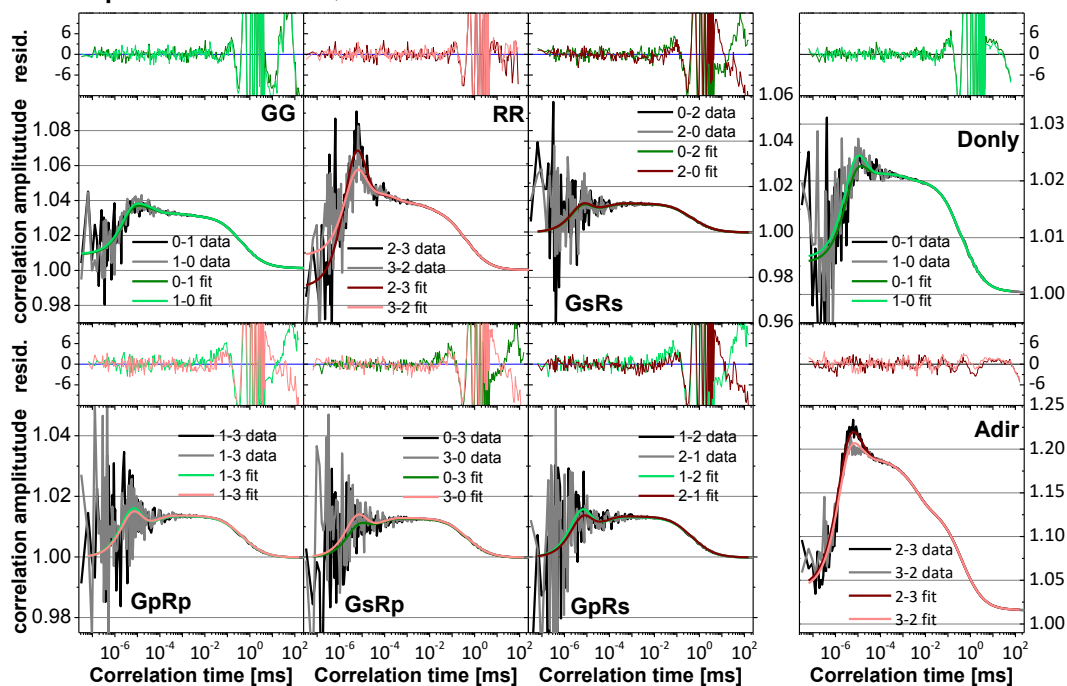
E K60pAcF N132C, 4 M urea



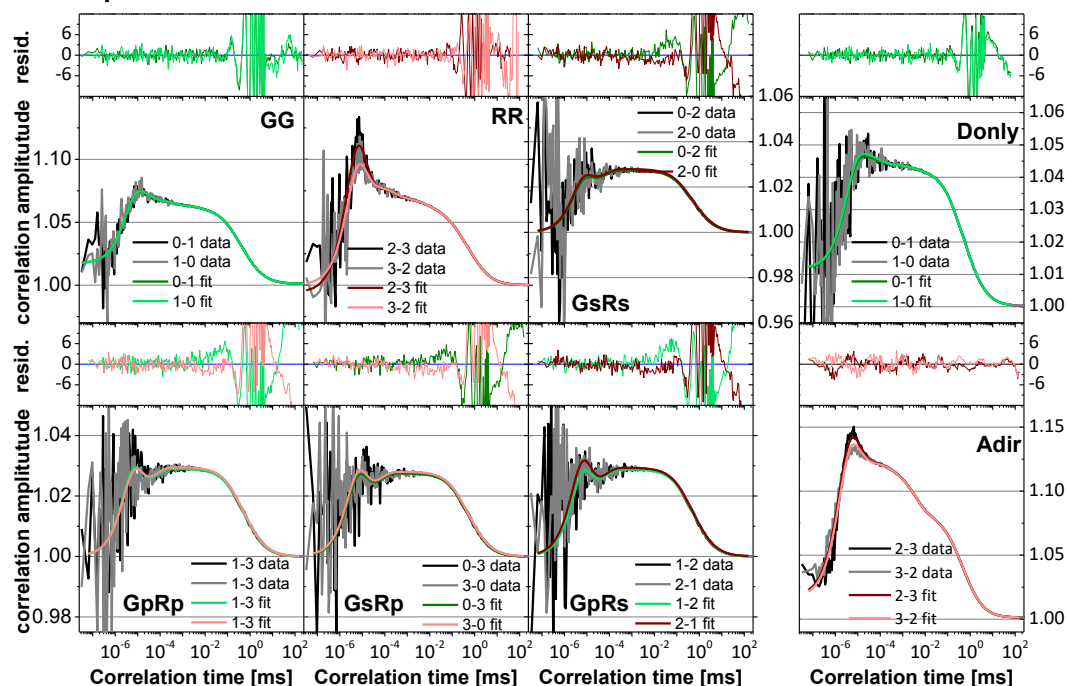
F K60pAcF N132C, 4.5 M urea



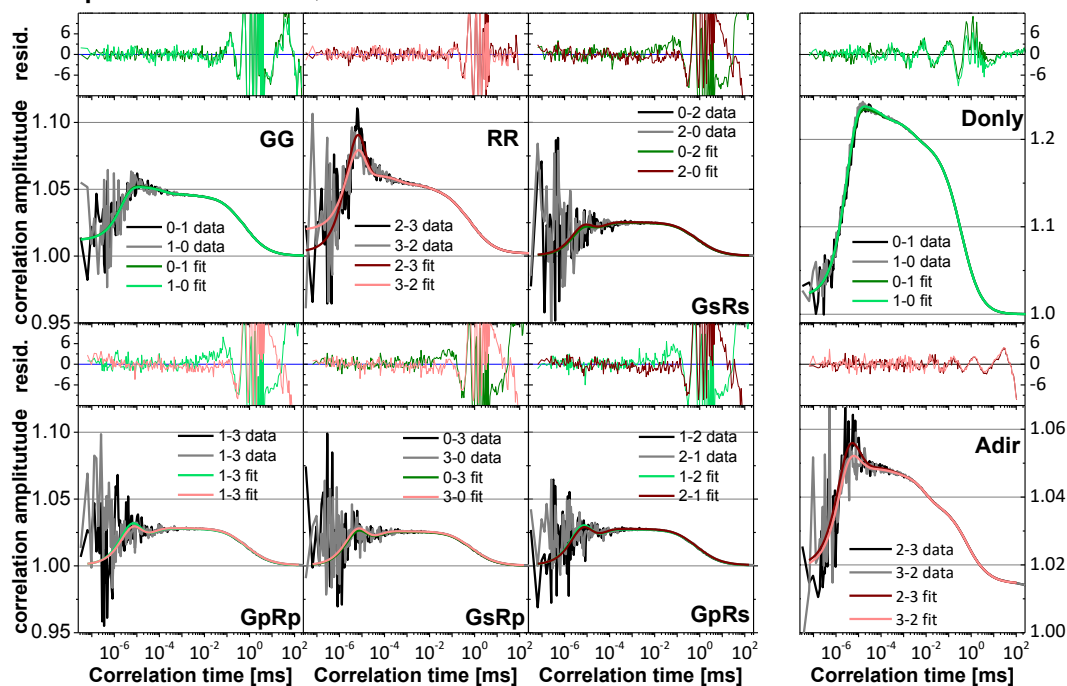
G K60pAcF N132C, 5 M urea



H K60pAcF N132C, 5.5 M urea



I K60pAcF N132C, 6 M urea



J K60pAcF N132C, 7.5 M urea

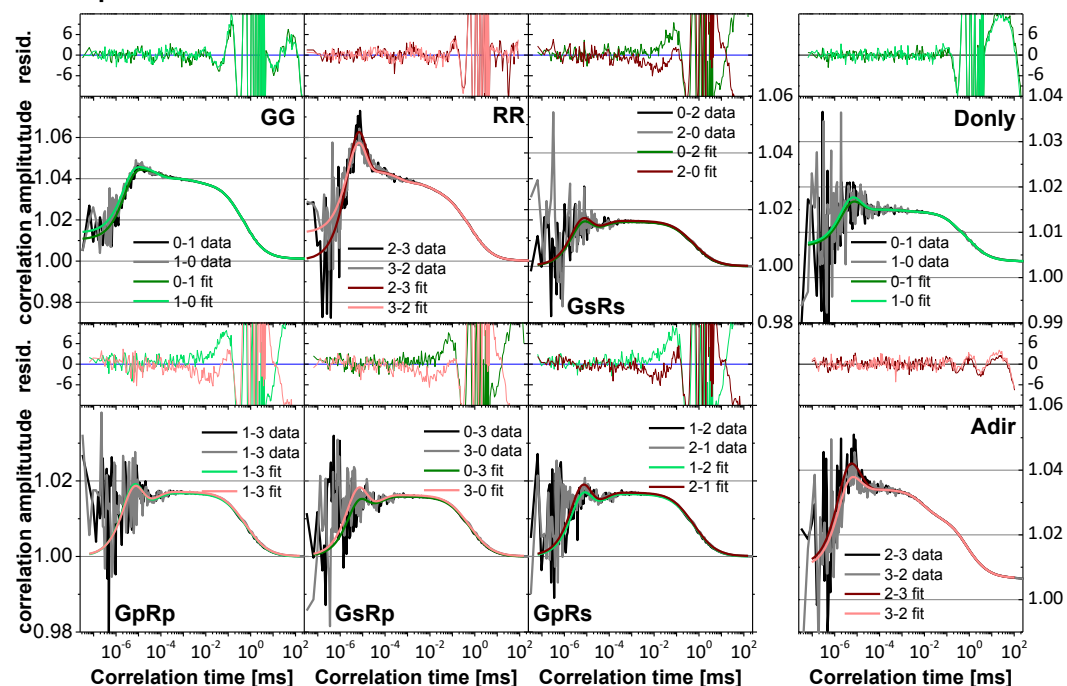


Figure 4.3.10 fullFCS curves of the variant K60pAcF N132C. Fit results are summarized in Table 4.3.5.

UNFOLDING PATHWAY OF T4L

Table 4.3.5 Fit results for fullFCS measurements of variant 60-132. Gray shaded cells indicate where values have been fixed for fitting. Uncertainties given are standard errors of the mean.

0 M																	
Variable	DOnly		DA		AOnly		DA		Variable	DA							
	g _s -g _p	g _p -g _s	g _s -g _p	g _p -g _s	r _s -r _p	r _p -r _s	r _s -r _p	r _p -r _s		g _p -r _p	g _p -r _s	g _s -r _p	g _s -r _s	r _p -g _p	r _p -g _s	r _s -g _p	r _s -g _s
<i>b</i>	1				1				<i>b</i>	1							
<i>N</i>	4.42		1.75		1.96		2.60		<i>N</i>	0.58	0.56	0.88	0.86	0.43	0.60	0.45	0.60
<i>t_d</i> [ms]	0.20		0.26		0.29		0.24		<i>t_d</i> [ms]	0.13							
<i>z₀/ω₀</i>	5.63		23.36		6.53		13.84		<i>z₀/ω₀</i>	3.50							
<i>x_{ab}</i>	0.93±0.06		0.84±0.004		1.00	0.81	1.00	0.89	<i>x_{ab}</i>	0.55							
<i>t_{ab}</i> [ms]	(3.85±0.12)*10 ⁻⁶		(2.94±0.11)*10 ⁻⁶		(1.73±0.04)*10 ⁻⁶		(2.07±0.05)*10 ⁻⁶		<i>t_{ab}</i> [ms]	(5.37±3.58)*10 ⁻⁶							
<i>x_{rc1}</i>	0.16	0.16	0.16	0.16	0.40	0.21	0.29	0.27	<i>x_{rc}</i>	0.48	0.50	0.29	0.32	0.52	0.39	0.51	0.40
<i>t_{rc1}</i> [ms]	(2.66±0.03)*10 ⁻³				(1.14±0.02)*10 ⁻⁵				<i>t_{rc}</i> [ms]	1.0*10 ⁻⁵							
<i>x_{rc2}</i>	0.11	0.33	0.10	0.21	0.04	0.04	0.11	0.11	<i>f</i>	0.89							
<i>t_{rc2}</i> [ms]	(2.09±0.09)*10 ⁻⁵				(2.31±0.04)*10 ⁻⁴				<i>x_{cab1}</i>	0.06	0.06	0.06	0.06	0.02	0.02	0.02	0.02
									<i>t_{cab1}</i> [ms]	(1.97±0.75)*10 ⁻⁶							
<i>x_{b1}</i>			0.21±0.006		0.16±0.002		0.29±0.003		<i>x_{cab2}</i>	0.016±0.003							
<i>t_{b1}</i> [ms]	0.314±0.002				0.104±0.002				<i>t_{cab2}</i> [ms]	(4.10±0.80)*10 ⁻⁵							
<i>x_{b2}</i>	0.60±0.009		0.42±0.003		0.10±0.003		0.19±0.003		<i>x_{cab3}</i>	0.003±1.5*10 ⁻⁴							
<i>t_{b2}</i> [ms]	0.083±0.001				0.019±5*10 ⁻⁴				<i>t_{cab3}</i> [ms]	(3.32±0.36)*10 ⁻³							
<i>x_{b3}</i>	0.40±0.010		0.23±0.003		0.38±0.002		0.29±0.001		<i>x_{cab4}</i>	0.07±5*10 ⁻⁴							
<i>t_{b3}</i> [ms]	0.020±2*10 ⁻⁴				(4.16±0.04)*10 ⁻³				<i>t_{cab4}</i> [ms]	0.417±0.004							
<i>x_{b4}</i>			0.14±0.002		0.36±0.001		0.23±0.003		<i>x_{cab5}</i>	0.85±0.05							
<i>t_{b4}</i> [ms]	2.67*10 ⁻⁴ ±7*10 ⁻⁶				(1.83±0.01)*10 ⁻³				<i>t_{cab5}</i> [ms]	7.16±1.25							
1 M																	
<i>b</i>	1				1				<i>b</i>	1							
<i>N</i>	15.31		4.87		1.97		10.5		<i>N</i>	3.02	2.97	4.18	4.33	2.09	2.73	2.19	2.78
<i>t_d</i> [ms]	0.25		0.18		0.27		0.25		<i>t_d</i> [ms]	0.13							
<i>z₀/ω₀</i>	7.53		7.55		6.36		4.88		<i>z₀/ω₀</i>	3.5							
<i>x_{ab}</i>	0.89		0.86		1.0	0.81	1.13	0.80	<i>x_{ab}</i>	0.42							
<i>t_{ab}</i> [ms]	(1.23±0.04)*10 ⁻⁵		(3.38±0.06)*10 ⁻⁶		(1.71±0.04)*10 ⁻⁶		(2.14±0.06)*10 ⁻⁶		<i>t_{ab}</i> [ms]	(1.12±0.19)*10 ⁻⁶							
<i>x_{rc1}</i>	0.14	0.29	0.17	0.19	0.39	0.21	0.38	0.33	<i>x_{rc}</i>	0.58	0.59	0.44	0.43	0.57	0.45	0.56	0.45
<i>t_{rc1}</i> [ms]	(2.26±0.03)*10 ⁻⁵				(1.05±0.02)*10 ⁻⁵				<i>t_{rc}</i> [ms]	1.0*10 ⁻⁵							
<i>x_{rc2}</i>	0.13	0.13	0.23	0.06	0.04	0.04	0.10	0.10	<i>f</i>	0.91							
<i>t_{rc2}</i> [ms]	(2.67±0.09)*10 ⁻³		(3.20±0.11)*10 ⁻⁴		(1.84±0.04)*10 ⁻⁴				<i>x_{cab1}</i>	0.20	0.19	0.18	0.18	0.09	0.08	0.09	0.08
									<i>t_{cab1}</i> [ms]	(7.26±0.24)*10 ⁻⁶							
<i>x_{b1}</i>			0.34±0.004		0.24±0.001		0.19±0.001		<i>x_{cab2}</i>	0.03±0.005							

UNFOLDING PATHWAY OF T4L

t_{b1} [ms]	(2.58±0.09)*10 ⁻³				(9.78±0.18)*10 ⁻⁴				t_{cab2} [ms]	(4.58±1.10)*10 ⁻⁵							
x_{b2}	0.25		0.17±0.002		0.46±0.001		0.31±0.002		x_{cab3}	0.01±2.4*10 ⁻⁴							
t_{b2} [ms]	0.018±0.0003				(2.87±0.02)*10 ⁻³				t_{cab3} [ms]	0.022±0.001							
x_{b3}	0.75		0.29±0.002		0.14±0.001		0.20±0.002		x_{cab4}	0.14±0.001							
t_{b3} [ms]	0.088±0.0007				0.011±2*10 ⁻⁴				t_{cab4} [ms]	0.580±0.005							
x_{b4}			0.20±0.001		0.16±0.002		0.30±0.002		x_{cab5}	0.63±0.03							
t_{b4} [ms]	1.69±0.03				0.091±0.001				t_{cab5} [ms]	9.69±0.81							
2 M																	
b	1.0		1.01		1				b	1							
N	18.23		6.48		2.78		8.95		N	1.99	1.98	3.00	3.23	1.93	2.86	1.96	2.79
t_d [ms]	0.30		0.27		0.30		0.39		t_d [ms]	0.17							
z_0/ω_0	6.54		9.22		6.93		5.97		z_0/ω_0	3.5							
x_{ab}	0.96		0.85		1.03	0.81	1.1	0.83	x_{ab}	1							
t_{ab} [ms]	(2.82±0.18)*10 ⁻⁶	(2.58±0.11)*10 ⁻⁶			(1.66±0.04)*10 ⁻⁶	(2.06±0.09)*10 ⁻⁶			t_{ab} [ms]	3.00*10 ⁻⁶							
x_{rc1}	0.09	0.09	0.12	0.12	0.38	0.19	0.41	0.35	x_{rc}	0.67	0.68	0.56	0.53	0.61	0.48	0.62	0.50
t_{rc1} [ms]	(2.16±0.05)*10 ⁻³				(1.03±0.03)*10 ⁻⁵				t_{rc} [ms]	1.0*10 ⁻⁵							
x_{rc2}	0.04	0.11	0.11	0.15	0.04	0.04	0.09	0.09	f	0.84							
t_{rc2} [ms]	(6.64±0.60)*10 ⁻⁵				(1.57±0.04)*10 ⁻⁴				x_{cab1}	0.11	0.11	0.10	0.10	0.06	0.04	0.06	0.04
									t_{cab1} [ms]	(2.87±0.49)*10 ⁻⁵							
x_{b1}	0.24±0.008		0.04±0.001		0.25±8*10 ⁻⁴		0.18±0.002		x_{cab2}	0.01±5.4*10 ⁻⁴							
t_{b1} [ms]	0.011±0.0003				(1.02±0.01)*10 ⁻³				t_{cab2} [ms]	(5.34±0.76)*10 ⁻³							
x_{b2}			0.27±0.006		0.47±5*10 ⁻⁵		0.23±0.001		x_{cab3}	0.17±0.001							
t_{b2} [ms]	0.015±0.002				(3.26±0.02)*10 ⁻³				t_{cab3} [ms]	0.64±0.10							
x_{b3}	0.76±0.007		0.69±0.006		0.12±0.001		0.22±0.003		x_{cab4}	0.72±0.02							
t_{b3} [ms]	0.90±0.001				0.012±3*10 ⁻⁴				t_{cab4} [ms]	7.07±1.10							
x_{b4}					0.15±0.002		0.36±0.003		x_{cab5}								
t_{b4} [ms]					0.094±0.001				t_{cab5} [ms]								
3 M																	
b	1				1.01		1.0		b	1							
N	17.85		12.16		3.32		14.32		N	3.47	3.51	6.36	6.05	3.63	4.66	3.73	4.88
t_d [ms]	0.37		0.44		0.34		0.48		t_d [ms]	0.19							
z_0/ω_0	14.46		11.35		8.22		8.99		z_0/ω_0	3.5							
x_{ab}	0.95		0.85		1.00	0.82	1.00	1	x_{ab}	1							
t_{ab} [ms]	(2.82±0.39)*10 ⁻⁶	(2.50±0.11)*10 ⁻⁶			(1.77±0.10)*10 ⁻⁶	(2.54±0.36)*10 ⁻⁶			t_{ab} [ms]	3.00*10 ⁻⁶							
x_{rc1}	0.05	0.07	0.09	0.11	0.43	0.23	0.66	0.46	x_{rc}	0.44	0.45	0.11	0.17	0.41	0.34	0.41	0.32
t_{rc1} [ms]	(1.04±0.06)*10 ⁻⁴				(1.04±0.08)*10 ⁻⁵				t_{rc} [ms]	1.0*10 ⁻⁵							

UNFOLDING PATHWAY OF T4L

x_{rc2}	0.08	0.08	0.08	0.08	0.04	0.04	0.09	0.10	f	0.85
t_{rc2} [ms]	(2.48±0.14)*10 ⁻³				(2.01±0.16)*10 ⁻⁴				x_{cab1}	0.01±0.001
									t_{cab1} [ms]	(4.98±1.24)*10 ⁻³
x_{b1}	0.15±0.009		0.17±0.007		0.32±0.002		0.18±0.01		x_{cab2}	0.12±5*10 ⁻⁴
t_{b1} [ms]	0.017±0.001				(1.54±0.02)*10 ⁻³					0.79±0.01
x_{b2}	0.85±0.01		0.59±0.005		0.42±0.004		0.18±0.02		x_{cab3}	0.87±0.07
t_{b2} [ms]	0.104±0.002				(4.68±0.12)*10 ⁻³					13.39±1.75
x_{b3}			0.25±0.009		0.09±0.005		0.22±0.006		x_{cab4}	
t_{b3} [ms]	0.479±0.002				0.020±0.001					
x_{b4}					0.16±0.005		0.43±0.006		x_{cab5}	
t_{b4} [ms]					0.131±0.001				t_{cab5} [ms]	
4 M										
b	1				1.09		1.0		b	1
N	64.59		27.18		2.59		37.15		N	4.08 4.14 5.62 5.86 4.30 4.74 3.93 5.26
t_d [ms]	2.50		1.13		0.35		0.67		t_d [ms]	0.22
z_0/ω_0	7				7				z_0/ω_0	3.5
x_{ab}	0.86		0.71		1.02	1	0.96	0.87	x_{ab}	1
t_{ab} [ms]	(3.00±1.01)*10 ⁻⁶		(2.97±0.61)*10 ⁻⁶		(1.48±0.03)*10 ⁻⁶		(3.00±0.35)*10 ⁻⁶		t_{ab} [ms]	3.00*10 ⁻⁶
x_{rc1}	0.15	0.34	0.26	0.32	0.31	0.16	1.00	0.78	x_{rc}	0.61 0.61 0.52 0.51 0.59 0.59 0.63 0.56
t_{rc1} [ms]	(2.45±0.52)*10 ⁻⁵				(1.24±0.03)*10 ⁻⁵				t_{rc} [ms]	1.0*10 ⁻⁵
x_{rc2}					0.07	0.07	0.06	0.06	f	0.88
t_{rc2} [ms]					(3.93±0.10)*10 ⁻⁴				x_{cab1}	0.02±4*10 ⁻⁴
									t_{cab1} [ms]	(2.60±0.10)*10 ⁻⁴
x_{b1}	0.07±0.002		0.07±0.02		0.00		0.12±0.003		x_{cab2}	0.10±0.001
t_{b1} [ms]	(1.55±0.42)*10 ⁻³				(1.51±0.01)*10 ⁻³				t_{cab2} [ms]	1.00±0.01
x_{b2}			0.06±0.02		0.61±0.001		0.01±2*10 ⁻⁴		x_{cab3}	0.88±0.01
t_{b2} [ms]	(6.86±2.18)*10 ⁻³				(2.35±0.02)*10 ⁻³				t_{cab3} [ms]	17.33±0.44
x_{b3}	0.46±0.004		0.49±0.004		0.20±0.001		0.26±0.001		x_{cab4}	
t_{b3} [ms]	0.125±0.001				0.010±3*10 ⁻⁴				t_{cab4} [ms]	
x_{b4}	0.47±0.004		0.37±0.007		0.20±0.002		0.62±0.004		x_{cab5}	
t_{b4} [ms]	0.897±0.01				0.132±0.001				t_{cab5} [ms]	
4.5 M										
b	1				1.01		1.0		b	1
N	33.18		27.02		5.77		30.1		N	4.02 3.68 4.21 6.12 3.53 4.96 3.89 4.53
t_d [ms]	0.57				0.43		0.60		t_d [ms]	0.25
z_0/ω_0	14.61				10.55		512		z_0/ω_0	3.5

UNFOLDING PATHWAY OF T4L

x_{ab}	0.83	0.74	0.99	0.78	1.03	0.89	x_{ab}	1
t_{ab} [ms]	$(3.31 \pm 1.17) * 10^{-6}$	$(2.75 \pm 0.60) * 10^{-6}$	$(1.90 \pm 0.07) * 10^{-6}$	$(2.75 \pm 0.32) * 10^{-6}$	$(2.75 \pm 0.32) * 10^{-6}$	$(2.75 \pm 0.32) * 10^{-6}$	t_{ab} [ms]	$3.00 * 10^{-6}$
x_{rc1}	0.18	0.17	0.13	0.19	0.44	0.22	x_{rc}	0.59
t_{rc1} [ms]	$(4.56 \pm 0.77) * 10^{-5}$	$(9.06 \pm 0.40) * 10^{-6}$	$(9.06 \pm 0.40) * 10^{-6}$	$(9.06 \pm 0.40) * 10^{-6}$	$(9.06 \pm 0.40) * 10^{-6}$	$(9.06 \pm 0.40) * 10^{-6}$	t_{rc} [ms]	$1.0 * 10^{-5}$
x_{rc2}			0.04	0.04	0.12	0.11	f	0.90
t_{rc2} [ms]			$(1.39 \pm 0.06) * 10^{-4}$	$(1.39 \pm 0.06) * 10^{-4}$	$(1.39 \pm 0.06) * 10^{-4}$	$(1.39 \pm 0.06) * 10^{-4}$	x_{cab1}	0.04 ± 0.006
x_{b1}		0.18 ± 0.03	0.19 ± 0.001	0.10 ± 0.002	0.19 ± 0.001	0.10 ± 0.002	t_{cab1} [ms]	$(6.76 \pm 1.50) * 10^{-5}$
t_{b1} [ms]	$(6.80 \pm 0.77) * 10^{-4}$	$(6.80 \pm 0.77) * 10^{-4}$	$(1.07 \pm 0.03) * 10^{-3}$	$(1.07 \pm 0.03) * 10^{-3}$	$(1.07 \pm 0.03) * 10^{-3}$	$(1.07 \pm 0.03) * 10^{-3}$	x_{cab2}	$0.09 \pm 7 * 10^{-4}$
x_{b2}	0.22 ± 0.007	0.14 ± 0.004	0.46 ± 0.001	0.17 ± 0.002	0.46 ± 0.001	0.17 ± 0.002	t_{cab2} [ms]	0.95 ± 0.008
t_{b2} [ms]	$(7.15 \pm 0.60) * 10^{-3}$	$(7.15 \pm 0.60) * 10^{-3}$	$(3.54 \pm 0.03) * 10^{-3}$	$(3.54 \pm 0.03) * 10^{-3}$	$(3.54 \pm 0.03) * 10^{-3}$	$(3.54 \pm 0.03) * 10^{-3}$	x_{cab3}	0.87 ± 0.07
x_{b3}	0.78 ± 0.01	0.69 ± 0.02	0.15 ± 0.002	0.17 ± 0.002	0.15 ± 0.002	0.17 ± 0.002	t_{cab3} [ms]	18.90 ± 2.15
t_{b3} [ms]	0.139 ± 0.003	0.139 ± 0.003	$0.013 \pm 5 * 10^{-4}$	$0.013 \pm 5 * 10^{-4}$	$0.013 \pm 5 * 10^{-4}$	$0.013 \pm 5 * 10^{-4}$	x_{cab4}	
x_{b4}			0.20 ± 0.003	0.57 ± 0.005	0.20 ± 0.003	0.57 ± 0.005	t_{cab4} [ms]	
t_{b4} [ms]			0.124 ± 0.002	0.124 ± 0.002	0.124 ± 0.002	0.124 ± 0.002	x_{cab5}	
							t_{cab5} [ms]	
5 M								
b	1	1	1.02	1.0	1.02	1.0	b	1
N	48.12	28.11	6.00	25.54	6.00	25.54	N	4.47
t_d [ms]	0.95	1.26	0.76	0.87	0.76	0.87	t_d [ms]	0.27
z_0/ω_0	7	7	7.54	11.82	7.54	11.82	z_0/ω_0	3.5
x_{ab}	0.80	0.80	0.97	0.79	1.11	0.89	x_{ab}	1
t_{ab} [ms]	$(7.23 \pm 2.12) * 10^{-6}$	$(2.81 \pm 0.70) * 10^{-6}$	$(1.67 \pm 0.10) * 10^{-6}$	$(2.38 \pm 0.41) * 10^{-6}$	$(1.67 \pm 0.10) * 10^{-6}$	$(2.38 \pm 0.41) * 10^{-6}$	t_{ab} [ms]	$3.00 * 10^{-6}$
x_{rc1}	0.73	1.00	0.40	0.18	1.33	0.73	x_{rc}	0.56
t_{rc1} [ms]	$(1.16 \pm 0.26) * 10^{-5}$	$(1.16 \pm 0.26) * 10^{-5}$	$(1.15 \pm 0.07) * 10^{-5}$	$(1.15 \pm 0.07) * 10^{-5}$	$(1.15 \pm 0.07) * 10^{-5}$	$(1.15 \pm 0.07) * 10^{-5}$	t_{rc} [ms]	$1.0 * 10^{-5}$
x_{rc2}			0.05	0.05	0.12	0.14	f	0.89
t_{rc2} [ms]			$(3.27 \pm 0.20) * 10^{-4}$	$(3.27 \pm 0.20) * 10^{-4}$	$(3.27 \pm 0.20) * 10^{-4}$	$(3.27 \pm 0.20) * 10^{-4}$	x_{cab1}	0.03 ± 0.006
x_{b1}		0.12 ± 0.004	0.32 ± 0.002	$0.05 \pm 4 * 10^{-4}$	0.32 ± 0.002	$0.05 \pm 4 * 10^{-4}$	t_{cab1} [ms]	$(1.00 \pm 0.30) * 10^{-4}$
t_{b1} [ms]	$(8.20 \pm 3.82) * 10^{-5}$	$(8.20 \pm 3.82) * 10^{-5}$	$(2.79 \pm 0.04) * 10^{-3}$	$(2.79 \pm 0.04) * 10^{-3}$	$(2.79 \pm 0.04) * 10^{-3}$	$(2.79 \pm 0.04) * 10^{-3}$	x_{cab2}	0.12 ± 0.001
x_{b2}	0.11 ± 0.003	0.17 ± 0.004	0.25 ± 0.001	0.20 ± 0.01	0.25 ± 0.001	0.20 ± 0.01	t_{cab2} [ms]	0.94 ± 0.01
t_{b2} [ms]	$(1.73 \pm 0.37) * 10^{-3}$	$(1.73 \pm 0.37) * 10^{-3}$	$(9.51 \pm 0.20) * 10^{-3}$	$(9.51 \pm 0.20) * 10^{-3}$	$(9.51 \pm 0.20) * 10^{-3}$	$(9.51 \pm 0.20) * 10^{-3}$	x_{cab3}	0.85 ± 0.08
x_{b3}	0.49 ± 0.007	0.35 ± 0.006	0.12 ± 0.002	0.45 ± 0.006	0.12 ± 0.002	0.45 ± 0.006	t_{cab3} [ms]	14.89 ± 2.31
t_{b3} [ms]	0.117 ± 0.002	0.117 ± 0.002	0.102 ± 0.002	0.102 ± 0.002	0.102 ± 0.002	0.102 ± 0.002	x_{cab4}	
x_{b4}	0.39 ± 0.02	0.37 ± 0.007	0.20 ± 0.002	0.30 ± 0.003	0.20 ± 0.002	0.30 ± 0.003	t_{cab4} [ms]	
t_{b4} [ms]	0.753 ± 0.007	0.753 ± 0.007	0.513 ± 0.001	0.513 ± 0.001	0.513 ± 0.001	0.513 ± 0.001	x_{cab5}	
							t_{cab5} [ms]	

UNFOLDING PATHWAY OF T4L

5.5 M																	
<i>b</i>	1				1				<i>b</i>	1							
<i>N</i>	23.28		13.54		8.63		14.32		<i>N</i>	3.54	3.78	3.96	5.03	3.78	3.72	3.20	4.74
<i>t_d</i> [ms]	0.89				0.65		0.89		<i>t_d</i> [ms]	0.31							
<i>z₀/ω₀</i>	9.37				5.40		9.88		<i>z₀/ω₀</i>	3.5							
<i>x_{ab}</i>	0.79		0.89		1.10	0.81	1.04	1.00	<i>x_{ab}</i>	1							
<i>t_{ab}</i> [ms]	(2.80±0.87)*10 ⁻⁶		(3.09±0.38)*10 ⁻⁶		(1.53±0.06)*10 ⁻⁶		(2.84±0.34)*10 ⁻⁶		<i>t_{ab}</i> [ms]	3.00*10 ⁻⁶							
<i>x_{rc1}</i>	0.06	0.07	0.07	0.07	0.36	0.16	1.17	0.69	<i>x_{rc}</i>	0.43	0.40	0.40	0.24	0.39	0.43	0.48	0.27
<i>t_{rc1}</i> [ms]	(3.78±0.82)*10 ⁻⁴				(1.09±0.04)*10 ⁻⁵				<i>t_{rc}</i> [ms]	1.0*10 ⁻⁵							
<i>x_{rc2}</i>					0.06	0.06	0.13	0.14	<i>f</i>	0.86							
<i>t_{rc2}</i> [ms]					(2.98±0.11)*10 ⁻⁶				<i>x_{cab1}</i>	0.04±0.009							
									<i>t_{cab1}</i> [ms]	(6.99±2.20)*10 ⁻⁵							
<i>x_{b1}</i>			0.31±0.006		0.46±0.003		0.08±0.005		<i>x_{cab2}</i>	0.13±0.001							
<i>t_{b1}</i> [ms]	(2.64±0.77)*10 ⁻⁵				(2.82±0.06)*10 ⁻³				<i>t_{cab2}</i> [ms]	0.99±0.01							
<i>x_{b2}</i>	0.04±0.003		0.07±0.007		0.22±0.001		0.21±0.001		<i>x_{cab3}</i>	0.82±0.10							
<i>t_{b2}</i> [ms]	(6.08±0.89)*10 ⁻³				0.010±1.4*10 ⁻⁴				<i>t_{cab3}</i> [ms]	13.25±2.18							
<i>x_{b3}</i>	0.51±0.013		0.32±0.006		0.12±0.002		0.40±0.004		<i>x_{cab4}</i>								
<i>t_{b3}</i> [ms]	0.118±0.003				0.092±0.002				<i>t_{cab4}</i> [ms]								
<i>x_{b4}</i>	0.45±0.016		0.30±0.007		0.20±0.001		0.31±0.002		<i>x_{cab5}</i>								
<i>t_{b4}</i> [ms]	0.638±0.018				0.379±0.001				<i>t_{cab5}</i> [ms]								
6 M																	
<i>b</i>	1				1.01		1.0		<i>b</i>	1							
<i>N</i>	4.48		21.6		30.91		19.10		<i>N</i>	1.55	.65	1.96	2.69	1.78	1.82	1.81	2.42
<i>t_d</i> [ms]	0.40		1.83		2.67		2.50		<i>t_d</i> [ms]	0.37							
<i>z₀/ω₀</i>	5.48		10.46		7.00		9.36		<i>z₀/ω₀</i>	3.5							
<i>x_{ab}</i>	0.93		0.78		1.02	0.73	0.99	0.92	<i>x_{ab}</i>	0.57	0.55	0.49	0.32	0.50	0.53	0.50	0.38
<i>t_{ab}</i> [ms]	(3.05±0.11)*10 ⁻⁶		(2.04±0.56)*10 ⁻⁶		(1.85±0.31)*10 ⁻⁶		(3.53±0.46)*10 ⁻⁶		<i>t_{ab}</i> [ms]	1*10 ⁻⁵							
<i>x_{rc1}</i>	0.06	0.06	0.13	0.12	0.73	1	4.35	1.54	<i>x_{rc}</i>	1							
<i>t_{rc1}</i> [ms]	(1.49±0.13)*10 ⁻⁵				(6.98±0.73)*10 ⁻⁶				<i>t_{rc}</i> [ms]	1.0*10 ⁻⁵							
<i>x_{rc2}</i>					0.07	0.06	0.13	0.13	<i>f</i>	0.77							
<i>t_{rc2}</i> [ms]					(4.00±0.42)*10 ⁻⁴				<i>x_{cab1}</i>	0.03±2.7*10 ⁻⁵							
									<i>t_{cab1}</i> [ms]	(6.07±1.50)*10 ⁻⁵							
<i>x_{b1}</i>	0.51±0.01		0.05±0.002		0.35±0.003		0.11±0.005		<i>x_{cab2}</i>	0.08±2*10 ⁻⁴							
<i>t_{b1}</i> [ms]	(2.52±0.08)*10 ⁻³				(9.23±0.22)*10 ⁻³				<i>t_{cab2}</i> [ms]	1.23±0.14							
<i>x_{b2}</i>	0.20±0.009				0.11±0.005		0.26±0.007		<i>x_{cab3}</i>	0.89±0.001							
<i>t_{b2}</i> [ms]	0.019±0.001				0.095±0.004				<i>t_{cab3}</i> [ms]	20.56±3.74							

UNFOLDING PATHWAY OF T4L

x_{b3}	0.28±0.004		0.38±0.005		0.33±0.004		0.39±0.007		x_{cab4}								
t_{b3} [ms]	0.127±0.002				0.507±0.011				t_{cab4} [ms]								
x_{b4}			0.57±0.006		0.21±0.005		0.24±0.015		x_{cab5}								
t_{b4} [ms]	0.816±0.007				2.06±0.002				t_{cab5} [ms]								
7.5 M urea																	
b	1				1.01		1.0		b	1							
N	89.1		25.5		37.1		26.1		N	4.30	5.17	6.10	5.61	4.55	4.58	4.40	5.09
t_d [ms]	2.1255		0.95418		2.22		2.84		t_d [ms]	0.28							
z_0/ω_0	7				25.4		2.45		z_0/ω_0	3.5							
x_{ab}	0.62321		0.72691		0.96	0.94	1.0	0.91	x_{ab}	0.53	0.44	0.36	0.42	0.49	0.51	0.51	0.47
t_{ab} [ms]	(1.00±1.07)*10 ⁻⁶		(2.10±0.22)*10 ⁻⁶		(1.14±0.12)*10 ⁻⁵		(6.17±0.64)*10 ⁻⁶		t_{ab} [ms]	1*10 ⁻⁵							
x_{rc1}	0.05	0.10	0.12	0.17	5.06	3.88	3.51	2.73	x_{rc}	1							
t_{rc1} [ms]	(1.02±0.11)*10 ⁻⁴		(6.53±0.69)*10 ⁻⁵		(6.82±0.36)*10 ⁻⁶				t_{rc} [ms]	3*10 ⁻⁶							
x_{rc2}					0.05	0.04	0.15	0.13	f	0.88							
t_{rc2} [ms]					(4.06±0.22)*10 ⁻⁴				x_{cab1}	0.03±0.005							
									t_{cab1} [ms]	(7.5±1.60)*10 ⁻⁵							
x_{b1}	0.64±0.01		0.43±0.008		0.18±0.003		0.24±0.015		x_{cab2}	0.14±0.001							
t_{b1} [ms]	0.849±0.014				2.42±0.002				t_{cab2} [ms]	0.830±0.007							
x_{b2}	0.31±0.008		0.43±0.005		0.32±0.007		0.37±0.009		x_{cab3}	0.82±0.07							
t_{b2} [ms]	0.126±0.002				0.584±0.002				t_{cab3} [ms]	10.7±1.19							
x_{b3}	0.05±0.009		0.14±0.002		0.11±0.013		0.27±0.007		x_{cab4}								
t_{b3} [ms]	(3.65±0.19)*10 ⁻³				0.105±0.004				t_{cab4} [ms]								
x_{b4}					0.38±0.007		0.12±0.008		x_{cab5}								
t_{b4} [ms]					0.011±4*10 ⁻⁴				t_{cab5} [ms]								

4.3.2.7. Additional filteredFCS results

Figure 4.3.11A shows the t_{rel} obtained in the different urea concentrations. Figure 4.3.11B-C show examples of species-selective autocorrelation curves. It is evident from the right shift of the diffusion term that also here the diffusion time t_{diff} increases in higher urea concentrations as expected (see also Figure 4.3.11E). Furthermore, it is directly visible from the shape of the $sACF$ that the slower relaxation times diminish while the fraction of fast dynamics increases. The amplitude of the $sACF$ curves show how the ratio between LF and HF species found is reversed at higher urea concentration (see also Figure 4.3.11B-C). In fact, the development of species fractions shows a typical two state behavior with MoT's of 4.7 M (HF) and 5.4 M (LF), i.e. the amount of HF species decreases much faster than the fraction of LF species is increasing (Figure 4.3.11D). The fraction of molecules showing dynamic behavior reduces from $\sim 80\%$ to 60% . How the HF and LF species are defined and the decays and filters are obtained is shown exemplary in Figure 4.3.11G-I.

As the determined t_{diff} is not only affected by the shape of the molecule, but also by the viscosity μ of the surrounding medium and the shape of the detection volume (\sim determined by the index of refraction η of the measurement buffer), we corrected the obtained fit results for t_{diff} for both parameters. Kawahara and Tanford (143) determined that the viscosity of a buffer with increasing concentrations of urea changes according to the eq. 4.3.29:

$$\frac{\mu(cM)}{\mu(0M)} = 1 + 0.0375 \cdot c + 0.00315 \cdot c^2 + 0.00031 \cdot c^3 \quad 4.3.29$$

with c being the respective urea concentration in [M]. Thus, the t_{diff} one should obtain without any further changes than in viscosity can be calculated as follows:

$$t_{diff}(xM) = t_{diff}(0M) \frac{\mu(cM)}{\mu(0M)} \quad 4.3.30$$

However, as also the shape of the detection volume changes, we measured as a reference for Rhodamine110 how its t_{diff} changes with increasing urea concentration and extracted the additional change next to the increase by viscosity (eq. 4.3.29). Finally, we compared our measured $t_{diff,meas}$ to the calculated $t_{diff,calc}$ expected, when no change in shape or size of T4L would occur (Figure 4.3.11E):

$$t_{diff,calc,xM} = -0.148 + 1.50 \cdot t_{diff,meas,0M} \cdot t_{diff,meas,xM} \quad 4.3.31$$

$$t_{diff,rel,xM} = \frac{t_{diff,meas,xM}}{t_{diff,calc,xM}}$$

$t_{diff,rel}$ displays a 2-state behavior with a MoT of 5.5 M urea (Figure 4.3.11E).

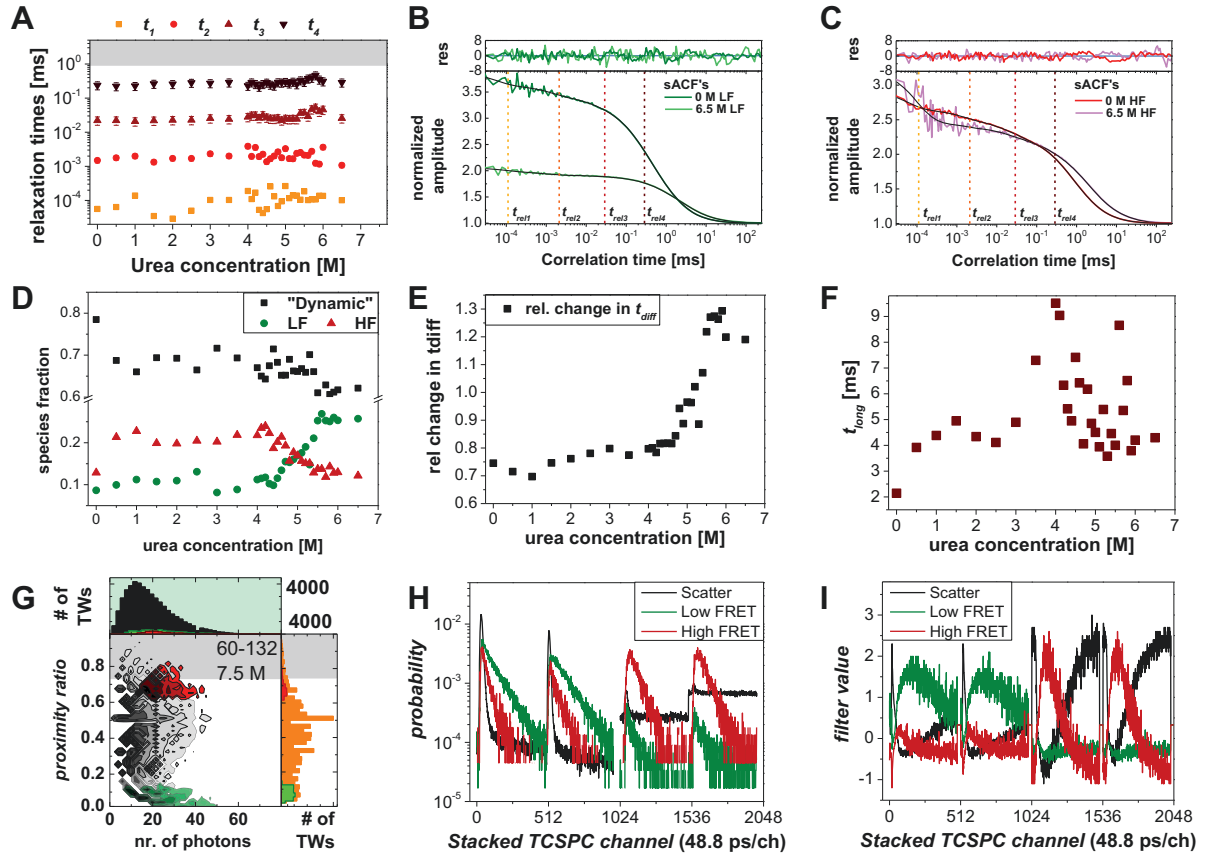
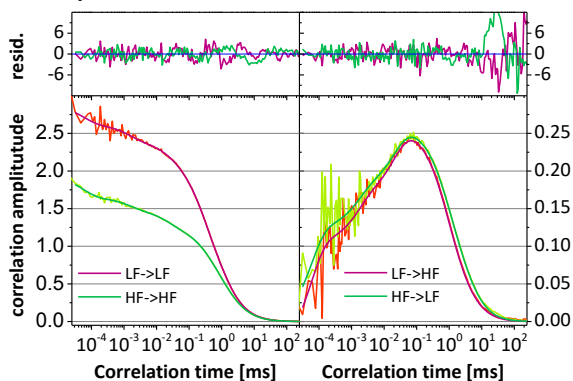


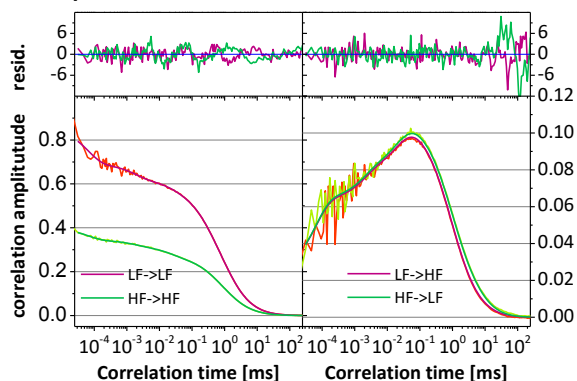
Figure 4.3.11 Exemplary filteredFCS results of variant K60pAcF N132C. (A) Relaxation times (t_{rel}) obtained in sCCF curves depending upon the urea concentration. Error bars indicate average standard error of the mean and gray shaded area times slower than the diffusion time. (B, C) Example of species-selective autocorrelation curves (sACF's) for the variant K60pAcF N132C-(DA) in 0 M and 6.5 M urea, filters were generated from single molecule bursts belonging to low FRET (LF) and high FRET (HF) displaying species (G). Residual for the global fit are shown on top. Vertical lines are placed at average relaxation times. (D) Species fraction of fFCS-fits, next to the LF and HF species used to generate the filter, the major fraction is of the molecules is dynamic. (E) Relative change in diffusion time corrected for changes in viscosity and index of refraction effects. (F) Fraction of long t_{rel} . (G) Time-window ID's of low- (green) and high- (red) FRET species were selected in a 2D frequency histogram of number of photons vs. proximity ratio. (H) Generated decays for the two pseudo-species low FRET and high FRET in addition to the scatter profile. (I) Filters were calculated according to Eq. 3.3.24 in Chapter 3 using the decays from graph (E).

The slow t_{rel} found (t_{long}) is much longer than t_{diff} and increases up ~ 10 ms in 4 M urea (Figure 4.3.11F), at higher urea concentration t_{long} is scattered between 4-8 ms. It might be that we catch here an additional slow unfolding process, however as the time scale of this process exceeds the t_{diff} by far, we cannot characterize it thoroughly.

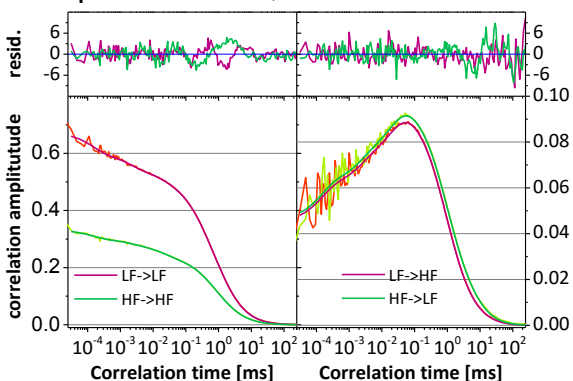
K60pAcF N132C, 0 M urea



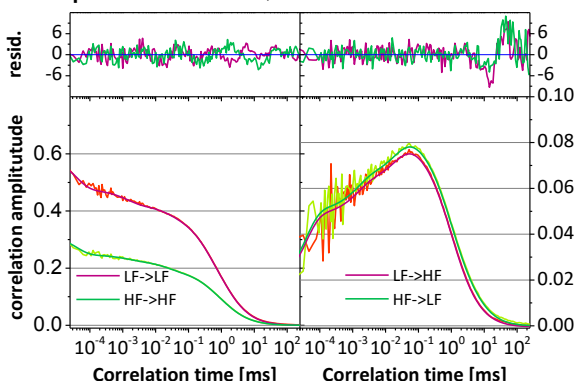
K60pAcF N132C, 0.5 M urea



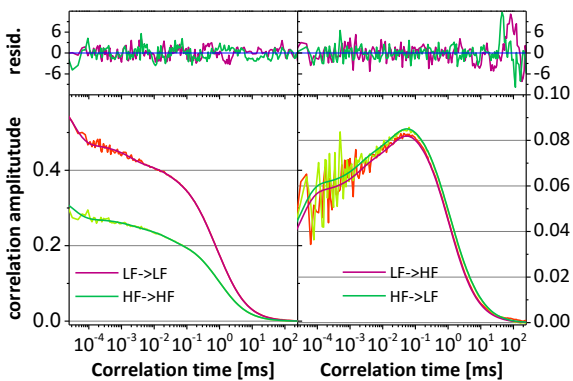
K60pAcF N132C, 1 M urea



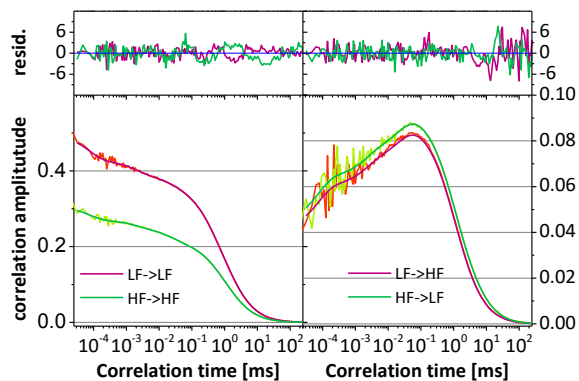
K60pAcF N132C, 1.5 M urea



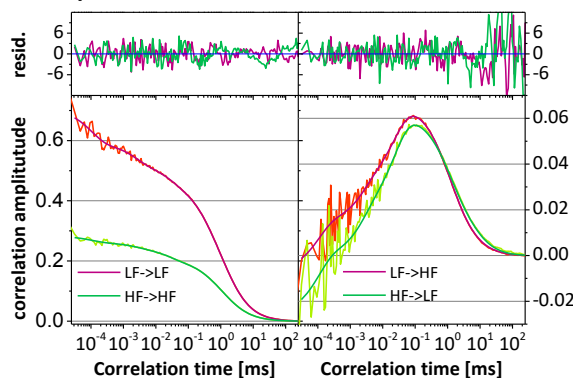
K60pAcF N132C, 2 M urea



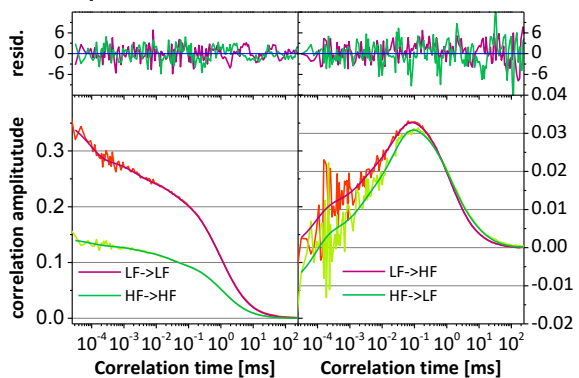
K60pAcF N132C, 2.5 M urea

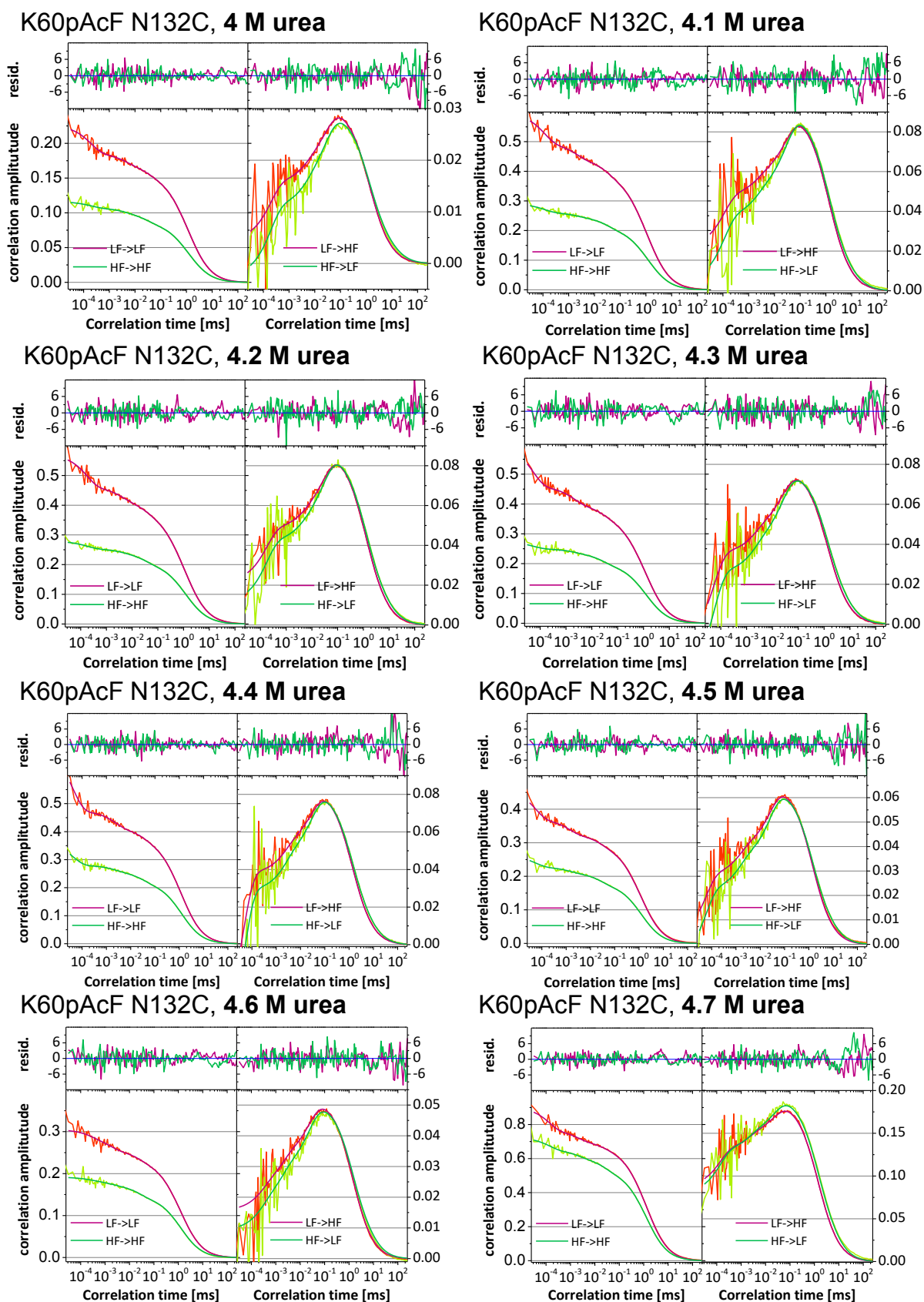


K60pAcF N132C, 3 M urea

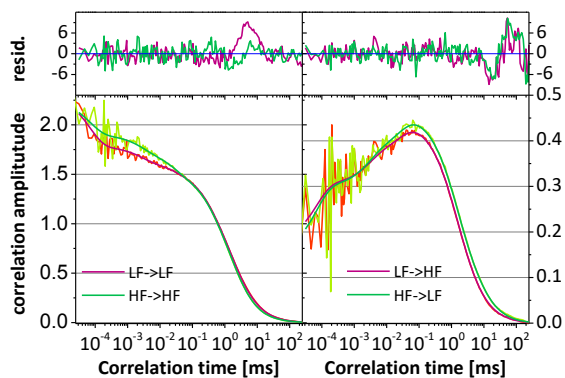


K60pAcF N132C, 3.5 M urea

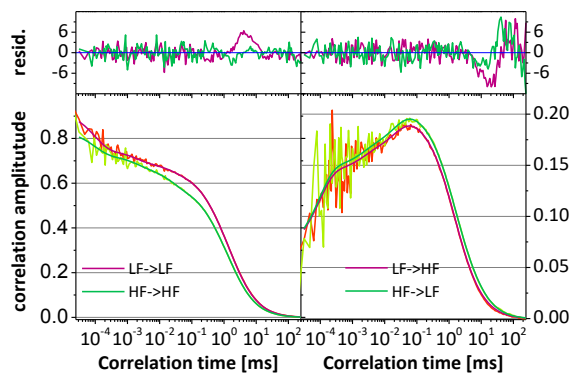




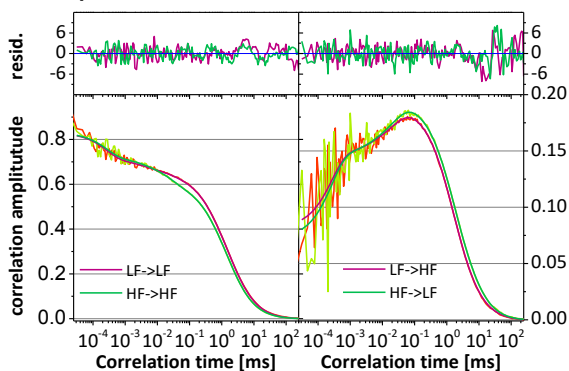
K60pAcF N132C, 4.8 M urea



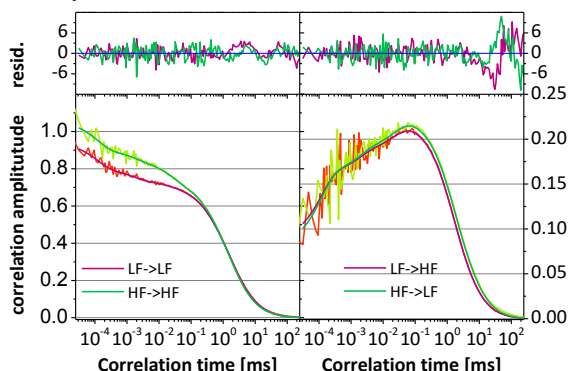
K60pAcF N132C, 4.9 M urea



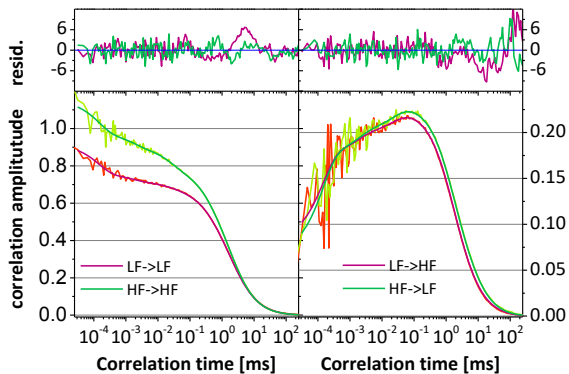
K60pAcF N132C, 5.0 M urea



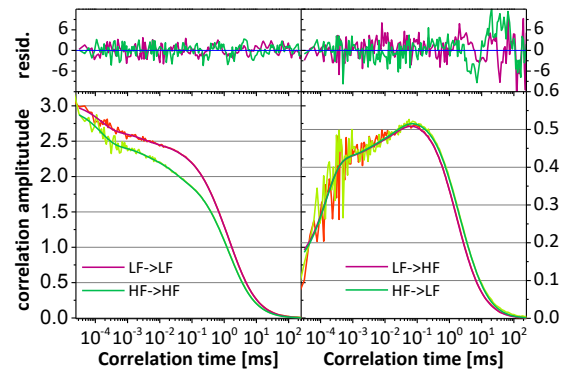
K60pAcF N132C, 5.1 M urea



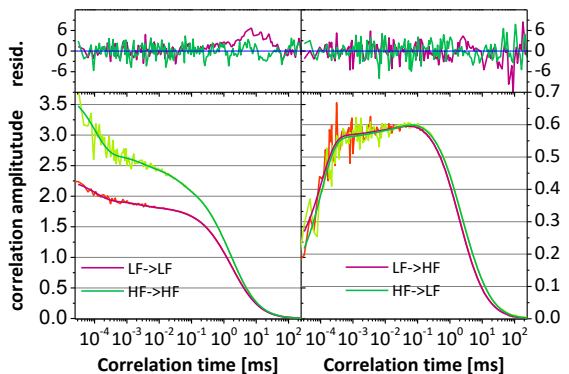
K60pAcF N132C, 5.2 M urea



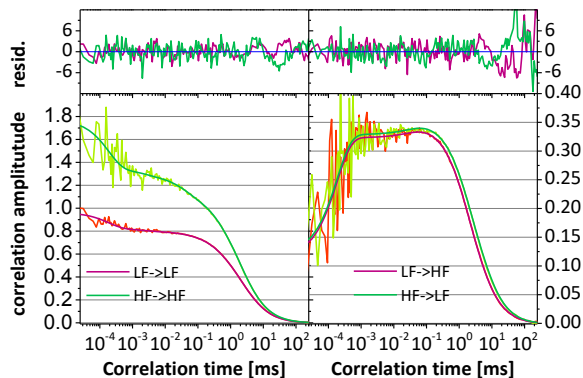
K60pAcF N132C, 5.3 M urea



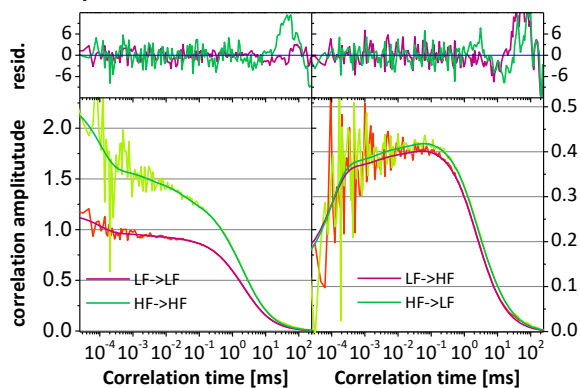
K60pAcF N132C, 5.4 M urea



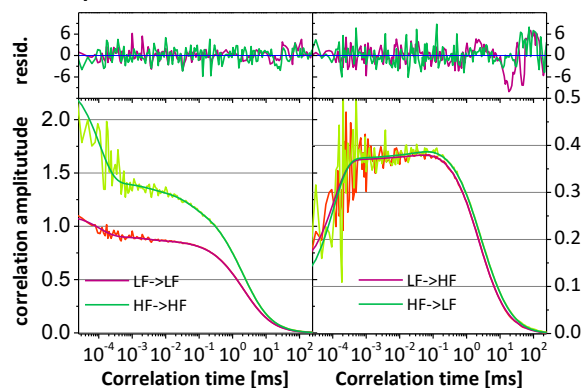
K60pAcF N132C, 5.5 M urea



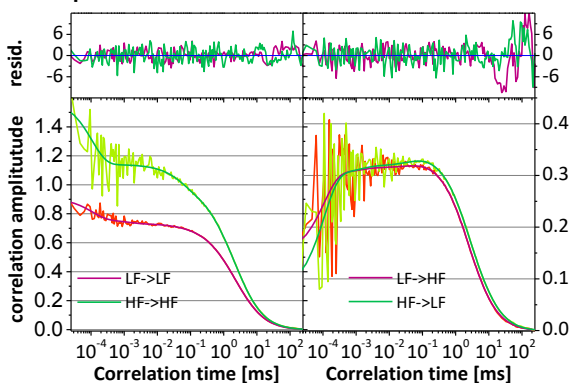
K60pAcF N132C, 5.6 M urea



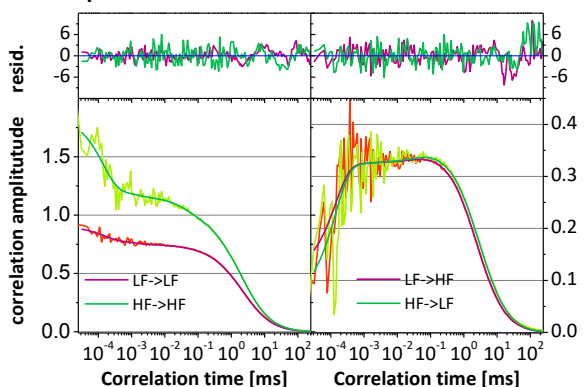
K60pAcF N132C, 5.7 M urea



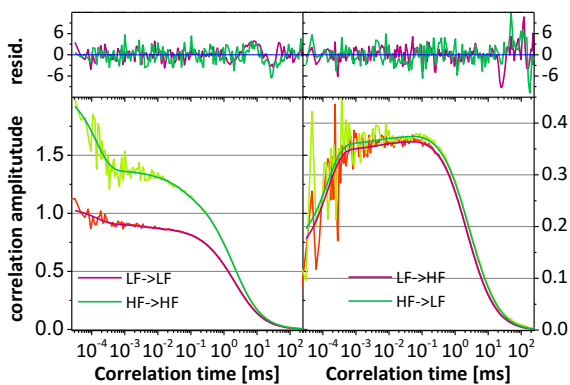
K60pAcF N132C, 5.8 M urea



K60pAcF N132C, 5.9 M urea



K60pAcF N132C, 6 M urea



K60pAcF N132C, 6.5 M urea

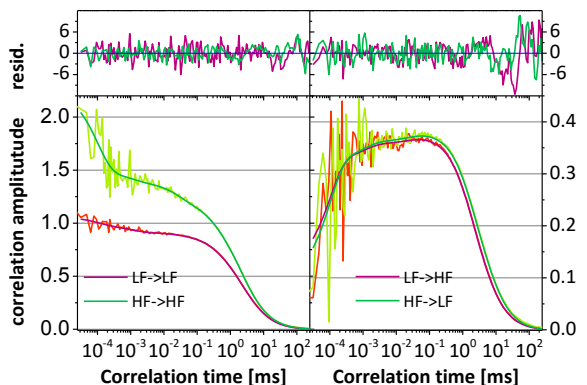


Figure 4.3.12 filteredFCS results of variant K60pAcF N132C. Fit results are summarized in Table 4.3.6.

UNFOLDING PATHWAY OF T4L

Table 4.3.6 Fit results for filteredFCS measurements of variant 60-132. Uncertainties given are standard errors of the mean. For a_4 , the average uncertainty of a_1 - a_3 are given (21.2 %).

[M urea]		b	N	t_d	z_0/ω_0	f	t_{rel1} [ms]	a_1	t_{rel2} [ms]	a_2	t_{rel3} [ms]	a_3	t_{rel4} [ms]	a_4	a_{long}	t_{long} [ms]
0	LF-HF	1.03	3.62	0.94 ± 0.05	4.95	1.17	(5.63 ± 2.07) E-05	0.46	(1.49 ± 0.18) E-03	0.18	(2.20 ± 1.06) E-02	0.30	0.237 ± 0.101	0.06	0.00	2.14 ± 0.09
	HF-LF	1.03	2.64			1.04		± 0.08		± 0.004		± 0.13		± 0.01	0.27	
	LF-LF	1.02	0.35					0.09		0.07		0.06		0.19	0.25	
	HF-HF	1.04	0.51					0.15		0.08		0.11		0.02	0.05	
0.5	LF-HF	1.00	9.40	0.92 ± 0.004	6.19	0.77	(6.40 ± 1.42) E-05	0.47	(1.77 ± 0.27) E-03	0.18	(2.20 ± 0.07) E-02	0.34	0.222 ± 0.014	0.01	0.00	3.92 ± 0.12
	HF-LF	1.00	6.67			0.78		± 0.06		± 0.03		± 0.04		± 0.002	0.28	
	LF-LF	1.00	1.16					0.19		0.08		0.05		0.08	0.06	
	HF-HF	1.00	2.49					0.14		0.06		0.11		0.03	0.01	
1	LF-HF	1.00	10.32	0.92 ± 0.005	7.24	0.54	(1.39 ± 0.23) E-04	0.30	(1.99 ± 0.27)E-03	0.21	(2.08 ± 0.10) E-02	0.46	0.227 ± 0.009	0.03	0.00	4.38 ± 0.14
	HF-LF	1.00	7.09			0.54		± 0.05		± 0.03		± 0.05		± 0.007	0.29	
	LF-LF	1.00	1.48					0.11		0.08		0.05		0.10	0.05	
	HF-HF	1.00	3.00					0.10		0.06		0.12		0.03	0.00	
1.5	LF-HF	1.00	12.31	1.01 ± 0.02	7.78	0.84	(3.49 ± 3.04) E-05	0.52	(1.33 ± 0.90) E-03	0.20	(2.19 ± 0.28) E-02	0.28	0.263 ± 0.027	0.00	0.00	4.95 ± 0.01
	HF-LF	1.00	8.93			0.84		± 0.16		± 0.08		± 0.06		0.00	0.25	
	LF-LF	1.00	1.64					0.22		0.09		0.05		0.09	0.05	
	HF-HF	1.00	3.04					0.25		0.06		0.10		0.04	0.01	
2	LF-HF	1.00	11.37	1.06 ± 0.01	9.83	0.77	(2.93 ± 0.55) E-05	0.55	(1.71 ± 0.54) E-03	0.17	(2.26 ± 0.07) E-02	0.28	0.258 ± 0.014	0.00	0.00	4.34 ± 0.23
	HF-LF	1.00	8.34			0.74		± 0.07		± 0.08		± 0.18		0.00	0.24	
	LF-LF	1.00	1.56					0.27		0.07		0.04		0.08	0.03	
	HF-HF	1.00	2.81					0.24		0.06		0.10		0.04	0.04	
2.5	LF-HF	1.00	11.14	1.12 ± 0.01	8.49	0.61	(5.01 ± 0.71) E-05	0.45	(1.70 ± 0.86) E-03	0.20	(2.27 ± 0.31) E-02	0.34	0.281 ± 0.029	0.01	0.00	4.11 ± 0.17
	HF-LF	1.00	8.49			0.60		± 0.21		± 0.06		± 0.08		± 0.002	0.20	
	LF-LF	1.00	1.93					0.16		0.08		0.05		0.09	0.06	
	HF-HF	1.00	3.02					0.18		0.06		0.11		0.02	0.03	
3	LF-HF	1.00	14.67	1.17 ± 0.002	10.36	1.10	(1.03 ± 0.15) E-04	0.28	(2.33 ± 0.28) E-03	0.22	(2.54 ± 0.08) E-02	0.48	0.273 ± 0.009	0.03	0.00	4.89 ± 0.24
	HF-LF	1.00	10.27			1.41		± 0.01		± 0.02		± 0.01		± 0.006	0.34	
	LF-LF	1.00	1.41					0.16		0.10		0.08		0.07	0.10	
	HF-HF	1.00	3.52					0.08		0.06		0.14		0.03	0.06	
3.5	LF-HF	1.00	27.01	1.18 ± 0.001	9.96	0.99	(1.03 ± 0.10) E-04	0.29	(2.00 ± 0.16) E-03	0.19	(2.41 ± 0.08) E-02	0.48	0.287 ± 0.002	0.04	0.00	7.30 ± 1.54
	HF-LF	1.00	18.03			1.27		± 0.01		± 0.02		± 0.05		± 0.008	0.37	
	LF-LF	1.00	2.87					0.18		0.09		0.08		0.07	0.05	
	HF-HF	1.00	7.09					0.07		0.06		0.14		0.04	0.06	

UNFOLDING PATHWAY OF T4L

4	LF-HF	1.00	31.59	1.25 ±0.004	7.05	0.86	(1.90 ±0.20) E-04	0.40 ±0.02	(3.88 ±0.71) E-03	0.12 ±0.02	(2.79 ±0.16) E-02	0.42 ±0.02	0.229 ±0.025	0.05 ±0.01	0.00 0.34	9.52 ±0.01
	HF-LF	1.00	21.34			1.08		0.18		0.06		0.07		0.05	0.00	
	LF-LF	1.00	4.42					0.08		0.05		0.13		0.03	0.01	
	HF-HF	1.00	8.60													
4.1	LF-HF	1.00	10.80	1.27 ±0.006	10.17	0.76	(1.09 ±0.15) E-04	0.36	(1.95 ±0.32) E-03	0.14	(3.06 ±0.12) E-02	0.45	0.288 ±0.050	0.04 ±0.009	0.00 0.23	9.04 ±0.03
	HF-LF	1.00	8.16			0.95		0.17		0.09		0.09		0.06	0.05	
	LF-LF	1.00	1.68					0.10		0.05		0.13		0.03	0.06	
	HF-HF	1.00	3.43													
4.2	LF-HF	1.00	10.79	1.25 ±0.009	9.57	0.76	(1.69 ±0.23) E-04	0.39 ±0.03	(3.57 ±0.63) E-03	0.10 ±0.02	(3.09 ±0.24) E-02	0.49 ±0.05	0.258 ±0.039	0.03 ±0.006	0.04 0.25	6.33 ±0.01
	HF-LF	1.00	8.39			0.87		0.19		0.07		0.07		0.06	0.02	
	LF-LF	1.00	1.75					0.09		0.04		0.13		0.03	0.02	
	HF-HF	1.00	3.57													
4.3	LF-HF	1.00	12.43	1.31 ±0.003	9.14	1.13	(5.31 ±0.70) E-05	0.51 ±0.01	(1.93 ±0.23) E-03	0.13 ±0.01	(2.60 ±0.15) E-02	0.32 ±0.02	0.215 ±0.022	0.04 ±0.009	0.00 0.23	5.42 ±0.80
	HF-LF	1.00	9.59			1.37		0.23		0.10		0.06		0.05	0.06	
	LF-LF	1.00	1.67					0.09		0.05		0.13		0.03	0.05	
	HF-HF	1.00	3.63													
4.4	LF-HF	1.00	11.77	1.32 ±0.003	10.93	1.60	(4.26 ±0.54) E-05	0.66 ±0.01	(2.26 ±0.29) E-03	0.11 ±0.003	(3.14 ±0.13) E-02	0.22 ±0.01	0.259 ±0.026	0.01 ±0.003	0.00 0.22	4.95 ±0.39
	HF-LF	1.00	9.24			1.91		0.31		0.08		0.05		0.05	0.04	
	LF-LF	1.00	1.45					0.23		0.06		0.10		0.02	0.06	
	HF-HF	1.00	2.75													
4.5	LF-HF	1.00	14.85	1.33 ±0.003	7.98	0.97	(5.53 ±0.52) E-05	0.40 ±0.01	(1.37 ±0.21) E-03	0.17 ±0.03	(2.69 ±0.11) E-02	0.39 ±0.01	0.284 ±0.022	0.04 ±0.009	0.00 0.27	7.41 ±0.36
	HF-LF	1.00	10.94			1.11		0.19		0.10		0.07		0.05	0.00	
	LF-LF	1.00	2.18					0.13		0.06		0.13		0.03	0.04	
	HF-HF	1.00	3.81													
4.6	LF-HF	1.00	18.36	1.34 ±0.009	6.91	0.71	(2.58 ±0.41) E-04	0.19 ±0.01	(1.67 ±0.30) E-03	0.25 ±0.05	(2.49 ±0.28) E-02	0.48 ±0.08	0.247 ±0.030	0.07 ±0.02	0.00 0.28	6.42 ±0.05
	HF-LF	1.00	13.46			0.83		0.06		0.10		0.08		0.05	0.00	
	LF-LF	1.00	3.28					0.04		0.05		0.15		0.03	0.03	
	HF-HF	1.00	5.22													
4.7	LF-HF	1.00	5.29	1.40 ±0.007	10.90	0.54	(1.32 ±0.26) E-04	0.42 ±0.04	(2.01 ±0.37) E-03	0.18 ±0.05	(2.47 ±0.34) E-02	0.37 ±0.06	0.232 ±0.061	0.03 ±0.006	0.00 0.21	4.06 ±0.16
	HF-LF	1.00	4.04			0.59		0.15		0.07		0.05		0.05	0.00	
	LF-LF	1.00	1.11					0.10		0.06		0.10		0.04	0.07	
	HF-HF	1.01	1.38													
4.8	LF-HF	1.00	2.25	1.57 ±0.009	51.54	0.60	(6.85 ±2.34) E-05	0.47 ±0.04	(2.65 ±0.56) E-03	0.23 ±0.04	(2.57 ±0.24) E-02	0.29 ±0.06	0.290 ±0.013	0.01 ±0.002	0.00 0.24	6.18 ±0.31
	HF-LF	1.00	1.65			0.67		0.23		0.07		0.04		0.07	0.00	
	LF-LF	1.00	0.43					0.16		0.07		0.09		0.06	0.19	
	HF-HF	1.00	0.44													

UNFOLDING PATHWAY OF T4L

4.9	LF-HF	1.00	5.01	1.49 ±0.01	10.45	0.68	(9.40± 3.21) E-05	0.58 ±0.05	(1.79 ±0.49) E-03	0.13 ±0.06	(2.22 ±0.16) E-02	0.28 ±0.06	0.284 ±0.018	0.01 ±0.002	0.00 0.20	4.85 ±0.25
	HF-LF	1.00	3.90			0.69										
	LF-LF	1.00	1.08													
	HF-HF	1.00	1.19													
5	LF-HF	1.00	5.24	1.63 ±0.008	7.40	0.57	(2.64 ±0.47) E-04	0.60 ±0.05	(3.30 ±0.78) E-03	0.07 ±0.05	(2.47 ±0.23) E-02	0.32 ±0.04	0.297 ±0.021	0.01 ±0.003	0.00 0.19	4.50 ±0.14
	HF-LF	1.00	4.13			0.63										
	LF-LF	1.00	1.20													
	HF-HF	1.00	1.20													
5.1	LF-HF	1.00	4.48	1.64 ±0.01	7.84	0.60	(1.16 ±0.26) E-04	0.55 ±0.05	(1.69 ±0.68) E-03	0.17 ±0.08	(2.41 ±0.21) E-02	0.25 ±0.04	0.269 ±0.019	0.03 ±0.006	0.00 0.18	3.94 ±0.22
	HF-LF	1.00	3.59			0.64										
	LF-LF	1.00	1.06													
	HF-HF	1.00	0.94													
5.2	LF-HF	1.00	4.27	1.75 ±0.01	9.06	0.65	(1.39 ±0.38) E-04	0.61 ±0.05	(1.66 ±0.43) E-03	0.14 ±0.04	(2.46 ±0.14) E-02	0.17 ±0.02	0.302 ±0.019	0.07 ±0.01	0.00 0.20	5.39 ±0.20
	HF-LF	1.00	3.31			0.72										
	LF-LF	1.00	1.09													
	HF-HF	1.00	0.87													
5.3	LF-HF	1.00	1.76	1.53 ±0.03	5.97	0.79	(1.39 ±0.14) E-04	0.66 ±0.03	(2.40 ±0.13) E-03	0.08 ±0.005	(2.42 ±0.19) E-02	0.17 ±0.01	0.254 ±0.008	0.09 ±0.02	0.00 0.23	3.58 ±0.20
	HF-LF	1.00	1.35			0.81										
	LF-LF	1.00	0.33													
	HF-HF	1.00	0.33													
5.4	LF-HF	1.00	1.56	1.87 ±0.05	5.23	0.74	(1.05 ±0.14) E-04	0.84 ±0.04	(2.55 ±0.47) E-03	0.03 ±0.008	(3.61 ±0.26) E-02	0.09 ±0.007	0.331 ±0.012	0.05 ±0.01	0.00 0.23	4.46 ±0.12
	HF-LF	1.00	1.18			0.85										
	LF-LF	1.00	0.44													
	HF-HF	1.00	0.27													
5.5	LF-HF	1.00	2.80	2.14 ±0.02	7.67	0.68	(1.84 ±1.84) E-04	0.86 ±0.24	(2.13 ±0.51) E-03	0.00 ±0.001	(3.67 ±0.77) E-02	0.09 ±0.008	0.298 ±0.032	0.05 ±0.01	0.00 0.20	4.00 ±0.21
	HF-LF	1.00	2.20			0.67										
	LF-LF	1.00	1.04													
	HF-HF	1.00	0.56													
5.6	LF-HF	1.00	2.17	2.25 ±0.01	42.78	0.65	(1.02 ±0.46) E-04	0.76 ±0.27	(2.10 ±1.34) E-03	0.09 ±0.06	(2.90 ±0.56) E-02	0.09 ±0.03	0.374 ±0.021	0.06 ±0.01	0.07 0.27	8.66 ±0.76
	HF-LF	1.00	1.63			0.70										
	LF-LF	1.00	0.86													
	HF-HF	1.00	0.44													
5.7	LF-HF	1.00	2.48	2.28 ±0.01	6.78	0.69	(1.07 ±0.27) E-04	0.88 ±0.04	(2.75 ±1.06) E-03	0.01 ±0.004	(4.69 ±0.80) E-02	0.08 ±0.03	0.394 ±0.032	0.03 ±0.007	0.00 0.18	5.35 ±0.23
	HF-LF	1.00	1.98			0.80										
	LF-LF	1.00	0.90													
	HF-HF	1.00	0.42													

UNFOLDING PATHWAY OF T4L

5.8	LF-HF	1.00	2.95	2.42 ±0.02	6.46	0.57	(1.07 ±0.60) E-04	0.82 ±0.13 0.16 0.29	(1.23 ±0.94) E-03	0.06 ±0.04 0.03 0.00	(5.23 ±1.01) E-02	0.08 ±0.03 0.02 0.09	0.464 ±0.029	0.04 ±0.009 0.05 0.05	0.00 0.17 0.00 0.11	6.51 ±0.40
	HF-LF	1.00	2.36			0.80										
	LF-LF	1.00	1.10													
	HF-HF	1.00	0.63													
5.9	LF-HF	1.00	2.83	2.35 ±0.03	6.57	0.67	(1.38 ±0.32) E-04	0.88 ±0.03 0.13 0.34	(1.16 ±0.63) E-03	0.02 ±0.01 0.04 0.03	(4.01 ±3.6) E-02	0.07 ±0.06 0.02 0.07	0.349 ±0.083	0.03 ±0.007 0.05 0.06	0.01 0.15 0.00 0.07	3.79 ±0.16
	HF-LF	1.00	2.37			0.82										
	LF-LF	1.00	1.10													
	HF-HF	1.00	0.54													
6	LF-HF	1.00	2.56	2.20 ±0.02	7.60	0.66	(1.37 ±0.38) E-04	0.84 ±0.05 0.14 0.34	(3.63 ±1.05) E-03	0.04 ±0.01 0.03 0.01	(4.55 ±1.73) E-02	0.07 ±0.02 0.01 0.07	0.310 ±0.054	0.06 ±0.01 0.05 0.05	0.00 0.18 0.00 0.10	4.20 ±0.18
	HF-LF	1.00	2.06			0.61										
	LF-LF	1.00	0.95													
	HF-HF	1.00	0.48													
6.5	LF-HF	1.00	2.54	2.28 ±0.02	6.99	0.69	(1.01 ±0.34) E-04	0.75 ±0.04 0.07 0.34	(1.07 ±0.42) E-03	0.11 ±0.06 0.07 0.04	(2.56 ±0.63) E-02	0.06 ±0.02 0.02 0.07	0.286 ±0.021	0.08 ±0.02 0.05 0.06	0.00 0.19 0.00 0.09	4.30 ±0.17
	HF-LF	1.00	2.01			0.76										
	LF-LF	1.00	0.94													
	HF-HF	1.00	0.44													

4.3.2.8. Results of Two-state fitting

Table 4.3.7 Results of fitting the different observables with a two-state fit. Uncertainties given are standard errors of the mean.

Method [Signal unit]	y_N	m_N [1/M]	y_D	m_D [1/M]	ΔG^0 [kcal/M]	m [kcal/M ²]	MoT [M]
R8pAcF P86C							
CD [mdeg]	-10391 ± 214 (-7860 ± 162)	348±119 (859 ±294)	-4391 ± 257	455±35	10.4±2.3 (29.7 ± 6.5)	-2.82±0.6 (-4.99 ±1.10)	3.7 (6.0)
$\langle\tau\rangle_f$ [ns]	2.71±0.03	-0.019 ±0.01	3.01 ±0.11	0.017 ±0.02	10.73 ±2.41	-2.62±0.59	4.1
Gaussian fit, $\langle R_{DA} \rangle$ [Å]	42.3±2.8	0.043 ±0.15	42.2 ±11.3	3.65 ±2.0	471±19.4	-102±4.22	4.6
Gaussian fit, w_{RDA} [Å]	9.24±2.89	0.30 ±1.42	48.6 ±9.58	-2.16 ±1.51	15.5±6.60	-3.71±1.60	4.2
WLC fit, L [Å]	66.0±17.6	0.15 ±3.60	74.2 ±32.0	6.30 ±5.49	41.0±20.2	-11.9±5.85	3.5
Residual anisotropy $r_{ss,D0}$	0.18 ±0.003	-0.0072 ±0.003	0.074 ±0.04	0.0040±0 .01	5.53±1.51	-1.12±0.33	4.9
K19pAcF P86C							
CD [mdeg]	-9834 ±243	1.46 ±35.8	-3746 ±1379	402 ±213	11.5 ±1.69	-2.52 ±0.37	4.6
$\langle\tau\rangle_f$ [ns]	2.23±0.04	0.0040 ±0.02	2.78 ±0.19	0.013±0. 03	25.7±10.5	-5.46±2.27	4.7
Gaussian fit, $\langle R_{DA} \rangle$ [Å]	49.7±10.3	0.40 ±1.86	548 ±5543	-62.8 ±710	6.40±4.81	-0.82±1.45	(7.8)
Gaussian fit, w_{RDA} [Å]	9.39±2.93	-0.40 ±2.02	64.7 ±12.8	-5.96 ±4.17	7.47±3.17	-1.40±0.60	5.3
WLC fit, L [Å]	69.2±1.81	-0.39 ±1.90	102 ±8.99	0.79 ±0.88	15.8±0.15	-3.24±0.03	4.9
WLC fit, κ [-]	0.91±0.04	0.02 ±0.02	0.34 ±0.13	0.04 ±0.05	12.8±0.31	-2.74±0.07	4.7
static PDA, $\langle R_{DA} \rangle_E$ [Å]	51.5±0.66	0.83 ±0.19	64.5 ±4.69	-0.32 ±0.69	298±0.43	-54.3±0.08	5.5
S44pAcF R119C							
CD [mdeg]	-10350 ±97.1	62.9 ±26.4	-2193 ±548	169 ±62.0	10.4±0.66	-2.06±0.13	5.0
$\langle\tau\rangle_f$ [ns]	2.95±0.02	-0.019 ±0.01	3.30 ±0.10	0.0071±0 .01	13.15 ±2.51	-2.71±0.51	4.9
Gaussian fit, $\langle R_{DA} \rangle$ [Å]	52.9±1.84	-0.37 ±1.10	457 ±604	-51.0 ±77.6	5.95±1.33	-0.84±0.33	(7.1)
Gaussian fit, w_{RDA} [Å]	9.15±2.66	0.66 ±0.84	58.1 ±12.5	-4.58 ±1.92	145±0.92	-28.7±0.18	5.0
WLC fit, L [Å]	72.7±1.97	-0.12 ±0.76	151 ±21.7	-8.26 ±3.07	9.43±2.73	-1.80±0.54	5.2
WLC fit, κ [-]	0.88±0.06	0.005 ±0.02	0.56 ±0.24	0.032 ±0.05	28.2±316	-6.02±69.2	4.7
Residual anisotropy $r_{ss,D0}$	0.14±0.13	0.0039 ±0.005	-0.087 ±0.18	0.02 ±0.04	1.88±4.27	-0.34±0.77	5.4
static PDA, $\langle R_{DA} \rangle_E$ [Å]	52.9±0.62	0.98 ±0.20	59.5 ±3.87	0.58 ±0.81	195±6.79	-37.8±1.31	5.2
static PDA, hw_{DA} [Å]	3.57±0.34	0.47 ±0.11	9.24 ±1.58	-0.20 ±0.24	135±0.75	-26.8±0.15	5.0

UNFOLDING PATHWAY OF T4L

S44pAcF I150C							
CD [mdeg]	-9548 ±70.9	0.81 ±18.2	-964 ±526	-4.78 ±116	9.97±0.38	-2.06±0.08	4.8
$\langle\tau\rangle_f$ [ns]	2.55±0.06	0.016 ±0.03	2.66 ±0.31	0.087 ±0.05	18.2±11.2	-3.84±2.37	4.7
Gaussian fit, $\langle R_{DA} \rangle$ [Å]	51.7±1.91	0.42 ±0.80	64.3 ±38.3	0.58 ±5.32	7.66±6.43	-1.41±1.28	5.4
Gaussian fit, w_{RDA} [Å]	9.65±1.38	0.10 ±0.55	49.3 ±12.0	-4.14 ±1.73	12.2±4.92	-2.36±0.97	5.2
WLC fit, L [Å]	71.9±8.54	0.05 ±2.89	134 ±66.2	-3.43 ±9.66	11.2±0.84	-2.16±0.16	5.2
WLC fit, κ [-]	0.88±0.06	0.0022 ±0.01	0.26 ±0.42	0.0057 ±0.05	87.1 ±129.3	-17.4±25.9	5.0
Residual anisotropy $r_{ss,D0}$	0.14±0.25	0.0039 ±0.05	-0.087 ±1.08	0.020 ±0.38	1.88±0.07	-0.34±0.01	5.4
N55pAcF R119C							
CD [mdeg]	-13251 ±68.1	100 ±27.7	-2406 ±489	154 ±72.9	10.1±0.29	-2.11±0.06	4.8
$\langle\tau\rangle_f$ [ns]	3.18±0.02	-0.0086 ±0.01	3.36 ±0.05	-0.0014 ±0.01	19.8±10.4	-4.82±2.58	4.1
Gaussian fit, $\langle R_{DA} \rangle$ [Å]	32.3±15.0	-7.05 ±12.7	358 ±5435	-26.6 ±420	1.38±4.99	-0.19±0.69	(7.1)
Gaussian fit, w_{RDA} [Å]	12.0±2.28	0.59 ±1.09	24.1 ±5.44	1.37 ±0.89	187±0.75	-46.6±0.19	4.0
WLC fit, L [Å]	80.4±4.16	0.36 ±4.95	119 ±15.6	-2.01 ±2.51	484±13.2	-108±2.93	4.5
WLC fit, κ [-]	0.95±0.03	0.0051 ±0.01	0.61 ±0.10	0.006 ±0.006	142±3.18	-32.1±0.72	4.4
Residual anisotropy $r_{ss,D0}$	0.14 ±0.003	-0.0064 ±0.001	0.043 ±0.02	0.0051 ±0.003	14.6 ±0.19	-2.88±0.04	5.1
K60pAcF R119C							
CD [mdeg]	-8725 ±79.4	57.8 ±31.4	-1723 ±644	135 ±97.3	10.7±0.55	-2.19±0.11	4.9
$\langle\tau\rangle_f$ [ns]	2.40±0.01	0.024 ±0.01	2.96 ±0.04	0.028 ±0.01	9.89±0.79	-2.33±0.19	4.2
Gaussian fit, $\langle R_{DA} \rangle$ [Å]	49.7±0.58	0.86 ±0.22	50.6 ±2.76	1.89 ±0.42	19.7±9.68	-4.21±2.11	4.7
Gaussian fit, w_{RDA} [Å]	9.04±0.48	0.34 ±0.25	35.1 ±2.97	-1.83 ±0.43	8.06±1.16	-1.76±0.26	4.6
WLC fit, L [Å]	71.7±0.77	0.56 ±0.35	124 ±3.71	-0.98 ±0.57	11.2±1.03	-2.48±0.23	4.5
WLC fit, κ [-]	0.76±0.01	0.02 ±0.004	0.36 ±0.03	0.03 ±0.004	10.9±1.08	-2.65±0.27	4.1
Residual anisotropy $r_{ss,D0}$	0.18 ±0.002	-0.0035 ±0.0009	0.10 ±0.01	-0.0007 ±0.001	19.8±15.4	-4.44±3.42	4.5
K60pAcF N132C							
CD [mdeg]	-7286 ±51.0	-29.8 ±19.4	-1344 ±532	94.4 ±76.4	10.1±0.38	-1.96±0.08	5.1
$\langle\tau\rangle_f$ [ns]	2.03±0.02	0.035 ±0.009	2.92 ±0.11	0.044 ±0.016	10.3±1.23	-2.36±0.28	4.4
$\langle\tau\rangle_{f,MFD}$ [ns]	2.04±0.03	0.017 ±0.017	2.67 ±0.21	0.020 ±0.032	10.1±2.87	-2.25±0.64	4.5
Gaussian fit, $\langle R_{DA} \rangle$ [Å]	45.8±0.53	0.942 ±0.25	61.5 ±5.10	1.94 ±0.75	10.3±1.80	-2.16±0.38	4.8
Gaussian fit, w_{RDA} [Å]	8.54±0.39	-0.14 ±0.18	18.3 ±2.98	0.40 ±0.45	11.8±2.96	-2.57±0.65	4.6
WLC fit, L [Å]	66.7±2.40	0.58 ±1.06	71.6 ±15.0	4.83 ±2.30	11.6±0.51	-2.56±0.11	4.5

UNFOLDING PATHWAY OF T4L

WLC fit, l_p [Å]	45.0±1.34	2.91 ±0.78	27.4 ±4.02	6.49 ±0.67	11.0±1.40	-3.01±0.38	3.7
Residual anisotropy $r_{ss,D0}$	0.18± 0.002	-0.0029 ±0.0009	0.11 ±0.01	-0.0029 ±0.002	9.86±2.10	-2.18±0.47	4.5
static PDA, $\langle R_{DA} \rangle_E$ [Å]	45.5±0.26	0.65 ±0.14	51.3 ±2.18	1.14 ±0.32	7.92±1.38	-1.72±0.31	4.6
static PDA, hw_{DA} [Å]	2.89±0.13	-0.072 ±0.061	4.54 ±0.46	0.097 ±0.078	19.4±5.88	-4.62±0.06	4.2
fFCS, amplitude LF species [-]	0.11±0.01	0.00027 ±0.02	3.45 ±3.49	-0.38 ±0.59	8.95±2.78	-1.65±1.63	5.4
fFCS, amplitude HF species [-]	0.70±0.04	0.036 ±0.01	0.89 ±0.73	-0.13 ±0.30	14.1±7.47	-2.98±1.63	4.7
fFCS, amplitude t_{rel1} [-]	0.38±0.15	-0.023 ±0.04	3.02 ±0.71	-0.34 ±0.05	12.9±15.1	-2.44±2.76	5.3
fFCS, amplitude t_{rel3} [-]	0.60±0.07	0.069 ±0.04	-1.10 ±0.89	0.17 ±0.12	9.60±11.5	-1.87±2.25	5.1
fFCS, t_{diff} [ms]	0.02 ±0.0008	0.04 ±0.01	3.63 ±1.34	-0.44 ±0.29	15.1±4.14	-2.76±0.80	5.5
Q69pAcF N132C							
CD [mdeg]	-8732 ±0.01	57.4 ±0.002	-3027 ±0.05	305 ±0.01	10.9 ±4E-5	-2.15 ±9E-6	5.0
$\langle \tau \rangle_f$ [ns]	2.92±0.02	-0.053 ±0.02	2.99 ±0.10	0.044 ±0.02	6.45±2.56	-1.71±0.64	3.8
Gaussian fit, $\langle R_{DA} \rangle$ [Å]	39.7±0.75	1.95 ±0.24	51.6 ±5.21	1.63 ±0.77	44.6 ±171	-8.71±34.2	5.1
Gaussian fit, w_{RDA} [Å]	12.7±2.20	1.33 ±7.97	9.62 ±3.04	-0.13 ±0.07	-157±4.49	39.3±1.12	4.0
Residual anisotropy $r_{ss,D0}$	0.76±0.02	0.090 ±0.02	0.11 ±0.01	-0.0023 ±0.0009	-1.53 ±0.12	-0.24±0.02	(6.3)

*This variant displays a three state behavior; values in brackets denote the values for the intermediate state.

5. The denatured state of T4L

5.1. Fluorescence spectroscopic toolkit resolves the heterogeneity of denatured states

The manuscript is based on collaborative work in a joint project of the following persons: Katherina Hemmen, Dmitro Rodnin, Hugo Sanabria, Igor Markovic, Thomas Peulen, Suren Felekyan, Ralf Kühnemuth and Claus A.M. Seidel.

The contributions to this manuscript are as follows: IM conducted experiments under supervision of **KH** and DmR during his 10 weeks BioSoft Summer school. DmR measured a part of MFD and filteredFCS. TP programmed the polymer models and did the FPS. SF helped in analyzing filteredFCS data. RK prepared the setup for FCS. **KH** purified and labeled the protein and conducted eTCSPC and FCS measurements and missing MFD and filteredFCS. **KH** analyzed the data. HS gave support on data analysis. HS and CAMS designed the study. **KH**, HS and CAMS wrote the manuscript.

KH: 50 %

DmR: 12 %

IM: 5 %

TP: 7 %

SF: 6 %

RK: 6 %

HS: 7 %

CAMS: 7 %

5.2. Manuscript

Abstract

About 30 % of human proteins do not fold into a stable 3D arrangement of secondary structure elements, but stay predominantly unfolded -similar to proteins under highly denaturing conditions. These proteins are involved in many cell signaling processes. Their characterization poses a great challenge for current experimental methods as they consist of an ensemble of rapidly interconverting conformations. Intense debate exists on the possibility that they show, to certain extent, residual structure, which might facilitate folding or enhance ligand binding. To study the unfolded state conformational heterogeneity using Förster resonance energy transfer (FRET), we used the lysozyme from the phage T4 (T4L) in denaturing conditions as a model system. We built an elastic network model that spans T4L's topology in order to evaluate local and global conformational changes by combining ensemble (ensemble time-resolved fluorescence lifetime and anisotropy) and single-molecule spectroscopic (multiparameter fluorescence detection, photon distribution analysis, (filtered) fluorescence correlation spectroscopy) methods. Through extensive comparison of models, we identified regions with apparent residual structure under highly denaturing conditions, which might serve as folding nuclei; and additionally we showed that chemically denatured T4L is not a random coil as previously thought. By using obtained distance restraints we determined that denatured T4L shows a native-like mean structure, albeit larger in size compared to the native state. We demonstrate here the necessity of careful data interpretation, but also the potential of a multidimensional approach to characterize an ensemble of states, which can be applied generally to unstructured or denatured proteins.

Significance Statement

About 30 % of human proteins do not fold into a stable three-dimensional arrangement of secondary structure elements, but stay predominantly unfolded -similar to proteins under highly denaturing conditions. These greatly dynamic and flexible proteins involved in many cell signaling processes are referred to as intrinsically disordered proteins (IDP's). Their flexibility poses a great challenge for current experimental methods of structure characterization. Here, we present a combinational approach of fluorescence spectroscopic tools that can resolve the conformational heterogeneity of denatured states of proteins and IDPs over eight orders of magnitude in time– from ps to ms. The distribution of states is probed with ensemble time-correlated single photon counting, whereas the local and global

dynamics are probed with fluorescence anisotropy, photon distribution analysis and (filtered) fluorescence correlation spectroscopy.

5.2.1. Introduction

About 30 % of the human proteins (144) do not adopt the stable 3-dimensional fold, which was long thought of being necessary for the successfully functioning of proteins. These proteins, which fulfill important tasks within e.g. the cell signaling or cell cycle, either lack folded sections completely (referred to as intrinsically disordered proteins, IDPs) or contain disordered regions (IDRs) (144, 145). Due to their apparent lack of stable structural elements, they sample an ensemble of rapidly interconverting conformational states akin to those of denatured proteins (ensemble of denatured states, DSE). This makes their detailed structural characterization a challenge.

Many standard methods, which have been employed to characterize IDPs and DSEs (50), can only determine average ensemble features like size, (pre-) molten globule behavior or residual secondary structure content (50). This limited experimental resolution reduces the description of the state ensembles to simple models. Nowadays, NMR analysis of chemical shifts, relaxation dispersion or order parameters offer the advantage of residue-specific information. The local mobility, the surrounding secondary structure or short distances between residues ≤ 5 Å apart in space can be determined. The usage of EPR with appropriate spin labels can extend this distance up to ~ 50 Å. Nevertheless, for a mixture of dynamic structures still only the ensemble average for the respective observable is obtained.

In recent years, fluorescence spectroscopic methods arose as “tools-of-choice” when studying IDPs or the DSE (130, 146-148). They have numerous advantages such as: *i*) The low concentration of sample needed for single-molecule fluorescence spectroscopy experiments prevents possible aggregation. *ii*) Additionally, Förster Resonance Energy Transfer (FRET) probes distances between the labeling sites and their fluctuations in equilibrium between two protein sites up to 100 Å apart. Hence, structure and dynamics of a sample can be obtained simultaneously. *iii*) When combining high-resolution single molecule spectroscopy with multiparameter fluorescence detection (MFD) (25, 57, 149), not only the full heterogeneity of the samples can be caught, but also sub-ensembles can be identified and species-specific correlation analysis can reveal the samples’ kinetics up to seven orders of magnitude in time (sub- μ s - ms, (47, 70, 72)) without the need of external synchronization as in non-equilibrium experiments. *iv*) Moreover, the use of single-labeled variants and polarized excitation-

emission permits probing the local mobility of the protein backbone via fluorescence anisotropy (25, 53, 149).

IDPs and DSEs are often described using simple polymer models, implicitly assuming that no significant secondary or tertiary structure exists (12, 13, 18). In contrast, for both IDPs and DSEs the formation of at least transiently ordered structure is anticipated to facilitate the binding to ligands (50) or the folding into the proteins' native structure (150). Thus, a more complex behavior should be expected.

Here, we use the lysozyme from the phage T4 (T4L) as a model system for which many studies concerning protein stability, function and folding have been conducted (20, 21, 27-31, 35, 38-44) (Chapter 4). T4L is a thorough and well-characterized enzyme of which the enzymatic function is known (35)(Chapter 3), and how it reacts to the introduction of mutations, insertions and deletions within its amino acid sequence (27) and how it unfolds via at least two intermediates *I* and *J* (20) (Chapter 4). Structurally T4L consists of two subdomains (NTsD and CTsD), which are connected by a long α -helix (Figure 5.2.1A) and differ in stability (33, 41).

In this study we focused on the characterization of the DSE of T4L in presence of high concentrations of the chemical denaturant urea by a combination of ensemble (fluorescence lifetime and anisotropy) and single-molecule (MFD, photon distribution analysis (PDA), full and filtered fluorescence correlation spectroscopy (fullFCS and filteredFCS, respectively)) fluorescence methods (Figure 5.3.1). We will firstly characterize the polymer behavior of chemically denatured T4L, and then try to identify possible nucleation sites and use obtained mean distances to deduce the overall shape and size of T4L's DSE. Here, the behavior of a network (Chapter 3) of 24 variants (Figure 5.2.1A) labeled site-specifically with a FRET dye pair in ensemble was analyzed with different fit models to validate its polymer behavior. We showed that simple polymer models are not applicable to describe the properties of T4L's DSE. Additionally, we compared T4L's DSE to the conformer equilibrium under native conditions (Chapter 3) to identify apparent residual structure and possible nucleation sites. The obtained apparent distances, which were used to gain an approximate description of the size and form of chemically denatured T4L, revealed that its mean structure is native-like, albeit it is larger than the native state conformation. However, to increase the resolution of the DSE, we would have to optimize our labeling network and probe additionally intra-subdomain FRET distances. Finally, we analyzed the site-specific local mobility and global kinetics with correlation methods to show the rough energy landscape within the DSE.

We demonstrate here the necessity of careful data interpretation, but also the potential of a multidimensional approach to characterize an ensemble of states. This is of major interest for denatured proteins and even more for the studies of IDP's, which so far eluded deeper structural characterization by traditional biophysical methods.

5.2.2. Results

5.2.2.1. *Denatured T4L does not behave like a polymer*

To examine the polymer behavior of the DSE of T4L, we performed ensemble fluorescence lifetime measurements in 7.5 M urea of all 24 T4L-FRET pairs (Figure 5.2.1A). Two polymer models, the Gaussian chain (GC) and worm like chain model (WLC) (Section 5.3.1.3, Table 5.3.1A-B), were used to fit the time-resolved fluorescence intensities of all double-labeled variants in denaturing conditions. The GC model (eq. 5.3.11) assumes that the interdyne distances in the denatured state behave as a random coil. In this model each amino acid residue is treated as a monomer of fixed size (here: 3.6 Å, the size of an amino acid residue), and that adjacent bond angles (i.e. the orientation of the monomers to each other) are uncorrelated (14, 16). The sole fit parameter is the number of monomers between the two label sites. The WLC model (eq. 5.3.12, 5.3.13), conversely, assumes that the denatured amino acid chain behaves like a semi-flexible rod of total length L with stiffness κ . Thus, the chain is being comprised of longer, uncorrelated links (12, 141) instead of the fixed monomers in the GC model.

Considering the variant 19-119, we found that the GC model was not able to describe our data (Figure 5.2.1B). Given the heterogeneous monomer sequence (i.e. amino acid chain) of T4L, it is also not expected that T4L behaves like an ideal random coil. The WLC model described the data best possible with a flat distribution of residuals and a χ^2_r of 1.03 ($\chi^2_{r,avg} = 1.33$ for the GC, and 1.14 for the WLC, Table 5.3.1B). Keeping in mind that our fluorophores are coupled to T4L via long and flexible linkers, which can move freely, and the non-random-coil behavior of denatured T4L, we decided to decouple the linker-induced fluorophore motion from the biomolecules motion. Thus, we modified the WLC model accordingly (eq. 5.3.14, 5.3.15). This led to a constant increase of κ and decrease of L (Figure 5.3.3C-F) compared to the model without explicit linker-treatment. From the difference in L for the standard WLC and the linker-corrected WLC-model, we determined an offset of ~ 4 Å due to the fluorophore linkers (Figure 5.3.3C, Table 5.3.1B). κ of the denatured T4L was distributed between 0.54 and 1.01 (Figure 5.2.1C), which identifies it as semi-flexible chain

(141). Both κ and L (61 Å to 107 Å) increased with increasing number n of amino acid residues between the labeling sites (Figure 5.3.3D-E). The topology diagram in Figure 5.2.1C showed how κ and L are distributed for the individual variants. For κ (upper triangle), the value on the x -axis marks the donor fluorophore position and the y -value the acceptor fluorophore position. For L in the lower triangle, the order is reversed (acceptor = x -axis, donor = y -axis). The closer the data points are to the diagonal line (black), the closer the two fluorophores are in sequence. In the majority of variants, stiffness and chain length were correlated (similar colors in Figure 5.2.1C); however, two variants deviated strongly from that trend: 60-150 and 69-150 (lower red circle in Figure 5.2.1D). They have the acceptor position I150C in common, but the donor position is located at the beginning (K60pAcF) or the middle (Q69pAcF) of α -helix C (Figure 5.2.2A).

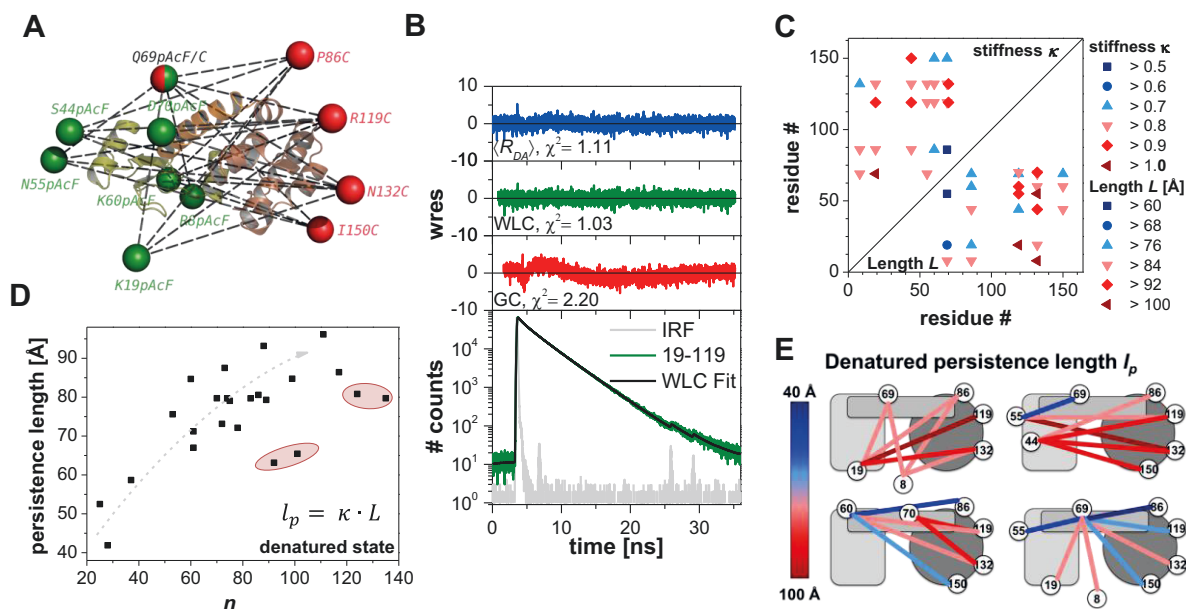


Figure 5.2.1 T4L network shows Worm-like chain behavior. (A) Distance network of 24 DA FRET pairs. At donor positions (green), the inserted unnatural amino acid pAcF was labeled with Alexa488-hydroxylamine; the cysteine at the acceptor positions (red) was labeled with Alexa647-maleimide. Position 69 was used both donor and acceptor position. Spheres represent mean dye position of the PDB-structure 172L as calculated with the AV-approach (53). (B) Time-resolved fluorescence intensity of double-labeled variant 19-119. On top, the residual for fitting the data with a Gaussian distributed mean distance, the linker-corrected WLC model or the GC model are shown. (C) Total chain length L and stiffness κ for the WLC fit. (D) Persistence length calculated from length and stiffness for the linker-corrected WLC fit. (E) Distribution of persistence length l_p from WLC model depicted schematically on a model of T4L (light grey: N-terminal subdomain, Middle grey: Helix c, dark grey: C-terminal subdomain).

We calculated the persistence length l_p from κ and L for the variants (Figure 5.2.1C-E). l_p is defined as the chain length over which correlated chain motions decay (13, 141). It rose from 42 Å to 97 Å with increasing number of residues n between the labeling sites (The fluorophore linkers, each ~ 20 Å, were assumed to be equivalent to 11 additional residues

(53)). l_p decreases especially for variants with very large n : 19-132 and 8-132 (upper red circle in Figure 5.2.1D). Here, the donor fluorophore was placed either in the β -sheet region of T4L or in α -helix A, which structurally-wise belongs to the CTsD (Figure 5.2.5A) (33, 41). For a protein obeying the WLC model ideally we would firstly expect that the fitted total chain length L_{fit} resembles roughly the theoretical length L_{theo} of the amino acid chain between the two labeling sites ($L_{theo} = n * 3.6 \text{ \AA}$). Figure 5.3.3B shows that in denatured T4L $L_{fit} \ll L_{theo}$. Secondly, the fact that the longer the chain is, the stiffer it gets, seemed at first pass contradictory. Finally, l_p was strongly decreasing for residues with either a large sequence separation or with one labeling site close to the C-terminus (I150C) (Figure 5.2.1C-E). Altogether, these observations indicated the presence of (apparent) residual structure shortening the distance between the two label sites. The longer the chain, the larger is the probability of apparent residual structure. Additionally, we could conclude that the marked variants (red circle in Figure 5.2.1D) might reflect the softness of the CTsD in the unfolded ensemble and the increased likelihood of looping structures for long chains, in which the chain ends are close together (151).

From eTCSPC we concluded here that the chemically denatured ensemble of T4L does not consist of a mainly featureless amino acid chain but that tertiary interactions and secondary structure, of either residual or persistent nature, led to deviations from the anticipated polymer behavior.

5.2.2.2. Heterogeneity on the single-molecule level

To understand the results obtained from the polymer model fitting of the ensemble data, we determined the variants' appearance and behavior using single-molecule multiparameter fluorescence detection (MFD) (25). The two-dimensional MFD histograms relate two FRET indicators: *i*) the donor to acceptor fluorescence intensity ratio (F_D/F_A) and *ii*) the fluorescence-weighted fluorescence lifetime of the donor in presence of acceptor ($\langle \tau_{D(A)} \rangle_f$). Exemplary, the MFD histogram for a variant with a large (19-119) and a small (55-69) interdye distance are shown in Figure 5.2.2. The variant 19-119 with $n = 100$ exhibited, under both native (black) and denaturing (violet) conditions, a narrow distributed population centered on the static FRET line (black/ violet-dashes). When unfolding the variant 19-119, $\langle \tau_{D(A)} \rangle_f$ and F_D/F_A increased from $\sim 2 \text{ ns}$ (ratio ~ 2.5) to $\sim 3 \text{ ns}$ (ratio ~ 9) (Figure 5.2.2B), indicative for a larger distance separation between the two label positions. Additionally, a tailing effect towards lower F_D/F_A ratio (i.e. the DOnly region) was observed. In contrast, the

variant 55-69 ($n = 14$) spread over a wide range of $\langle\tau_{D(A)}\rangle_f$ and F_D/F_A values under both native and denaturing conditions (Figure 5.2.2C).

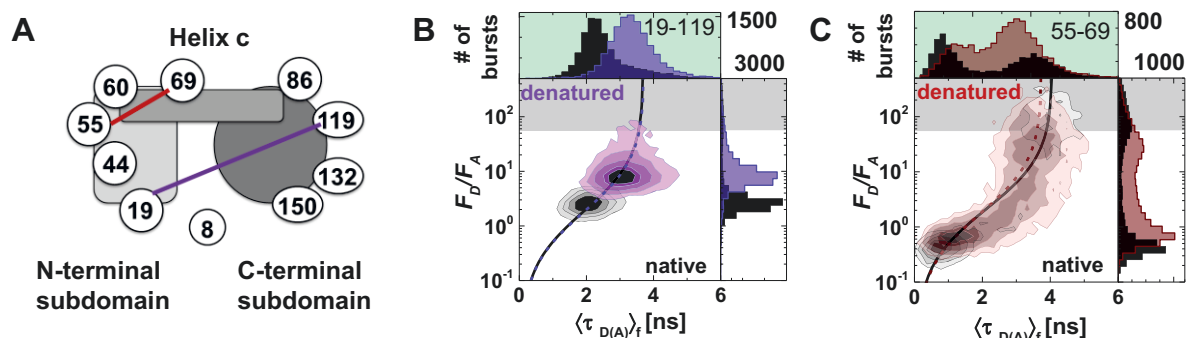


Figure 5.2.2 Heterogeneity on the single-molecule level. (A) Schematically model of T4L (light grey: N-terminal subdomain, Middle grey: Helix c, dark grey: C-terminal subdomain). Labeling sites for variants displayed to the right are shown in violet and wine, respectively. (B) MFD histogram of 19-119-(DA) labeled T4L in 0 M (black) and 7.5 M urea (violet). Two dimensional histogram of ratio of green over red fluorescence signal 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. Static FRET line is shown in black and dashed violet, respectively. Pure donor and acceptor fluorescence (F_D and F_A) are corrected for background, spectral cross-talk and detection efficiency ratio (Table 5.3.4). Shaded area in gray is the region of donor only. (C) Same as (B) for 55-69-(DA). Broad smeared distribution of data indicates averaging on ms time scale.

The main populations were found at very short lifetimes (< 1 ns), reflecting the extraordinary short interresidue distance still present also when the protein was denatured. The second population showed a high F_D/F_A value, characteristic for molecules either lacking an active acceptor (bleached molecules due to very high FRET occurring) or a very large interdyer distance. Both populations are connected by some “smear” on the right side of the static FRET-line which reflects a slow interconversion between those two populations. This might be caused by acceptor blinking process (photophysical artefact) or by a repeatedly slow unfolding-refolding of the helix c between the two label positions.

Taken the MFD results shown in Figure 5.2.2 and Figure 5.3.6 together, we could conclude that each variant displayed its own unique behavior and that the shape and distribution of the data is a clear indication of the variants’ heterogeneous behavior.

5.2.2.3. Single-molecule kinetics depends upon chain length

Given the complexity of T4L’s behavior anticipated from the polymer model analysis and the visual inspection of our MFD plots, we decided to describe the population distribution with two empirical models using photon distribution analysis (PDA) (48, 63-65). This facilitates the comparison with the native state data as for a correct quantitative description an altered model function would be required. Kinetic information can be extracted from MFD experiments by binning the experimentally obtained photon trace in different time-windows

(TWs). Dynamic exchange between two or more limiting states leads to a narrowing or broadening of the data histogram, respectively (48, 63, 65). This is shown exemplary for the variant 19-119 in 7.5 M for the two TWs 1 ms and 4 ms in Figure 5.2.3A.

In the first model, we used a 1 ms TW by using static photon distribution analysis (sPDA) (SI Section 5.3.1.7) and used a Gaussian distributed mean distance $\langle R_{DA} \rangle_E$ with half-width $hw_{DA,s}$ (and a DOnly-fraction) as an empirical description of the shape of the F_D/F_A distribution. Under denaturing conditions both $\langle R_{DA} \rangle_E$ and $hw_{DA,s}$ are larger than in the native state (Figure 5.2.3B-C). $\langle R_{DA} \rangle_E$ increased with larger separation n of labeling position, but the half-width $hw_{DA,s}$, however, was reduced for larger n .

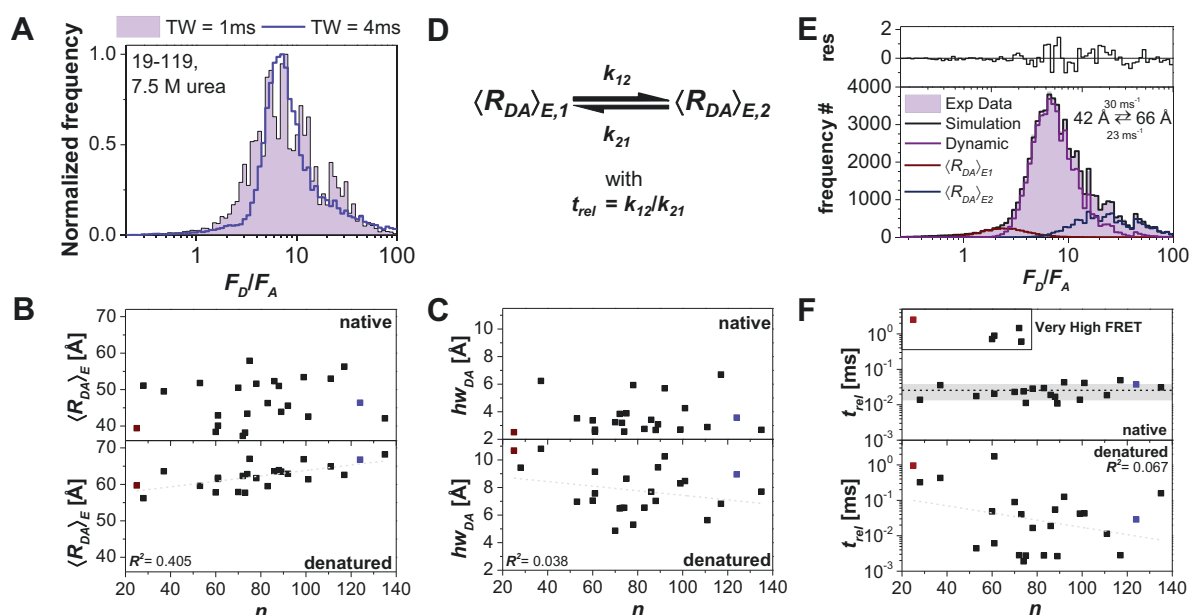


Figure 5.2.3 Single-molecule kinetics depends upon chain length. (A) Millisecond dynamics induces narrowing/broadening of histograms when analyzed with different time windows (TWs). (B-C) Under denaturing conditions, with increasing number of amino acid residues between labeling sites n , the mean distance determined in 1 ms TW analysis of static PDA increases and the half-width decreases. (D) Fit model dynamic PDA. (E) Example of dynamic photon distribution analysis with different time windows, with the fit model shown in (F); shown is the 2 ms TW. (F) Relaxation time constants t_{rel} , i.e. the exchange occurs faster. With the exception of five variants (8-69, 19-69, 55-69, 70-119 and 70-132), the t_{rel} in the absence of urea are distributed narrowly $\sim 20 \mu s$. Violet and wine data point represents the variant 19-119 and 55-69, respectively.

Next, we assumed that the population indeed consists of two limiting states $\langle R_{DA} \rangle_{E1}$ and $\langle R_{DA} \rangle_{E2}$ with limited half-width $hw_{DA,d}$, which exchange with certain apparent relaxation rate constants, k_{12} and k_{21} (Figure 5.2.3D-E). Here, we used dynamic PDA (dPDA) and analyzed globally three different TWs (2 ms, 3 ms and 4 ms). dPDA has been shown to deduce the limiting states and their dynamic interconversion with high precision when the rate of exchange is similar to the burst duration (63, 65).

With increasing n the apparent relaxation time t_{rel} ($= 1/(k_{12}+k_{21})$) decreased under denaturing conditions (Figure 5.2.3F), i.e. the exchange between the two states $\langle R_{DA} \rangle_{E1}$ and $\langle R_{DA} \rangle_{E2}$ occurred faster. This agrees with the decreased $hw_{DA,s}$ observed in sPDA (Figure 5.2.3C); a small $hw_{DA,s}$ indicates only small changes in FRET-efficiency or D-A distance correlated with only minor internal protein movements during the burst duration. These can occur with a higher frequency (faster t_{rel}) than larger (sub-(domain)) motions. Under native conditions, the mean relaxation times t_{rel} was in a narrow range $\sim 20 \mu s$ – except for five deviating variants (Figure 5.2.3F). Those variants showed very high FRET under native conditions, thus, dPDA fitted here as dominant process the slow (ms) “bleaching” of the acceptor fluorophore.

From sPDA and dPDA we can conclude that each of the 24 variants displayed their own, unique kinetic behavior when unfolded (Figure 5.3.6); thus the analysis displays the overall nature of the heterogeneity found in the ensemble of denatured states.

5.2.2.4. Characterization of the unfolded ensemble: Distance effects & Heterogeneity

Having established the non-ideal, heterogeneous behavior of the chemically denatured state of T4L, we next used the ensemble time-resolved fluorescence intensity data to describe the average, apparent features of our ensemble using on the one hand the mean FRET efficiency and on the other hand a distance distribution fit.

To compare the ensemble of conformations of the native states (Chapter 3) with the DSE, we calculated the mean transfer efficiency $\langle E \rangle$ (Figure 5.2.4A-B). For the native state we used the previously published data (Chapter 3). Here, $\langle E \rangle$ showed a bowl-shaped distribution – reflecting that fluorophores with large sequence separation are close in their 3-dimensional structure (Figure 5.2.1A, Figure 5.2.4A). In average, $\langle E \rangle$ was higher, i.e. the two fluorophores are closer together, in the native than in the denatured state. $\langle E \rangle$ of the denatured protein showed roughly an exponential decline (Figure 5.2.4B). In concord with their shorter l_p , the variants 60-150 and 69-150 displayed a higher FRET efficiency (shorter distance between fluorophores) as would be expected from their labeling site distance. We overlaid the data points with two curves; the dotted curve represents the development of $\langle E \rangle$ for an increasing chain, when assuming the average stiffness κ_{avg} obtained from the fit to the linker-corrected WLC fit of all 24 FRET pairs. A better description of the average behavior is achieved, when the empirical value of $\kappa \sim 0.6$ is used for the calculation. However, also here the description for short chain lengths is still bad.

As the WLC model is not an appropriate model to describe the native distribution of states in folded proteins, we decided to empirically describe the fluorescence lifetime decay of the T4L FRET pairs under native and denaturing conditions additionally through a Gaussian distribution with mean distance $\langle R_{DA} \rangle$ and width w_{DA} (Figure 5.2.1B, Figure 5.2.4C-E). In denatured conditions, the average width of the distribution more than doubles from 9 Å to 21 Å and the average mean distance increases from 44 Å to 61 Å, which was an expected result when proteins are unfolded (Figure 5.2.4C). The increasing width shows that the protein itself was more flexible and populated conformations with a lifetime > 10 ns, because dye movements alone will not be able to reach a width > 10 Å (53). From these observations, it was possible to infer the existence of conformational fluctuations of the protein. In concord with $\langle R_{DA} \rangle_E$, for $\langle R_{DA} \rangle$ an increasing n led to an in average larger distance between the two fluorophores (Figure 5.2.4E). Most informative were here those variants, where either no change in $\langle R_{DA} \rangle$ occurs or those, in which $\langle R_{DA} \rangle$ was rising significantly. Under native conditions, e.g. the fluorophore in both 55-69 and 70-119 were close together. Under denaturing conditions, they were still for 55-69 ($n = 14$, $\langle R_{DA} \rangle = 41$ Å), whereas for the variant 70-119 ($n = 49$) we obtained a large distance ($\langle R_{DA} \rangle = 69$ Å). $\langle R_{DA} \rangle$ was correlated with the total chain length L and persistence length l_p (Figure 5.3.3I).

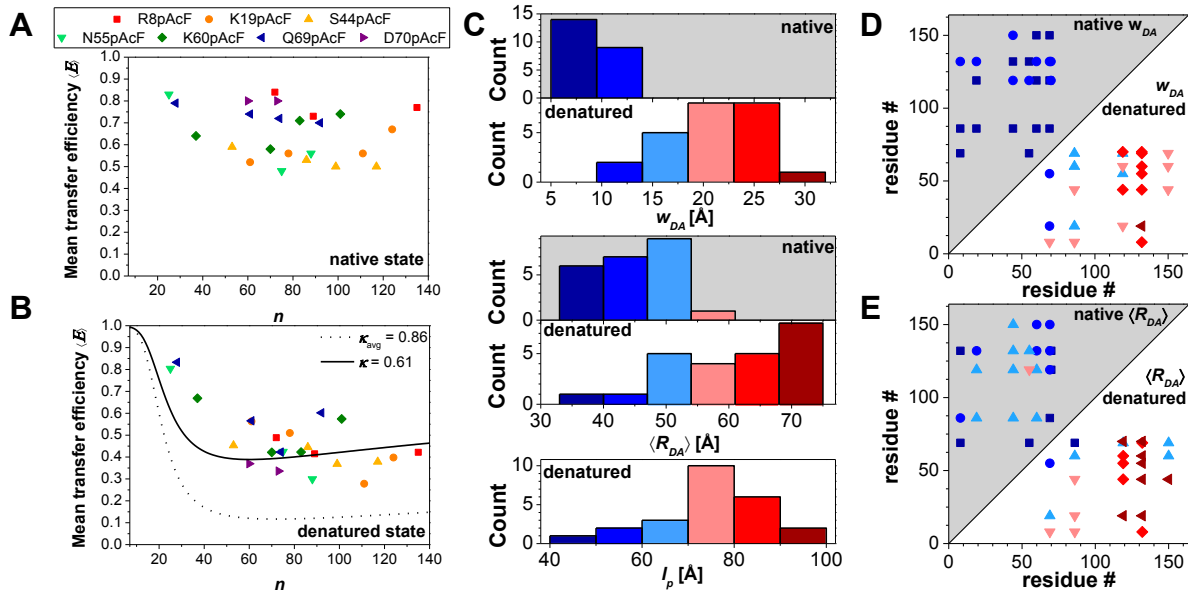


Figure 5.2.4 Ensemble characterization of T4L with eTCSPC. (A,B) Mean Transfer Efficiency $\langle E \rangle$ in 0 M and 7.5 M urea containing buffer. 0 M values were calculated from previously reported results (Chapter 3), 7.5 M points were calculated from fitting of the eTCSPC data with linker-corrected Worm-like chain model. (C) Histogram for obtained parameters in the WLC and Gauss fit. (D, E) Topology diagrams for mean distance $\langle R_{DA} \rangle$ and width w_{DA} obtained from distribution fit in the absence and presence of urea. Color code from (C) is kept for (D, E).

Alike $\langle R_{DA} \rangle$, w_{DA} also increases with increasing n (Figure 5.2.4D). This is most likely a consequence of the correlation between $\langle R_{DA} \rangle$ and w_{DA} (Increasing $\langle R_{DA} \rangle$ lead to larger w_{DA} due to increasing uncertainties, Thomas Peulen, personal communication). Nevertheless, some labeling positions (8, 44, 70, 86, 132, 150) showed a homogeneous w_{DA} ranging from 20 -27 Å (16 – 23 Å for 86). For position 69, however, the obtained w_{DA} ranged from 9 -25 Å. We conclude here that as anticipated the overall dimensions of T4L are increased in denaturing conditions and that it was more flexible than in native conditions.

5.2.2.5. Characterization of the proteins segment mobility

To identify the stiffest regions of T4L's backbone or regions with the highest probability of residual structure, we used fluorescence anisotropy as the local readout of backbone mobility. From arrangement of the fluorophores in all variants (Figure 5.2.5A), only two fluorophore positions (19: β -sheet, 55: loop) do not belong to α -helices, which are generally thought to be more stable than other secondary structural elements (28, 30, 101). As usually the side-chain mobility follows the backbone mobility, fluorescence anisotropy measurements can report with high confidence the backbone flexibility.

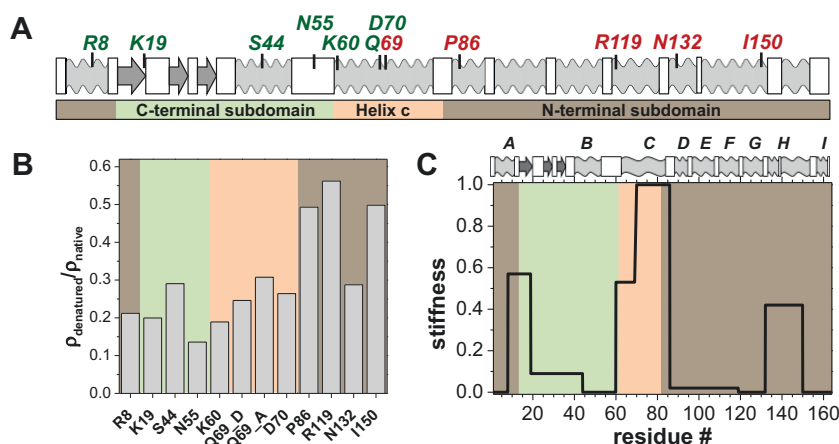


Figure 5.2.5 Different mobility of protein segments. (A) Secondary structure composition of folded T4L. White boxes resemble loop region, light gray waves α -helices and dark gray arrows the β -sheets. Green and red numbers indicate residues used for site-selective donor (green) or acceptor (red) fluorophore coupling; Q69 was used as both, donor and acceptor position. (B) Relative rotational correlation time ρ_{rel} of single labeled variants calculated as $\rho_{\text{av}}(7.5 \text{ M})/\rho_{\text{av}}(0 \text{ M})$ averaged over shared dye positions (SI 5.3.1.4, eq. 5.3.18). (C) Stiffness of the individual protein segment probed as calculated from the WLC model (details see 5.3.2.3).

As an indicator for local mobility we calculated from our ensemble measurements the average rotational correlation time ρ_{av} from the fluorescence-weighted average fluorescence lifetime $\langle \tau \rangle_f$ (eq. 5.3.3) and the steady-state fluorescence anisotropy r_{ss} of our single-labeled variants using the Perrin-equation (SI Section 5.3.1.4, eq. 5.3.18). By dividing the viscosity corrected ρ_{av} from the denatured samples by the values obtained under native conditions

(Figure 5.2.5B), we determined label-position specific, to what extent the rotational motion of the linker-connected fluorophore, and thus the local side-chain mobility, was enhanced by denaturation. Generally, ρ_{av} is decreased significantly for all labeling positions. For three residues in the CTsD (86, 119 and 150) labeled with the acceptor fluorophore Alexa647, ρ_{av} is halved, while for all other positions (8 to 70 (NTsD), 132), the ρ_{av} decreased much more (Figure 5.2.5B). The strong decrease in both ρ_{av} and r_{ss} (Figure 5.3.4C) for the NTsD goes in concord with prior studies reporting an intermediate state, in which the NTsD is unstructured at less denaturing conditions (20, 21, 41-44).

Moreover, anisotropy experiments can be used to report the effect on the stability of both mutation sites. For example, a group of variants sharing one dye position does not behave homogeneously because the introduced mutation in the other subdomain may or may not influence the population of a certain conformation (20). This is the case for the position 86 labeled with Alexa647, which behaves nearly as in folded protein when combined with the second, unlabeled mutation in position 8 (both 0 M and 7.5 M $r_{ss} = 0.24$), whereas when combined with the mutation at site Q69pAcF, its steady state anisotropy is strongly decreased (0 M: $r_{ss} = 0.26$; 7.5 M: $r_{ss} = 0.15$) (Figure 5.3.4C-I, Table 5.3.2C-D).

To gain insight into the segmental mobility, i.e. the mobility of the individual secondary structure elements, we used the stiffness values obtained from the linker-corrected WLC fit (Table 5.3.1B) and constructed an harmonic network model, where each labeling pair is considered as a spring where the spring constant between each pair is the WLC stiffness (SI section 5.3.2.4). A series of springs with spring constants k_i can be simplified by one equivalent spring with spring constant K (152) according to equation 5.3.32. Thus, the obtained stiffness values κ for a labeling variant resemble K . A protein segment with spring constant k_i is defined by the amino acid residues lying in between two labeling positions. For example, consider the FRET pair 8-69 where the total interdye segment (S_T) is composed of the sum of five segments: S_1 [8 – 19], S_2 [20– 44], S_3 [45- 55], S_4 [56-60] and S_5 [61-69]. The numbers in brackets are amino acid positions in the sequence of T4L (SI section 5.3.2.4). The same is done for all other FRET pairs. In total, a linear system of 24 equations is built. By solving this inverse problem it was possible to determine the stiffness per segment (Figure 5.2.5C) and identify regions with higher stiffness or higher propensity of formation of residual secondary structure. The stiffness is highest in the second half of the connecting α -helix “c” and in the CTsD it is generally higher than in the NTsD. These results correlate well

with previous studies (20, 21, 41-44), which have shown that the CTsD is more resistant against unfolding.

Unexpectedly, the results obtained from fluorescence anisotropy experiments and the build stiffness matrix locate the regions with highest probability of residual structure in different proteins segments. This might be indicative for a more complex behavior of the denatured amino acid chain as can be captured with these two (averaging) methods.

5.2.2.6. *Fluorescence correlation spectroscopy probes chain dynamics*

To study the kinetics of the denatured ensemble of states, which we inferred already from the shape and distribution of our MFD plots and the dPDA analysis, in more detail, we applied fluorescence correlation spectroscopy (FCS). The heterogeneity found in the ensemble and single-molecule experiments led us to anticipate a rough energy landscape within the DSE. This would lead to a complex kinetic behavior of the chemically denatured T4L (4, 19, 130). Full fluorescence correlation spectroscopy (fullFCS) with its high-time resolution and large dynamic range is able to capture dynamics on the timescales from the lifetime of the observed fluorophore (~ 4 ns for Alexa488, ~ 1.2 ns for Alexa647) up to the diffusion time of the biomolecule of interest (here: for T4L ~ 0.2 ms) (69). Here, we probed the fluorescence fluctuations of a selected subset of five variants out of the 24 variants encompassing the full network (Figure 5.2.6A). Figure 5.2.6B shows exemplarily the donor-signal autocorrelation $G^{(DX)}_{DD}$ curve of the double-labeled T4L variant 60-132 in the absence (black) and presence of 7.5 M urea (gray). In order to identify FRET-induced donor quenching, we fitted both, the autocorrelation function (*ACF*) of donor-signal in the DOnly-labeled ($G^{(D0)}_{DD}$, *DOnly-ACF*) and DA-labeled sample ($G^{(DA)}_{DD}$, *DA-ACF*), globally and needed up to four relaxation time constants (SI Section 5.3.1.8, eq. 5.3.28) in addition to the diffusion- and photon antibunching (i.e. the fluorescence lifetime) component to obtain a flat distribution of the weighted residuals (Figure 5.3.9, Table 5.3.6). Generally, the relaxation time components t_R found in DOnly are also present in the DA-sample. The t_R already present in *DOnly-ACF* might have been due to fluorescence brightness fluctuations induced by environmental changes around the fluorophore (133) or local quenching by aromatic amino acids (131). However, the *DA-ACF*'s display additional dynamics induced by FRET (69), e.g. the very fast component in the 100 ns range (Figure 5.2.6E, hatched bars). The t_R were distributed between 100 ns and 1 ms covering the fast chain dynamics, local (sub-global) and global fluctuations of T4L (4, 19).

This result agrees very well with the cross-correlation functions (*CCF*) (Figure 5.2.6C) of the donor-signal with the (FRET-sensitized) acceptor signal ($G^{(DA)}_{D|A}$, *DA-CCF*), which showed only two time constants in addition to diffusion and photon antibunching (eq. 5.3.24, 5.3.25): a very fast component $t_{RI} \sim 100$ ns and a slow component, slower than the diffusion time, with $t_{R4} \sim 1$ ms (Figure 5.2.6E).

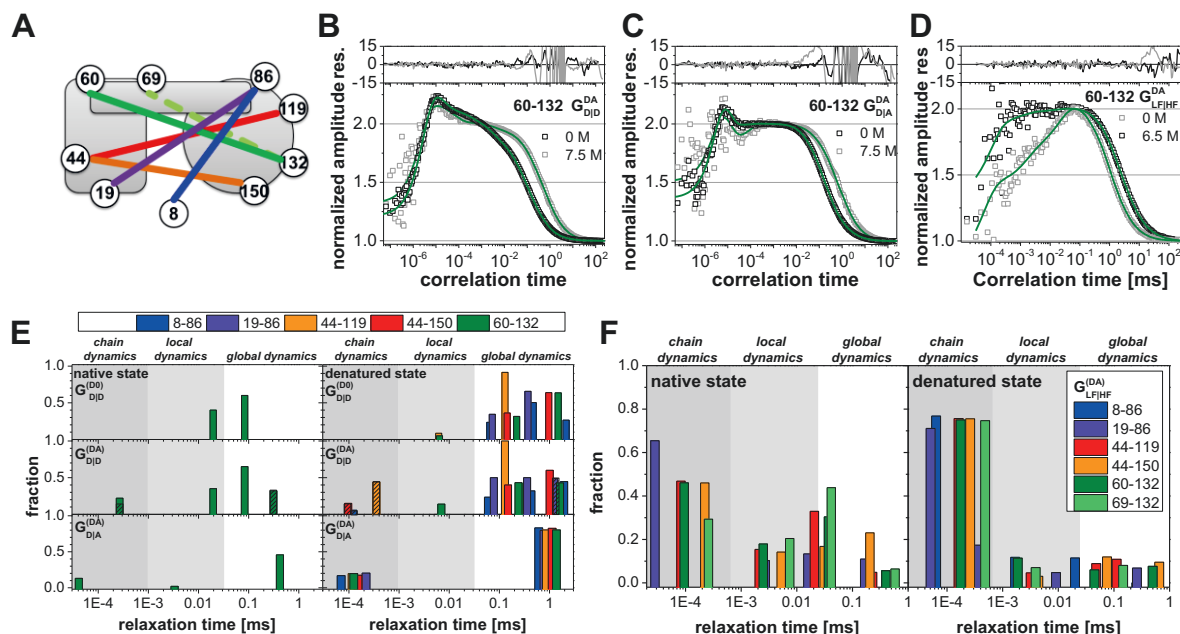


Figure 5.2.6 Native and unfolded T4L dynamics displays a wide range of time scales. (A) Variants used in FCS experiments; variant 69-132 was only measured in filtered FCS. The color code is kept throughout the figure. (B) Example of the donor-signal autocorrelation curve of full FCS for the variant 60-132-(DA) in 0 M and 7.5 M urea (*DA-ACF*). Residuals from global fitting are shown on top. (C) Example of donor-acceptor cross-correlation curve of full FCS for the variant 60-132-(DA) in 0 M and 7.5 M urea (*DA-CCF*). Residuals from global fitting are shown on top. (D) Example of species-selective cross-correlation curves (*sCCFs*) for the variant 60-132-(DA) in 0 M and 7.5 M urea, filters were generated from single molecule bursts belonging to low FRET (LF) and high FRET (HF) species (SI Section 5.3.1.9, Figure 5.3.10). Residuals for the global fit are shown on top. (E) Distribution and amplitude of time constants found in 7.5 M urea of full FCS measurements in donor-signal autocorrelation curves for DOnly and DA samples (Top, middle). Bottom row shows distribution and amplitude of time constants in donor-acceptor crosscorrelation curves. (F) Obtained relaxation time constants and their respective fractions for different variants in 0 M and 7.5 M urea of *sCCFs* from filtered FCS.

In summary, in agreement with the MFD histograms, fullFCS identifies dynamics on a broad range of timescales from the ns up to the ms, where especially FRET-fluctuations are found on either extreme site of the timescale.

5.2.2.7. Species-selective FCS shows a wide range of kinetics

Next, we performed species-selective filteredFCS experiments of double-labeled samples to “zoom-in” on processes changing the FRET-efficiencies from high FRET to low FRET. Here, only fluctuations that are specific to FRET populations are observed and those fluctuations e.g. induced by triplet state blinking of the fluorophores are being suppressed

(47, 70, 72). This is possible by using species-selective filter which amplify the transitions between two selected border species, here a low FRET and high FRET species (Figure 5.3.10), and thereby report on all intermediate processes passed, when the molecule transits from one border state to the other. The species-selective cross correlation curves ($G^{(DA)}_{LF|HF}$, $sCCF$'s) for the variant 60-132 in the absence (black) and presence (gray) of urea are displayed in Figure 5.2.6D. Under native conditions, the global fit of both species- ACF 's ($sACF$, $G^{(DA)}_{HF|HF}$ and $G^{(DA)}_{LF|LF}$) and $sCCF$'s ($G^{(DA)}_{LF|HF}$ and $G^{(DA)}_{HF|LF}$) needed four relaxation times (SI Section 5.3.1.9, eq. 5.3.26, 5.3.27) next to the diffusion component (here: ~ 1 ms): $t_{rel1} = 99.5 \pm 50$ ns, $t_{rel2} = 3.22 \pm 1.16$ μ s, $t_{rel3} = 23.3 \pm 3.29$ μ s and $t_{rel4} = 241 \pm 37.4$ μ s (Figure 5.2.6F). t_{rel2} and t_{rel4} have been identified in a previous study to represent the dynamics of the enzymatic cleavage of peptidoglycan by T4L (Chapter 3). The fastest t_{rel1} has also been found in the $DA-ACF$'s and $DA-CCF$'s in the fullFCS experiments (Figure 5.2.6E). Although, the t_{rel} 's are very similar for the variants examined, their amplitudes are label-position specific (Table 5.3.7). Under denaturing conditions, the relaxation times are very similar to those of the native state ($t_{rel1} = 159 \pm 48.4$ ns, $t_{rel2} = 2.49 \pm 1.44$ μ s, $t_{rel3} = 44.9 \pm 17.7$ μ s and $t_{rel4} = 444 \pm 181$ μ s); here, the amplitudes of t_{rel2} - t_{rel4} have decreased significantly and the amplitude vs. t_{rel} plot curves (Figure 5.2.6F) for the probed six variants show all a related trend. The amplitude of the fast chain motion t_{rel1} had the largest amplitude of ~ 80 %.

In summary, filteredFCS showed that the chain diffusion identified also by fullFCS, is the dominating process (amplitude ~ 80 %). Additionally, FRET-induced fluorescence fluctuations take place on intermediate timescales (μ s range), but their amplitude is low, and thus the contrast is small. The t_{rel} found for folded and unfolded T4L are very similar; however, their amplitudes change.

5.2.3. Discussion

5.2.3.1. Chemically denatured T4L is not a random coil but contains residual structure

In an ongoing debate, the existence of residual native or transient structure in chemically denatured proteins and IDP's is discussed extensively (153, 154). Our approach to this problem was to study a network of 24 FRET pairs under relatively strong denaturing conditions. We analyzed the time-resolved fluorescence intensities with various models and showed that it behaves like a non-ideal Worm-like chain, in which on average the parameters stiffness κ , total length L and persistence length l_p increase with increasing number n of

residues between the labeling sites (Figure 5.2.1). However, L is much shorter than would be expected. This polymeric behavior of denatured protein is commonly found (148, 155); and in contrast to other studies (148, 154, 155), the simple Gaussian chain model was not able to describe our data. This is supported by the analysis of O'Brien *et al.*, which could show that for both the simulated and experimental data of Protein L the GC model was inadequate (16). Additionally, they found that caution should be taken when inferring ensemble characteristics like the radius of gyration from experiments (16).

Often, researchers conclude by their observation that a polymer model can describe the experimentally obtained data, the absence of structure. Fitzkee and Rose showed in 2004 (18) that proteins with slightly enhanced flexibility can convincingly show polymer characteristics. This in concord with our previous study (Chapter 4), in which T4L was titrated with urea and even in the absence of urea the WLC model could fit the experimental data, although the model had no physical meaning and we only used it to compare how the parameter develop with increasing urea concentrations.

In light of the apparent residual structure, which might be present within a significant part of the vast number of conformations encompassed by the DSE, we have our doubts that polymer models are physically justified for chemically denatured proteins. Thus, we agree with previous studies, which stressed the need for a careful data interpretation when working with polymer models and unfolded proteins (16, 18). For T4L, we think that a better model might still be based on the WLC; however, due to the heterogeneity observed in the MFD plots, we should fit rather a distribution of stiffness's, i.e. persistence lengths, and keep the total chain length to a reasonable value, but that approach goes beyond the limits of this current work.

5.2.3.2. Identification of possible nucleation sites

To get a more detailed overview, where to find the apparent residual structure in T4L we used a stiffness matrix compiled from the WLC-fit results and the fluorescence anisotropy of single-labeled samples. This apparent residual structure might serve as nucleation point for structure formation and thus, could drive the folding of proteins (153, 156).

Najbar *et al.* (157) approached this question by dissecting T4L in its secondary structure elements (Figure 5.2.5A) and tested the isolated protein segments for their intrinsic propensity for structure (mainly α -helix) formation. They identified the α -helices A, C, E and the β -sheets as most likely nucleation sites, whereas the α -helices B, D, F, G and H did not show significant structure formation. This goes nicely in concord with the results, which we

obtained from our stiffness matrix (Figure 5.2.5C). Here, the stiffest segments were found to be in α -helices A, C and H and slightly elevated stiffness in the β -sheets and α -helix E.

For the local backbone mobility measured as average rotational correlation time ρ_{av} , which we determined using the residual anisotropy and the average fluorescence lifetime, the result differs. Here, in three positions in α -helices D (86), G (119) and at the end of helix H (150) ρ_{av} changed only marginally, which could be interpreted as still local structure being in present. However, the cyanine dye Alexa647 is well-known for being “sticky” and adhering to protein surfaces (158). This different behavior of donor and acceptor could be seen at position 69 labeled with both donor and acceptor dye, which display under native conditions a very different r_{ss} (Figure 5.3.4C, Table 5.3.2C-D). Comparing the relative ρ_{av} at position 69, we saw a similar decrease for both fluorophores. This indicated that the urea was able to dissolve the protein-dye interaction and thus, the still slow ρ_{av} for the three positions might indeed not be a fluorophore artefact. Hence, α -helices D, G and H might still be folded (Figure 5.2.5A). The time-resolved fluorescence anisotropy showed a diverse picture, but mainly an increase in the faster rotational correlation components (Figure 5.3.4). For example, when the donor is at position 60, the time resolved anisotropy decays much faster in the unfolded protein (Figure 5.3.4B), the dye and linker rotation are increased (Figure 5.3.4D-E, Table 5.3.2C). This leads to an in average strongly reduced steady-state anisotropy. When the acceptor is at position 119, only little change is observed in the time regime of < 5 ns (Figure 5.3.4B). However, close inspection of the fit of both the donor and acceptor samples (Table 5.3.2C-D) shows that at high urea concentrations the resulting rotational correlation times cannot be classified as “dye”, “linker” and “global” motion (Figure 5.3.4B), the time constants found are very heterogeneous when compared to the native state data (Table 5.3.2C-D). This indicates that the dyes are probing distinct environments that could represent a backbone with “native-” or “unfolded-” like characteristics.

This discrepancy between the two methods, the stiffness matrix, which measures average protein segment flexibility, and the rotational mobility of fluorophores, which probes the local backbone mobility, and in the comparison to the older study (157) might be based on two possible reasons: *i*) the averaging over inhomogeneous protein regions (stiffness matrix) or *ii*) the averaging over double mutations containing variants (residual anisotropy). Even, if we used single labeled variants, still the second mutation was inside the protein, which might have influenced the occupancy of the conformational states within the DSE (20), leading to the heterogeneous results observed for variants sharing one of the two label positions (Table

5.3.2C-D). A solution for the first problem might be the use of more label positions, which are placed to cover at least each secondary structure element once. Nevertheless, the outcome of both methods agrees with previously published results, which identified the NTsD as being unfolded easily (20, 21, 41-44). Thus, an in average lower stiffness and higher mobility would be anticipated there, which is exactly what we found (Figure 5.2.4).

5.2.3.3. Denatured T4L shows native-like mean structure

From fluorescence anisotropy measurements we found hints for restricted mobility for some of the fluorophores position, which is an indication for limited backbone mobility of the amino acid chain. Additionally, on the length-scale of FRET (20-80 Å) the denatured T4L cannot be described by polymer-like models. We interpret these unanticipated observations by differences in the stiffness. Knowing that all secondary structure is dissolved, we ask whether preferential arrangements of transiently formed conformations exist, which describe all experimentally measured FRET-distances. These distances between fluorophores can be interpreted by computational methods in terms of structural models. Here, structural pre-knowledge is combined with the experimental restraints. Typically, secondary- and tertiary structure is available (i.e. by crystallization of sub-domains) and thus, high resolution models can be determined (46). In case of the DSE, we do not have the luxury of structural pre-knowledge. Thus, we aim to determine mean position of fluorophores as opposed to a full structural model. The challenges in determining the probability distribution of mean fluorophore positions experimentally are well known (159).

In our distance network not all positions are fully trilaterated (incomplete network). Given our incomplete network, we anticipate ambiguous solutions of the mean fluorophore position. **Native State.** To study this effect, we used the experiments of the native state, where vast structural pre-knowledge is available. Due to the size of the mean fluorophore position distribution it is evident that the mean fluorophore positions are not well resolved by the FRET-restraints alone (Figure 5.2.7A). Next, we used as structural model an X-ray structure of T4L, which we dissected into four fragments (helix A, NTsD, helix C, CTsD, (see SI 5.3.1.10)). Using these fragments, which were held together by a flexible peptide bond, as rigid bodies, we tested, which mean fluorophore positions are possible within the docked structural ensemble. Similarly to the FRET-restraints alone, in the docked structures the fluorophore positions are widely distributed. However, surprisingly a combination of both resolves the mean fluorophore positions in space (Figure 5.2.7B). This is highlighted by

appearing peaks in Table 5.3.8. Additionally, a unique structural ensemble is resolved with Angström resolution. This demonstrates the emergent properties of a combination of simulations and experimental restraints, which stabilize recovered models and add new information (Figure 5.3.11, Table 5.3.8).

Denatured State. As the DSE was studied using the same distance network, we face the same problems. This illustrated in Figure 5.2.7C. Hence, by the current FRET-restraints alone we cannot hope to probe for structural features. Thus, here we made the reasonable assumption that the amino acid chain, which links neighboring labeling sites, behaves like a WLC. Here, we use the experimentally determined average stiffness of $\kappa_{avg} = 0.86$ and the calculated length given the separation in number of amino-acids ($L = 3.6 \text{ Å} * n$, the flexibility of the dye-linkers were accounted for within FPS, SI 5.3.1.10). For all but two labeling positions (8 and 19) this simple assumption stabilized our solutions of possible mean fluorophore positions significantly (Figure 5.2.7D-E). Thus, as in the native state, also in the DSE the combination of the introduced WLC-restraints and the FRET-restraints generates additional information not contained in neither of them individually (Figure 5.2.7F, Table 5.3.9).

In agreement with the loss in secondary structure and the anticipated loose packing of the amino-acid chain, highlighted by the short correlation times in FCS of approximately 100 ns (Figure 5.2.6), the distribution of the mean fluorophore positions within the DSE are markedly larger than for the native state. This is in agreement with the increased hydrodynamic radius (from $\sim 14 \text{ Å}$ to $\sim 36 \text{ Å}$) determined from the translational diffusion time in the filteredFCS experiments. Additionally, we find that the mean fluorophore positions of neighboring amino-acids within one subdomain are in close proximity, while the mean fluorophore positions of the CTsD and the NTsD are well separated (Figure 5.2.7E).

Previously, a similar behavior has been found for other small proteins by simulations (villin, TrpZip, BBA5, (160)) and experiments (Eglin C, staphylococcus nuclease (17, 161)). The mean fluorophore location of position 8 and 19 are apparently quite undefined either due their proximity to the end of the chain, missing experimental restraints or due their flexibility. The latter is corroborated by filteredFCS measurements, where 8-86 and 19-86 show the shortest correlation times. Our findings agree also well with previous findings, which showed that the NTsD is less stable than the CTsD and unfolds easily (20, 21, 41-44). Note, that the mean fluorophore positions of the DSE represent a subset of a large ensemble of conformational states, which are in dynamic exchange resulting in the observed complex distribution of correlation times (Figure 5.2.6).

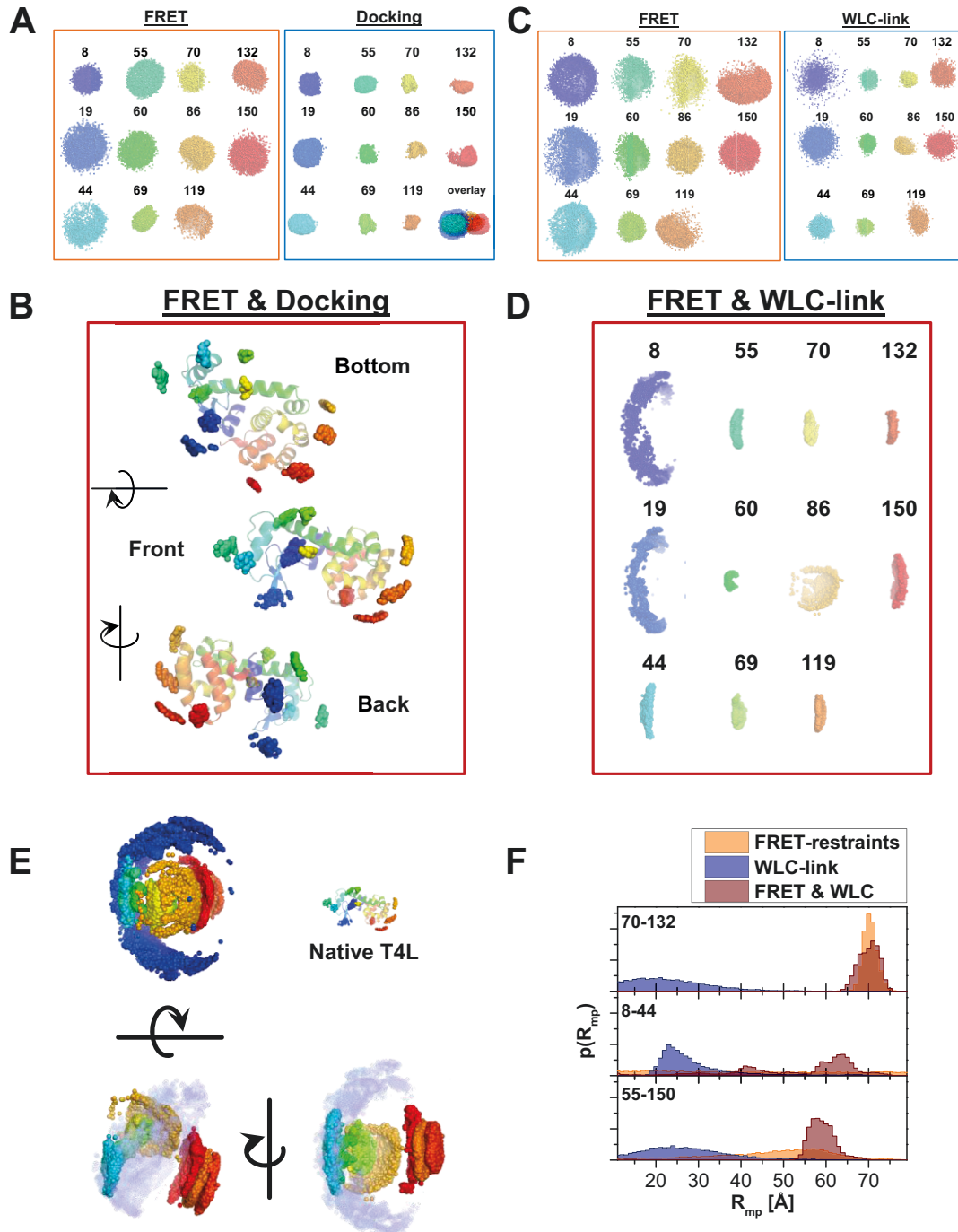


Figure 5.2.7 Mean fluorophore position distribution for the native (A-B) and denatured state (C-F). **Native state:** (A) FRET restraints alone (left) and restraints by simulation model alone (right). (B) Combination of FRET- and simulation model restraints resolves the structural model. **Denatured state:** (C) FRET restraints alone (left) and restraints by WLC links alone (right). (D, E) Combination of FRET- and WLC- restraints reduces spatial distribution of possible mean fluorophore positions. (F) Distance distributions for three selected (label-) pairs resulting from the individual restraints or the combination. Note, that only 70-132 was a measured distance, the other two were indirectly obtained. Details of modeling can be found in 5.3.1.10 and 5.3.2.9.

Two types of future experiments could improve the resolution of our model: *i*) intra-subdomain FRET-measurements using fluorophores with small Förster-radii could provide a more direct proof of ordered architecture of the DSE. *ii*) We could improve the current

labeling network with additional long-distance constraints, which might increase the resolution of the mean fluorophore positions and unequivocally answer the question whether the denatured CTsD and NTsD are well separated.

5.2.3.4. *Correlation analysis reveals the roughness of the energy landscape within the DSE*

In the fullFCS and filteredFCS measurements of the DSE of T4L we found relaxation time constants t_R in the sub- μ s, μ s and ms range (Figure 5.2.6). It is known that the time scales of the detected kinetics are related to the observed type of motions within the protein (4, 19, 128, 156). Fast chain dynamics occurs in the ns timescale, whereas the formation of transient structure takes place in the low μ s region. The concerted motion of smaller and larger partly or completely folded domains takes place in the longer μ s to ms timescale (Figure 5.2.8).

FullFCS measurements had the largest dynamic range, and by comparing single-label variants with double-labeled variants, we could decipher, which fluorescence fluctuations were due to local dynamics and which were based on longer range motions. In the Donly-samples, in which only fluctuations based on changes in the fluorophore environment or by quenching due to aromatic amino acids would be expected (131, 133), we found up to three relaxation time constants in the 10 μ s to \sim 1 ms region. This indicates that larger, and thus slower, motions within T4L are necessary to change the local fluorophore environment significantly. The *DA-CCFs* probed either very fast (few 100 ns) or very slow dynamics (\sim 1 ms). The fast dynamics might reflect the chain dynamics of the unfolded polymer chain. Unexpectedly, t_R did not depend upon the chain length between the labeling sites (Figure 5.3.8). Previous studies showed that the longer the chain, the faster the chain reconfiguration gets (130, 148, 162). However, it was also determined that the “loose ends” (i.e. residues before the first and after the second labeling site) might influence the measurements (16, 162). These differ for all our variants. The slower t_R might probe the diffusion from one unfolded well to another (Figure 5.2.8).

In filteredFCS, we zoomed in on all intermediates, which are on the way of a high FRET to a low FRET transition and hereby discarded the local fluorophore effects. We found in both, the native and unfolded state, four t_{rel} , whose amplitudes differed markedly among the variants in the native state, but not in the DSE. The fastest t_{rel} found with an amplitude of \sim 80 % resembles probably again the chain dynamics already found with the *DA-CCFs*. Also, here the anticipated chain length dependency was not detected. A disturbing factor might be

in both measurements the flexible fluorophore linker, which moves freely in the same timescale as an unfolded peptide chain. Although, two of the t_{rel} , which have been linked under native conditions to distinct steps in the enzymatic cleavage cycle of T4L, could also be found under denaturing conditions, we believe that within the DSE they reflect rather general faster or slower proteins motions on the sub-domain or global protein level (4, 19, 128, 156).

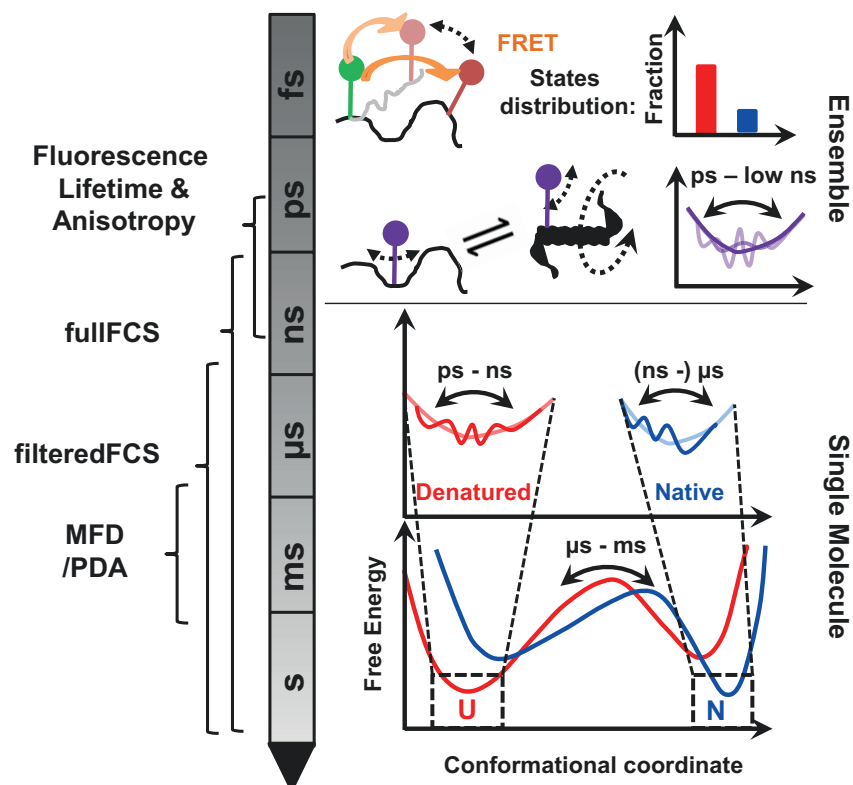


Figure 5.2.8 Energy landscape and the dynamics of the DSE. Fluorescence lifetime probes the static distribution of states, whereas fluorescence anisotropy experiments – preferably conducted with short linker dyes – describe the local, backbone dynamics of the sample. Single-molecule experiments in terms of multiparameter fluorescence detection (MFD) and photon distribution analysis (PDA) show the population distribution and its dynamics on timescales similar to burst duration (~ 1 ms). Fast local and global dynamics on timescales shorter than diffusion time can be resolved with fluorescence correlation methods (fullFCS and filteredFCS) from both single and double labeled molecules

Nevertheless, we could show here the complex energy landscape of a chemically denatured protein and by the use of a small network of single and double labeled variants as well as the specialized filteredFCS methods, we could disentangle local changes from global motions. A more stable fit result for the relaxation time constant or the amplitudes, and thus the chain length dependence of the chain reconfiguration time, might be obtained when fitting globally over all variants and not only (sub-)globally the correlation curves for each variant.

5.2.3.5. Outlook

Our goals in this study were to use a network of 24 FRET-pairs to decipher the polymer-behavior of the ensemble of denatured states of chemically denatured T4L, to identify possible nucleation sites for folding and finally describe the dynamics within the DSE. We combined site-specific ensemble fluorescence lifetime and anisotropy measurements with single-molecule MFD analysis and correlation methods. By comparing different fit models for our ensemble measurements, we could show that the DSE of T4L apparently behaves like a non-ideal WLC, but the obtained parameters are meaningless; thus, also due to the observed heterogeneity in the MFD plots, a better description might be obtained when introducing a distribution of stiffnesses to properly consider the changes in structural content of the individual conformations. Nevertheless, we could identify stiffer region in the chain of T4L, which might hint to possible nucleation sites. Using apparent distance restraints, we could further show that based on fluorophore distributions and further assumption on protein backbone behavior chemically denatured T4L has a native-like mean structure, albeit larger in overall size. Finally, we probed the rough energy landscape of the complex local and global kinetics within the DSE by correlation methods.

Overall, we could resolve the full heterogeneity of a protein's denatured state. This multidimensional approach could be used to characterize intrinsically disordered proteins and reveal the formation of transient secondary structure elements and the kinetics across the energy landscape. Additionally, a (extensive) data set was generated which can be used by simulation experts to screen their generated unfolded structures for an ensemble of most likely structures and thus fill the empty space between our fluorophore distributions. However, we could also show that the view of chemically denatured proteins as “simple” polymers is far too simplistic and care should be taken, when the behavior of DSEs or IDPs is interpreted in such ways.

5.2.4. Materials & Methods

Sample Preparation. Mutagenesis of the pseudo-wildtype (wt*) T4L gene in the pET11a vector was done according to standard procedures. Protein production, purification and labeling were performed as described elsewhere (94, 95) (Chapter 3). As FRET pair, the donor Alexa488 and the acceptor fluorophore Alexa647 (Life Technologies, USA) were used. More detailed measurement and data analysis protocols for all experiments can be

found in the supplementary information (5.3.1), a tabled summary of fitting results in SI Section 5.3.2.

Time-resolved fluorescence & anisotropy. Ensemble time-resolved measurements were performed either using an IBH-500U (IBH, Scotland) or a Fluotime 200 (Picoquant, Germany). Single and double labeled samples were both measured up to a peak count of 100'000 photons. By convolution with the instrument response function, the obtained fluorescence decay curves were fitted to several models (multi-exponential decay (MEx), Gaussian-distributed distances (R_{DA}), worm-like chain model (WLC)). Fluorescence lifetime measurements were conducted under magic angle conditions (excitation polarizer 0° , emission polarizer 54.7°) and for fluorescence anisotropy measurements with vertical and horizontal detection. Additionally, the steady-state fluorescence anisotropy and the g-factor of the system were determined.

Single-molecule fluorescence. Single-molecule fluorescence measurements (MFD) were performed at room temperature in aqueous solution (50 mM Na-phosphate, 150 mM NaCl, 0 or 7.5 M urea) as described elsewhere (Chapter 3) except that no TROLOX was added. For Photon Distribution Analysis (PDA) the MFD data was used (48, 63-65). Here, we applied PDA in two different modes, static (sPDA, 1 ms time windows (TW)) and dynamic (dPDA, global fit of 2 ms, 3 ms and 4 ms TW's). In sPDA, the data was fitted to a Gaussian distributed distance $\langle R_{DA} \rangle_E$, whereas dPDA assumes two dynamically exchanging states $\langle R_{DA} \rangle_{E,1}$ and $\langle R_{DA} \rangle_{E,2}$ and an additional static fraction. Both models include also a fraction of FRET inactive molecules.

Fluorescence correlation spectroscopy. All fullFCS measurements were conducted on a confocal fluorescence microscope (Olympus IX81, Hamburg, Germany). As excitation sources the parked beam of an Ar-ion laser (488 nm) or a laser diode (635 nm) (both Olympus) were used, both lasers were operated in continuous wave mode. The laser light was directed into a 60x water immersion objective (NA < 1.2) by a dichroic beam splitter and focused in the sample, illuminating a diffraction limited excitation volume. The emitted light was collected by the same objective and separated into the two polarizations (parallel and perpendicular, PBS 101, Thorlabs) relative to the exciting laser beam. The fluorescence signal is further divided into two spectral ranges (630 DCXR, Chroma), equipped with bandpass filters (HC 520/35 (Semrock) and HQ 715/120 (Chroma)) for donor (Alexa488) and acceptor (Alexa647) fluorescence, respectively. The signal from single photon sensitive detectors (MicroPhotonDevices PD5CTC for the green channels and PerkinElmer AQR-14

for the red channels) was recorded photon-by-photon in time-tagged-time-resolved (tttr) mode with picosecond accuracy (HydraHarp400, Picoquant, Berlin, Germany). Samples were diluted in NUNC chambers (Lab-Tek, Thermo Scientific, Germany) in a total of 500 μL volume in the respective measurement buffer containing 50 mM Na-phosphate, 150 mM NaCl and 7.5 M urea. Additionally, 1 μM of unlabeled T4L was added to coat the chamber surface. The sample concentration of single and double labeled protein was adjusted in the low nM range, yielding \sim three molecules in the focus at a time. Single labeled variants were measured for 2.5 hrs, double-labeled variants for 10 hrs.

The data was correlated in a time-window of 32 ps to 2 s and fitted with home written software to standard FCS models containing a diffusion term, a photon antibunching term, a term describing the rotational correlation times and additional bunching (autocorrelation functions, *ACF*) or antibunching (cross correlation functions, *CCF*) terms.

Filtered FCS. Filtered FCS measurements (47, 70, 72) were performed at room temperature in aqueous solution (50 mM Na-phosphate, 150 mM NaCl, 0 to 7.5 M urea) as described elsewhere (Chapter 3). The data was collected for 3 hrs, correlated by applying species-specific filters in a time-window of 16 ns up to 4 s and fitted to a similar model as described for FCS consisting of a diffusion term and several correlation (species cross correlation functions, *sCCF*) and anti-correlation terms (species auto correlation functions, *sACF*).

5.3. Supplementary Information

5.3.1. Methods

5.3.1.1. Experimental Design

We use our T4L network of FRET pairs developed in a previous work (Chapter 3) to study the behavior of proteins under strongly denaturing conditions (7.5 M urea). All variants – single and double labeled – are firstly measured under ensemble, equilibrium conditions using ensemble Time Correlated Single Photon Counting (eTCSPC) to characterize the samples' ensemble behavior and the local mobility of the individual segments on the ns time scale.

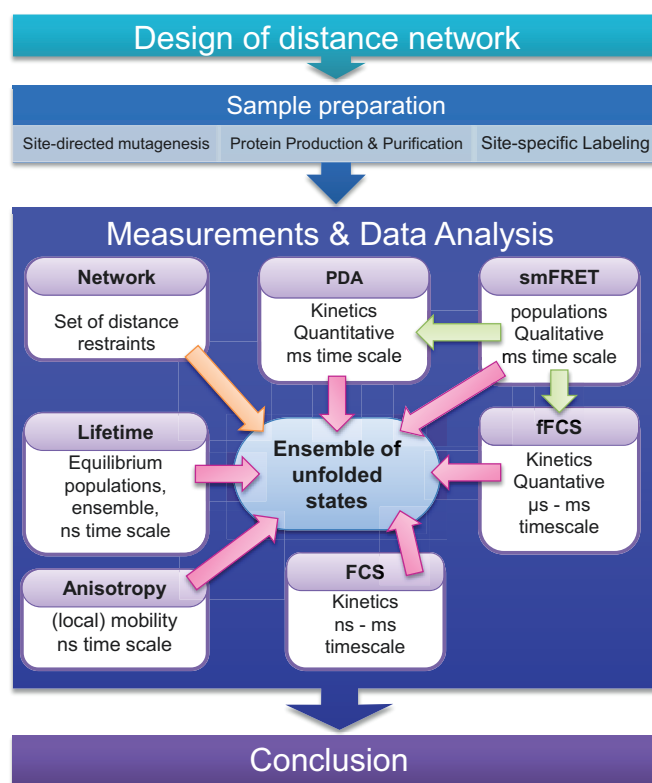


Figure 5.3.1 Flowchart of data acquisition and data analysis. The T4L network of variants generated within previous work (Chapter 3) are subjected to the following experiments: Time-resolved fluorescence lifetime and anisotropy analysis using ensemble Time correlated Photon Counting (eTCSPC) and multiparameter fluorescence detection (MFD). Photon distribution analysis (PDA) was used to analyze kinetics in the ms time range. Additionally, of selected variants fluorescence correlation spectroscopy (FCS) and *filtered* FCS measurements were conducted to resolve kinetics in the ns – ms (μ s – ms) time range.

Single-molecule measurements using multiparameter fluorescence detection (MFD) reveal the sample heterogeneity and Photon Distribution Analysis (PDA) of single-molecule bursts using multiple timewindows allows the identification of kinetics in the ms time range. Finally, full fluorescence correlation spectroscopy (FCS) was used to resolve the kinetics in the ensemble of unfolded state (DSE) in the ns to low ms time scale. A “zoom-in” on kinetic

processes influencing FRET was achieved by the use of *filtered* FCS (fFCS). A schematic overview for a better illustration is given in Figure 5.3.1.

5.3.1.2. Protein production, purification & labeling

The T4L variants were produced in *E. coli* and purified as described previously (Chapter 3). Labeling was done site-specifically using Alexa488-hydroxylamine as donor and Alexa647-maleimide as acceptor dye according to the protocol given in (94). 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.).

5.3.1.3. Ensemble Time Correlated Single Photon Counting with high precision

Data acquisition

Ensemble Time Correlated Single Photon Counting (eTCSPC) measurements were performed using either an IBH-500U (IBH, Scotland) or a Fluotime 200 (Picoquant, Germany) system.

The excitation source of the IBH machine were a 470 nm diode laser (LDH-P-C470, Picoquant, Germany) operating at 10 MHz for donor excitation and a 635 nm (LDH-(-1-126, Picoquant, Germany) for acceptor excitation. The excitation and emission slits were set to 2 nm and 16 nm, respectively. The excitation source of the Fluotime200 system was a white light laser (SuperK extreme, NKT Photonics, Denmark) operating at 20 MHz for both donor (485 nm) and acceptor (635 nm) excitation with excitation and emission slits set to 2 nm and 5 nm, respectively. Additionally, in both systems, cut-off filters were used to reduce the amount of scattered light (>500 nm for donor and >640 nm for acceptor emission).

For green detection, the monochromator was set to 520 nm and for red detection to 665 nm. All measurements were conducted under magic angle conditions (excitation polarizer 0°, emission polarizer 54.7°), except for anisotropy where the position of the emission polarizer was alternately set to 0° (VV) or 90° (VH).

In the IBH system, the TAC-histograms were recorded with a bin width of 14.1 ps within a time window of 57.8 ns, while the Fluotime200 was set to a bin width of 8 ps within a time window of 51.3 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* (~230 ps FWHM for the IBH, ~150 ps for the Fluotime200) 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.

For the IBH system, it was needed to perform before each measurement session a reference measurement with a continuous light signal 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.

Fitting procedure

Given different model functions described below the experimental fluorescence intensity decays F_{exp} 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, if necessary, 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 *Lin* (139). Given these corrections the experimental time-resolved fluorescence intensities of the samples F_{sample} are proportional to:

$$F_{sample}(t) \propto (F_{exp}(t) \otimes IRF + c) \cdot Lin \quad 5.3.1$$

Fluorescence lifetime and quantum yields of single labeled variants

The fluorescence decay of the single donor or acceptor labeled variants was multi-exponential, most likely due to local quenching and steric effects. To account for these effects, the single labeled variants were fitted by multi-exponential relaxation model:

$$F_{D(0)/A(0)}(t) = \sum_i x^{(i)} \exp(-t/\tau^{(i)}) \quad 5.3.2$$

Here, $x^{(i)}$ are pre-exponential factors and $\tau^{(i)}$ the fluorescence lifetimes. The species-averaged lifetime $\langle \tau \rangle_x$ and fluorescence averaged lifetime $\langle \tau \rangle_f$ of the sample are calculated as:

$$\langle \tau \rangle_x = \sum_i x^{(i)} \cdot \tau^{(i)} \quad \text{and} \quad \langle \tau \rangle_f = \frac{1}{\langle \tau \rangle_x} \sum_i x^{(i)} \cdot (\tau^{(i)})^2 \quad 5.3.3$$

Assuming only dynamic quenching, the fluorescence quantum yield of dyes Φ_{Dye} can be determined using the following equation:

$$\phi_{Dye} = \frac{\langle \tau \rangle_{x,Dye} \cdot \phi_{ref}}{\langle \tau \rangle_{x,ref}} \quad 5.3.4$$

As reference, single labeled DNA with either Alexa488 ($\Phi_{D0} = 0.8$, $\langle \tau \rangle_{x,D0} = 4$ ns) or Cy5 ($\Phi_{A0} = 0.32$, $\langle \tau \rangle_{x,A0} = 1.17$ ns) were used.

Data analysis of double-labeled variants

Determination of the Förster radius R_0

Depending upon the distance R_{DA} between the donor and the acceptor fluorophore in the biomolecule, the donor fluorescence might to be quenched due to Förster Resonance Energy Transfer (FRET) which results in a decreased species-weighted average fluorescence lifetime of the DA-sample when compared to the DOnly labeled sample and the appearance of measurable acceptor fluorescence. The amount of quenching, i.e. the FRET-efficiency E , is highly distance dependent as given by the Förster formula:

$$E = (1 + R_{DA}^6 / R_0^6)^{-1} \quad 5.3.5$$

with R_0 being the Förster radius, which depends on the refractive index of the surrounding medium η , the spectral overlap J between the donor emission and the acceptor absorption spectrum, the fluorescence quantum yield of the donor Φ_{D0} and the orientation factor κ^2 . The fluorophores are coupled to the biomolecule by a long, flexible linker. Therefore, we assume isotropic averaging of donor-acceptor dipole orientation and we use the approximation $\kappa^2 \approx 2/3$. This is supported by the anisotropy studies shown in (Chapter 3, herein Table 3.3.3). We found that η and J depend on the urea concentration, whereas Φ_{D0} is sample-dependent. The rate constant of the FRET-process is independent of the donor fluorescence quantum yield. Therefore, we define a reduced Förster-radius R_{0r} solely dependent on the spectral overlap, the orientation factor and the refractive index:

$$R_{0r} = \left[\frac{9(\ln 10)}{128\pi^5 \cdot N_A} \cdot \frac{J \cdot \kappa^2}{\eta^4} \right]^{\frac{1}{6}} = 0.2108 \cdot \text{\AA} \cdot \left[\frac{\kappa^2}{\eta^4} \cdot \left(\frac{J(\lambda)}{\text{mol}^{-1} \cdot \text{dm}^3 \cdot \text{cm}^{-1} \cdot \text{nm}^4} \right) \right]^{\frac{1}{6}} \quad 5.3.6$$

with N_A being Avogadro's constant. R_{0r} is 56 Å in the absence of urea and 54 Å in buffer containing 7.5 M urea. To precisely analyze the contribution of quenching due to FRET and distinguish from the possibly present local quenching of the donor, we analyze the DA-

sample globally with its respective DOnly-sample. The FRET-induced donor quenching can be described by different models adapted to the particular question. Here we use two different models: a formal description by Gaussian distributions and a Worm-like chain model.

We account for donor labeled molecules lacking an acceptor by including the fraction of DOnly- molecules (x_{DOnly}) in the analysis model using a modified version of eq. 5.3.1:

$$\begin{aligned} F_{FRET}(t) &\propto \left[(1 - x_{DOnly}) F_{D(A)}(t) + x_{DOnly} F_{D(0)}(t) \right] \otimes IRF + c \cdot Lin \\ F_{Ref}(t) &\propto \left(F_{D(0)}(t) \otimes IRF + c \right) \cdot Lin \end{aligned} \quad 5.3.7$$

Gaussian distributed distances-fitting (RDA)

The dyes are coupled to the biomolecules of interest via long and flexible linkers (~ 20 Å). This has to be accounted for in the analysis. The flexible linking assures that the dyes move freely. The inter-dye correlation time is approximately 100-200 ns (140). This is significantly faster than the integration time of the experiment but slower than the time-scale of fluorescence (4 ns). Therefore, the broadening of the distances distributions due to the linkers have to be considered. The distance R_{DA} between the FRET pair in one conformational state can then not be described by one distance but is instead modeled as a probability function $p(R_{DA})$ showing Gaussian distribution characteristics with a mean distance $\langle R_{DA} \rangle$ and width w_{DA} :

$$p(R_{DA}) = \frac{1}{w_{DA} \sqrt{\pi/2}} \exp \left(-2 \left[\frac{R_{DA} - \langle R_{DA} \rangle}{w_{DA}} \right]^2 \right) \quad 5.3.8$$

$$F_{D(A)}(t) = F_{D(0)} \cdot \int_{R_{DA}} p(R_{DA}) \cdot \exp \left(-t \cdot k_0 \left[1 + (R_0 / R_{DA})^6 \right] \right) dR_{DA} \quad 5.3.9$$

k_0 is the radiative rate constant of fluorescence ($k_0 = 0.224 \text{ ns}^{-1}$). The linkers used in the course of this work have a length of ~ 20 Å and it was estimated using FPS that a half-width of ~ 10 Å is to be expected in the folded protein (53).

Polymer model (GC & WLC)

Another class of fitting models used to describe a special case of biomolecules are polymer models. Derived from polymer theory, they are applied when it is likely to assume an unfolded biomolecule, i.e. under highly denaturing conditions, e.g. in denaturants like urea and guanidine hydrochloride. Under those conditions, the majority, but by no means all, of the tertiary and even secondary interactions are lost and the protein behaves as a polymer. Given the probability distribution function of the donor-acceptor distances $p(R_{DA})$ the mean transfer efficiency $\langle E \rangle$ is obtained by:

$$\langle E \rangle = \frac{\int_0^\infty E(R_{DA}) p(R_{DA}) dR_{DA}}{\int_0^\infty p(R_{DA}) dR_{DA}} = \frac{\int_0^\infty (1 + R_{DA}^6 / R_0^6)^{-1} p(R_{DA}) dR_{DA}}{\int_0^\infty p(R_{DA}) dR_{DA}} \quad 5.3.10$$

Here, $E(R_{DA})$ is the distance-dependent FRET which can be described by the Förster formula and $p(R_{DA})$ is the probability distribution of the polymer chain of having a population with the distance R_{DA} between the two dyes. The exact form of $p(R_{DA})$ depends on the assumed polymer model. In the simplest case, the peptide chain shows random coil or freely joint chain characteristics where the fluorescence decay can be approximated by the Gaussian Chain Model (14). This model assumes that each bond has a fixed length and that adjacent bond angles are uncorrelated. This results in a Gaussian distribution:

$$p(R_{DA}) = \frac{4\pi^2}{(2/3 \cdot \langle R_{DA}^2 \rangle)^{3/2}} \exp\left(-\frac{3R_{DA}^2}{2\langle R_{DA}^2 \rangle}\right) \quad 5.3.11$$

With $\sqrt{\langle R_{DA}^2 \rangle} = l \cdot n^{1/2}$ being the root-mean-square, end-to-end length of the chain (16). l is the bond length (3.6 Å per amino acid residue) and n the number of bonds between the two dyes. Thus, n is the only free fitting parameter.

However, for the model protein T4L used here, fitting with the above described model could not be achieved with a satisfactory result, i.e. flat distribution of weighted residuals and a $\chi^2 \sim 1$. Thus, it was necessary to use the more complex model, Worm-like chain. Here, the biopolymer is assumed to be a semi-flexible rod of length L and stiffness κ and is characterized by a so-called persistence length l_p , which is defined as the product of L and κ . Instead of a fixed segment length based on the monomers of the polymer, the chain is thought of being comprised of longer, uncorrelated links (12) and the persistence length is defined as this length over which the correlation between the bond angles perseveres. The bending energy E associated with a chain configuration $r(l)$ is given by:

$$\beta \cdot E = \frac{\kappa}{2} \int_0^L \|\partial_l t\|^2 dl \quad 5.3.12$$

where $\beta = 1/k_B T$ is the inverse temperature, l is the arclength, and $t = \delta_l r$ is the unit tangent vector ($\|t\| = 1$) (141). Unfortunately, it lacks an exact numerical solution and the approximations summarized and derived in (141) were thus implemented in the fitting routine. Inserting the given constants a , b , c , $(c_{ij})_{i,j}$ and d into equation 21 of the paper from Becker *et al* (141), leads to equation 5.3.13:

$$p(R_{DA}) = \frac{1 - c \cdot R_{DA}^2}{(1 - R_{DA}^2)^{5/2}} \cdot e^{\left(\frac{d \cdot \kappa \cdot a \cdot b(1+b)}{1 - (bR_{DA})^2} R_{DA}^2 \right)} \cdot e^{\left[\left(\frac{3}{4} \frac{1}{\kappa} - \frac{1}{2} \right) R_{DA}^2 + \left(\frac{23}{64} \frac{1}{\kappa} + \frac{17}{16} \right) R_{DA}^4 + \left(-\frac{7}{64} \frac{1}{\kappa} - \frac{9}{16} \right) R_{DA}^6 \right]} \cdot \frac{1}{1 - R_{DA}^2} \cdot I_0 \left(\frac{-d \cdot k \cdot a \cdot (1+b) \cdot R_{DA}}{1 - (bR_{DA})^2} \right) \quad 5.3.13$$

with:

$$a = 14.054$$

$$b = 0.473$$

$$c = 1 - (1 + (0.38 \cdot \kappa^{-0.95})^{-5})^{-1/5}$$

$$d = \begin{cases} 1 & \kappa < 1/8 \\ 1 - \frac{1}{0.177 / (\kappa - 0.111) + 6.40(\kappa - 0.111)^{0.783}} & \text{otherwise} \end{cases}$$

I_0 : Bessel - function of Order Zero

WLC model with explicit “linker” treatment

The WLC-model describes our biomolecule as a homogeneous system with a specific length and flexibility (\sim stiffness). However, to be able to measure this for different segments of the protein, we selectively coupled fluorophores to the two end points of the probed segment. These fluorophores are coupled via long, flexible linkers, whereas in the biomolecule residual structure (and/or tertiary interactions) is suspected. This was indicated by two facts (i) the fluorescence intensity decay curves could not be modeled with the Gaussian chain model and (ii) the obtained chain length L from the WLC model were much shorter than the theoretical expected one ($L_{theo} = n \cdot 3.6 \text{ \AA}$).

Hence, we decided to decouple the fluorophores from the biomolecule motion (WLC-like).

In accordance with the concept of accessible volumes, we describe the spatial probability distributions of the dyes by 3-dimensional Gaussians with uniform width for the donor σ_D and the acceptor σ_A ($\sigma_D = \sigma_A = 6 \text{ \AA}$) (53). The resulting distance distribution $p(R_L, R_{DA})$ between the donor and the acceptor for a single conformation with a separation distance R_L is given by eq. 5.3.14:

$$p(R_L, R_{DA}) = \begin{cases} \text{if } R_{DA} > 0 & \frac{R_L}{R_{DA}} \cdot [N(R_L, R_{DA}, \sigma_{AV}) - N(R_L, -R_{DA}, \sigma_{AV})] \\ \text{if } R_{DA} = 0 & 2 \left(\frac{R_L}{\sigma_{AV}} \right)^2 N(R_L, 0, \sigma_{AV}) \end{cases} \quad 5.3.14$$

$$N(R_L, R_{DA}, \sigma_{AV}) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{(R_L - R_{DA})^2}{2 \cdot \sigma_{AV}^2} \right)$$

$$\sigma_{AV} = \sqrt{\sigma_{Donor}^2 + \sigma_{Acceptor}^2}$$

Now, we consider for the WLC-distribution ($p_{WLC}(R_{DA})$) the dye-linker effects by eq. 5.3.15:

$$p(R_{DA}) = \int p_{WLC}(R_{DA}) \cdot p_L(R_L, R_{DA}) dR_{DA} \quad 5.3.15$$

5.3.1.4. Ensemble Fluorescence Anisotropy

The time-resolved anisotropy was analyzed as described in (53). Briefly, the time-resolved anisotropy decay $r(t)$ can be modeled by a three exponential decay:

$$r(t) = r_{\text{global}} \cdot \exp(-t / \rho_{\text{global}}) + r_{\text{lin ker}} \cdot \exp(-t / \rho_{\text{lin ker}}) + r_{\text{dye}} \cdot \exp(-t / \rho_{\text{dye}}) \quad 5.3.16$$

The three rotational correlation times ρ_i are related to the different motions the dyes are exposed to: Fastest motion is the rotation of the dye itself (< 0.1 ns), the motion of the dye-linker takes place at ~ 1 ns time scale, while the global protein rotation is slowest and depends upon protein size and label position (> 5 ns). The pre-exponential factors describe the extent to which each of the motions are probed by the label and their sum is normalized to the fundamental anisotropy r_0 (Alexa488: $r_0=0.375$, Alexa647: $r_0=0.39$). Experimentally, the time-resolved fluorescence decay of vertically ($I_{VV}(t)$) and horizontally ($I_{VH}(t)$) polarized light emitted from the sample after vertical excitation is measured and the anisotropy decay is calculated as:

$$r(t) = \frac{I_{VV}(t) \cdot (1 - r_{ss})}{I_{VH}(t) \cdot (2r_{ss} + 1)} \quad 5.3.17$$

Here, r_{ss} is the residual anisotropy measured under steady-state conditions. Using r_{ss} and $\langle \tau \rangle_f$, we used the Perrin-equation to calculate the average rotational correlation time ρ_{av} of our samples:

$$\rho_{av} = \frac{\langle \tau \rangle_f}{(r_0 / r_{ss}) - 1} \quad 5.3.18$$

5.3.1.5. Multiparameter Fluorescence Detection

Data acquisition and analysis for MFD was done as described elsewhere (Chapter 3)(25). As measurement buffer PBS, pH 7.5, containing either no (“0 M”) or 7.5 M urea (“7.5 M”) were used. Additionally, 1 μ M of unlabeled T4L was added to coat the surface of the measurement chamber and thus, minimize absorption of the labeled protein to it.

5.3.1.6. Guidelines for reading MFD histograms

To be able to properly read and thus understand MFD histograms correctly, one has to take care of some guidelines. A short list is presented here and for more detailed information the reader is referred to (25).

I) Donor only population is shown at high F_D/F_A ratio with lifetime ~ 4 ns (donor-only for Alexa488).

II) High FRET appears at shorter lifetimes when the fluorescence of acceptor is high ($F_D/F_A \rightarrow 0$).

III) Static FRET states distributions have elliptical 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 (163). *IV)* Static FRET states follow a theoretical line that accounts for dye linker mobility called "static FRET line" Eq. 3.3.3 (163).

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 averaged lifetime $\langle \tau_{D(A)} \rangle_f$. Therefore, dynamics in sub-ms range is expected when the shape is not elliptical. 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.

5.3.1.7. Photon Distribution Analysis

Photon or Probability distribution analysis (PDA) uses the photon traces of single-molecule measurements to deduce various FRET or fluorescence anisotropy related parameters (64) and displays them in occurrence histograms. The build histograms are then compared with theoretical distributions of the respective parameters (probability histograms). However, critical care has to be taken for used values of crosstalk, background, and fluorescence quantum yields (48). Thus, it is possible to distinguish an extra broadening of histograms due to measurement artefacts from actual broadening due to sample heterogeneities. Next to this so-called "static" analysis of the data, it is possible also to detect and describe simple dynamics in a sample. This is achieved by time-window analysis (TWA). Here, the photon trace is binned in different time-windows (TW's), i.e. 0.5 ms, 1 ms or even 5 ms (65). If a system undergoes dynamic exchange between two or more states in the time range of the diffusion time (\sim low ms range), which also limits the usable minimal/maximal timewindow for a sample, this can be easily detected by eye with TWA. The dynamics of this time range will induce a time-window dependent narrowing/broadening of obtained e.g. FRET-efficiency histograms. A detailed theoretical deduction can be found in (48, 64, 65).

In the course of this work, static PDA was used to estimate the mean and the half-width of averaged on ms time scale distances distribution between the dyes attached to the sample.

The mean distance was modeled as being Gaussian distributed with a limited half-width, i.e. the half-width was given in % of the distance and thus rising with an increasing distance between the FRET pair. Besides, a fraction of DOnly-labeled molecules was included. Additionally, in dynamic PDA, using TWA of three different TW's, a simple two state model was used to describe possible sample dynamics in the ms time range:

$$\langle R_{DA} \rangle_{E1} \xrightleftharpoons[k_{21}]{k_{12}} \langle R_{DA} \rangle_{E2} \quad 5.3.19$$

with $t_{rel} = 1/(k_{12}+k_{21})$ being the equilibrium constant of the exchange rate constants. $\langle R_{D(A)} \rangle_{E1}$ and $\langle R_{D(A)} \rangle_{E2}$ are mean distances of state 1 and 2, respectively. Furthermore a third, static state $\langle R_{D(A)} \rangle_{E3}$ and a fraction of DOnly-labeled molecules was included. The states were modeled same as for the static case.

5.3.1.8. Full Fluorescence Correlation Spectroscopy

Data acquisition

All fullFCS (from ps to s) measurements were conducted on a confocal fluorescence microscope (Olympus IX81, Hamburg, Germany). As excitation sources the parked beam of an Ar-ion laser (488 nm) or a laser diode (635 nm) (both Olympus) were used. Both lasers were operated in continuous wave mode. The laser light was directed into a 60x water immersion objective (NA = 1.2) by a dichroic beam splitter and focused in the sample, illuminating a diffraction limited excitation volume. The emitted light was collected by the same objective and separated into the two polarizations (parallel and perpendicular, PBS 101, Thorlabs) relative to the polarization of the exciting laser beam. The fluorescence signal is further divided into two spectral ranges (630 DCXR, Chroma), equipped with bandpass filters (HC 520/35 (Semrock) and HQ 715/120 (Chroma)) for donor (Alexa488) and acceptor (Alexa647) fluorescence, respectively. The signal from single photon counting detectors (MicroPhotonDevices PD5CTC for the green channels and PerkinElmer AQR-14 for the red channels) was recorded photon-by-photon in time-tagged-time-resolved (tttr) mode with picosecond accuracy (HydraHarp400, Picoquant, Berlin, Germany).

For fullFCS measurements, prior to each sample measurement, a diluted nM concentration of Rh110 and Atto647N, respectively, was measured for 1 min in buffer containing no urea and in buffer containing the respective urea concentration, and the collar ring was adjusted such that the counts per molecule got maximal, i. e. the detection volume minimal. These reference measurements were necessary to correct for the influence of urea on the index of refraction,

changing the size and shape of the observation volume and on the viscosity. Samples were diluted in NUNC chambers (Lab-Tek, Thermo Scientific, Germany) in a total of 500 μL volume in the respective measurement buffer containing 50 mM Na-phosphate, 150 mM NaCl and between 0 M and 7.5 M urea. Additionally, 1 μM of unlabeled T4L was added to coat the chamber surface. The sample concentration was adjusted in the low nM range, yielding in average \sim three molecules in the focus at a time.

For single labeled samples (DOnly and AOnly), photons were collected for 2.5 hrs resulting in average about $3 \cdot 10^9$ collected photons, while for double labeled samples (DA) the measurement time was increased to 10 hrs resulting in average about $8 \cdot 10^9$ collected photons. The power at objective for green excitation was 20 μW (485 nm) and for red excitation 10 μW (635 nm).

Data analysis

Prior to any data analysis, the saved data files ranging from 10 GB for the measurements on single labeled variants up 50 GB for those on double labeled variants had to be split in smaller files for being able to perform the subsequent correlations on a standard personal computer. Files were split in fractions of three million photons, resulting into split file size of \sim 11 MB. Splitting and correlation was performed using a Labview-based program. Photons were correlated within a time window of 32 ps up to 2 s. The correlation function is calculated as follows:

$$G_{A,B}(t_c) = 1 + \frac{\langle \delta S_A(t) \cdot \delta S_B(t + t_c) \rangle}{\langle S_A(t) \rangle \cdot \langle S_B(t) \rangle} \quad 5.3.20$$

where t_c is the correlation time, $S_X(t)$ represents the detected intensity signal (number of detected photons per time interval) at channel X , and $\delta S_X(t)$ corresponds to the fluctuation from the time average of the signal in channel X denoted as $\langle S_X(t) \rangle$ ($X = A, B$). The auto-correlation function is defined when the correlated channels are the same $A = B$ and it is called cross-correlation if $A \neq B$.

Keeping in mind our setup with four channels (green parallel (Gp), green perpendicular (Gs), red parallel (Rp) and red perpendicular (Rs)), the correlation of the double labeled samples yielded in total twelve curves: four color- autocorrelation curves (GpGs, GsGp, RpRs, RsRp) and eight color-cross correlation curves (GpRp, RpGp, GsRs, RsGs, GpRs, RsGp, GsRp, RpGs). Single labeled samples, of course, yielded only the two respective color-autocorrelation curves.

The general correlation function $G(t_c)$ of a solution with N fluorescent molecules in the focus, which displays no further fluctuation contributions due to e.g. enzyme kinetics, follows the analytical form (69):

$$G(t_c) = 1 + \frac{1}{N} \cdot G_{diff}(t_c) = 1 + \frac{1}{N} \cdot \left(1 + \frac{t_c}{t_{diff}}\right)^{-1} \cdot \left(1 + \left(\frac{\omega_0}{z_0}\right)^2 \cdot \frac{t_c}{t_{diff}}\right)^{-\frac{1}{2}}. \quad 5.3.21$$

As detection volume a 3-dimensional Gaussian shaped element with spatial distribution of the detection probabilities was assumed: $w(x, y, z) = \exp(-2(x^2 + y^2)/\omega_0^2) \cdot \exp(-2z^2/z_0^2)$. The $1/e^2$ radii in x and y or in z direction are denoted by ω_0 and z_0 , respectively. The diffusion time of particles, t_{diff} is related to the diffusion constant D as $t_{diff} = \omega_0^2/4D$. The amplitude of the correlation function at $t \rightarrow 0$ is scaled with the reciprocal of the average number of fluorescent particles N in the confocal volume.

The presence of additional processes apart from diffusion which lead to fluctuations in the fluorescence signal e.g. quenching kinetics, triplet formation and photon antibunching are inducing additional characteristic decay or rise terms in the correlation curve (69).

Thus, with assumption that all characteristic correlation times are well separated, the shape of the color- autocorrelation curves of our single and double labeled samples can be described by the following analytical formula:

$$G(t_c) = b + \frac{1}{N} \cdot G_{diff}(t_c) \cdot G_{abun}(t_c) \cdot G_{rc}(t_c) \cdot G_{bun}(t_c) \quad 5.3.22$$

with b being a constant offset, $G_{abun}(t_c)$ describing the photon antibunching term, $G_{rc}(t_c)$ a term associated with the rotational correlation of the molecule and $G_{bun}(t_c)$ describes all further (protein-motion induced) processes which lead to additional decay terms:

$$\begin{aligned} G_{abun}(t_c) &= 1 - x_{ab} \cdot e^{-t_c/t_{ab}} \\ G_{rc}(t_c) &= 1 + x_{rc} \left(\frac{1}{1 + x_{rc2}} \cdot e^{-t_c/t_{rc1}} + \frac{x_{rc2}}{1 + x_{rc2}} \cdot e^{-t_c/t_{rc1} \cdot t_{rc}} \right) \\ G_{bun}(t_c) &= 1 + x_{b1} - x_{b1} \cdot e^{-t_c/t_{b1}} + x_{b2} - x_{b2} \cdot e^{-t_c/t_{b2}} + x_{b3} - x_{b3} \cdot e^{-t_c/t_{b3}} + x_{b4} - x_{b4} \cdot e^{-t_c/t_{b4}} \end{aligned} \quad 5.3.23$$

a-c

The photon antibunching term is described by its amplitude x_{ab} (usually ~ 1) and time constant t_{ab} which is the reciprocal sum of the rate constants for excitation and fluorescence decay (~ 4 ns for the donor, ~ 2 ns for the acceptor). The rotational correlation term, whose amplitude x_{rc} is polarization dependent, contains up to two time constants t_{rc1} and $t_{rc2} = t_{rc} \cdot t_{rc1}$, their respective amplitudes described by the exponential prefactors. To be able to describe the

correlation function to a sufficient level (flat residuals), we needed additionally up to four bunching times t_{b1} to t_{b4} – depending on the urea concentration. The bunching amplitudes are defined by the respective prefactors (or fractions) x_{b1} to x_{b4} and normalized such that $x_{b1}+x_{b2}+x_{b3}+x_{b4} = 1$.

Corresponding color- autocorrelation curves of single and double labeled samples were fitted globally, i.e. the antibunching, the rotational correlation and all bunching times were identical for all curves while amplitudes of antibunching and bunching terms were kept identical for each pair of curves, e.g. for DOnly-GsGp and DOnly-GpGs as well as for DA-GsGp and DA-GpGs. In contrast, the amplitudes of the rotational correlation terms were left free for individual fitting due to their polarization dependency. In general, especially at low concentration where additional enzyme kinetic based fluorescence fluctuations were expected, for the DA samples more bunching times than for the single labeled ones were needed. Respective amplitudes were set to zero in the single labeled curves.

In the color- cross correlation curves, processes which are related to the FRET are detected as anticorrelation terms. Additionally, the photon antibunching term $G_{abun}(t_c)$ and a polarization-dependent bunching term $G_{rc,c}(t_c)$, related to the rotational correlation, are needed to fully describe the form of the color- cross correlation function:

$$G(t_c) = b + \frac{1}{N} \cdot G_{diff}(t_c) \cdot G_{abun}(t_c) \cdot G_{rc,c}(t_c) \cdot G_{cab}(t_c) \quad 5.3.24$$

where

$$\begin{aligned} G_{rc,c}(t_c) &= 1 - x_{rc} + x_{rc} \cdot e^{-t_c/t_{rc}} \\ G_{cab}(t_c) &= 1 - f \cdot (x_{cab1} \cdot e^{-t_c/t_{cab1}} + x_{cab2} \cdot e^{-t_c/t_{cab2}} + x_{cab3} \cdot e^{-t_c/t_{cab3}} + x_{cab4} \cdot e^{-t_c/t_{cab4}} + x_{cab5} \cdot e^{-t_c/t_{cab5}}) \end{aligned} \quad 5.3.25$$

$G_{diff}(t_c)$ and $G_{abun}(t_c)$ are identical to the respective terms of the color- autocorrelation function. The bunching term $G_{rc,c}(t_c)$ is described by its amplitude x_{rc} and time constant t_{rc} . The anticorrelation terms are described by a form factor f , the time constants t_{cab1} to t_{cab5} and their respective amplitudes (fractions) x_{cab1} to x_{cab5} which are normalized such that $x_{cab1}+x_{cab2}+x_{cab3}+x_{cab4} +x_{cab5}= 1$.

For fitting, all parameters except N , f and x_{rc} were set globally over all eight color- cross correlation curves. In low urea concentration, additionally another anticorrelation term could not be set globally as it also displayed a polarization dependent amplitude.

5.3.1.9. *Filtered Fluorescence Correlation Spectroscopy*

Data acquisition and analysis was performed as described previously (Chapter 3)(47).

Filtered FCS requires prior knowledge of the time-resolved fluorescence and polarization decays for each species (47, 70). In contrast to fullFCS here pulsed excitation is required. For a mixture of more than two species, we generated two “patterns” (filter) corresponding to two “pseudo-species” termed low FRET (LF) and high FRET (HF). For this, single-molecule measurements of the same molecules and conditions were used. The bursts were analyzed using 1 ms time- windows and plotted in a two- dimensional histogram of number of detected photons versus proximity ratio. Based on our experience, the low FRET species was defined as all TW’s displaying a proximity ratio between 0 and 0.2, while the filter for the high FRET species was generated from bursts showing a proximity ratio of ~ 0.6 up to 1. Here, the exact lower border was dependent upon variant and urea concentration.

The yielded correlated curves (two $sCCFs$ (LF-> HF, HF-> LF) and two $sACFs$ (LF->LF, HF-> HF) were fitted globally using similar equations as described in 5.3.1.8. Again, species cross-correlations display anticorrelation while species auto- correlations show bunching terms. Unlike fullFCS, for $sCCFs$ and $sACFs$ no photon antibunching or rotational correlation bunching terms are observed. Thus, the needed equations simplify to:

$$G_{sCCF}(t_c) = b + \frac{1}{N} \cdot G_{diff}(t_c) \cdot G_{ab}(t_c) \cdot G_{slow}(t_c) \quad 5.3.26$$

$$G_{sACF}(t_c) = b + \frac{1}{N} \cdot G_{diff}(t_c) \cdot G_b(t_c) \cdot G_{slow}(t_c) \quad 5.3.27$$

with $G_{ab}(t_c)$, $G_b(t_c)$ and $G_{slow}(t_c)$:

$$\begin{aligned} G_{ab}(t_c) &= 1 - f \cdot (x_{ab1} \cdot e^{-t_c/t_{ab1}} + x_{ab2} \cdot e^{-t_c/t_{ab2}} + x_{ab3} \cdot e^{-t_c/t_{ab3}} + x_{ab4} \cdot e^{-t_c/t_{ab4}}) \\ G_b(t_c) &= 1 - x_{ab1} + x_{ab1} \cdot e^{-t_c/t_{ab1}} - x_{ab2} + x_{ab2} \cdot e^{-t_c/t_{ab2}} - x_{ab3} + x_{ab3} \cdot e^{-t_c/t_{ab3}} - x_{ab4} + x_{ab4} \cdot e^{-t_c/t_{ab4}} \end{aligned} \quad \begin{matrix} 5.3.28 \\ \text{a,b} \end{matrix}$$

$G_{slow}(t_c)$ is modeled as an additional correlation ($sACF$) and anti-correlation ($sCCF$) term, respectively. All time constants are kept global, while amplitudes are only linked for both $sCCFs$. The sum of the exponential prefactors, i.e. amplitudes of time constants, is normalized to 1.

The hydrodynamic radius R_H , which gives an idea about the size of the molecule diffusing through the objective’s focus can be calculated from t_{diff} .

$$R_H = \frac{k \cdot T \cdot t_{diff}^{TAL}}{6 \cdot \pi \cdot \eta \cdot t_{diff}^{Rh110} \cdot D^{Rh110}} \quad 5.3.29$$

with k being Boltzmann's constant ($1.38 \text{ (kg}\cdot\text{m}^2/(\text{s}^2\cdot\text{K}))$), T the temperature (293 K) and η the viscosity of the surrounding medium ($1.002 \text{ mPa}\cdot\text{s}$ in the absence of urea and $1.592 \text{ mPa}\cdot\text{s}$ in the presence of 7.5 M urea). Additionally, the diffusion time from a reference measurement of Rhodamine110 (t_{diff}^{Rh110}) and its diffusion coefficient D^{Rh110} was needed.

5.3.1.10. FRET positioning and Screening

We position the AVs using the experimental constraints and the constraints imposed by the model used in the simulations (rigid body docking or flexible arrangement of fluorophores) and display the mean fluorophore positions. The accessible volume (AV) approach considers dyes as hard sphere models connected to the protein via flexible linkers (modeled as a flexible cylindrical pipe) (53). The overall dimension (width and length) of the linker is based on their chemical structures. For Alexa488 the five carbon linker length was set to 20 Å, the width of the linker is 4.5 Å and the dye radii $R_1 = 5 \text{ Å}$, $R_2 = 4.5 \text{ Å}$ and $R_3 = 1.5 \text{ Å}$. For Alexa647 the dimensions used were: length = 22 Å, width = 4.5 Å and three dye radii $R_1 = 11 \text{ Å}$, $R_2 = 3 \text{ Å}$ and $R_3 = 3.5 \text{ Å}$.

To account for dye linker mobility we generated a series of AV's for donor and acceptor dyes. For each pair of AV's, we calculated the distance between dye mean positions (R_{mp})

$$R_{mp} = \left| \left\langle \vec{R}_{D(i)} \right\rangle - \left\langle \vec{R}_{A(j)} \right\rangle \right| = \left| \frac{1}{n} \sum_{i=1}^n \vec{R}_{D(i)} - \frac{1}{m} \sum_{j=1}^m \vec{R}_{A(j)} \right| \quad 5.3.30$$

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. This results in a “cloud” for each dye with all possible spatial coordinates where the center of mass of the dye could be located. However, in ensemble TCSPC the mean donor-acceptor distance is observed:

$$\langle R_{DA} \rangle = \left| \left\langle \vec{R}_{D(i)} - \vec{R}_{A(j)} \right\rangle \right| = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \left| \vec{R}_{D(i)} - \vec{R}_{A(j)} \right| \quad 5.3.31$$

which can be modeled with the accessible volume calculation.

For the native state we firstly used solely the information about $\langle R_{DA} \rangle$ obtained from our FRET-measurements alone (Gaussian-distributed mean distance model, Section 5.3.1.3).

Next, we evaluated the effect of the structural information on our docking process and thus, determined the fluorophore probability distributions based these restraints alone and neglected the information we obtained from FRET. Finally, we included structural information into the modelling. For this, we dissected an exemplary X-ray structure of T4L

(PDBID: 172L) into the following separate structure elements: α -helix A (residues 1-13), N-terminal subdomain (NTsD, 14-60), α -helix C (61-81) and C-terminal subdomain (CTsD, 82-164) (Figure 5.3.2). The dyes were then attached to the $C\beta$ -atom of the respective labeling position and the structural elements were held together by flexible peptide bonds. The structural elements were subsequently docked such that they fulfill the FRET-restraints but avoid steric clashes among one another.

In the denatured state we proceeded similarly. First, we use solely the FRET restraints to determine possible mean fluorophore positions. Next, in the second approach, due to the lack of structural information, we assumed that neighboring labeling sites (residues 8-19, 19-44, 44-55, 55-60, 60-69, 69-70, 70-86, 86-119, 119-132 and 132-150) were held together by a WLC. This reduces the spatial uncertainty of the fluorophore with respect to another. To calculate the parameters of this WLC-link, we used for each amino acid between the two fluorophores n a size of 3.6 \AA (total chain length $L = n * 3.6$) and an average stiffness of $\kappa = \kappa_{avg}$ (0.86). From the simulated $p(R_{DA})$ using equation 5.3.13, we obtained the length of the WLC from the mean distance and the flexibility of the WLC-link was given by the width of the $p(R_{DA})$ distribution. In the subsequent modeling both FRET-restraints and the WLC-links had to be fulfilled. Finally, we again evaluated the effect of the WLC-links. Here, we simulated the possible fluorophore distributions, when using only the WLC-links but neglecting the experimental restraints given by FRET.

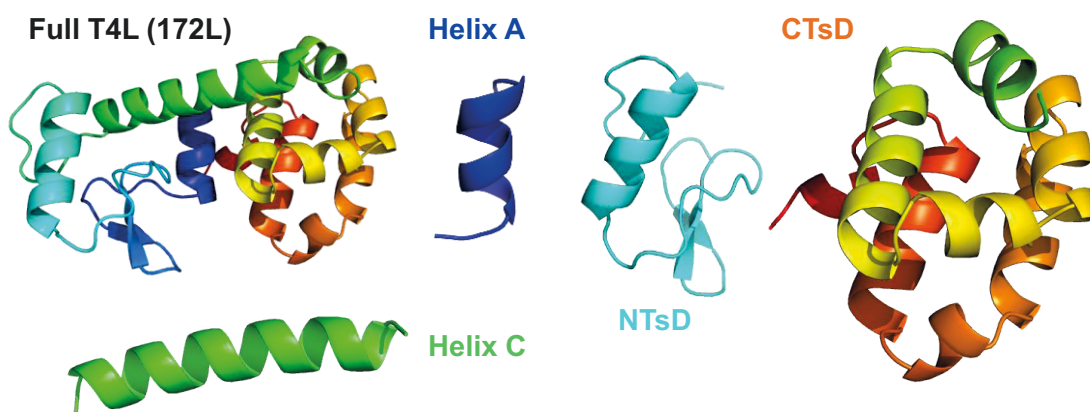


Figure 5.3.2 Exemplary X-ray structure of T4L (PDBID 172L) and the generated structural elements. Helix A encompasses residues 1-13, the NTsD 14-60, helix C 61-81 and the CTsD 82-164.

5.3.2. Supporting results

5.3.2.1. Fluorescence lifetime analysis of double labeled samples

All measured time-resolved fluorescence intensities were analyzed with various fit models (Table 5.3.1, Table 5.3.2). The measurements conducted in 0 M urea were taken from previously published data (Chapter 3) and re-analyzed using a single Gaussian distribution giving a single, average mean distance $\langle R_{DA} \rangle$, which is the contrary as was done in Chapter 3.

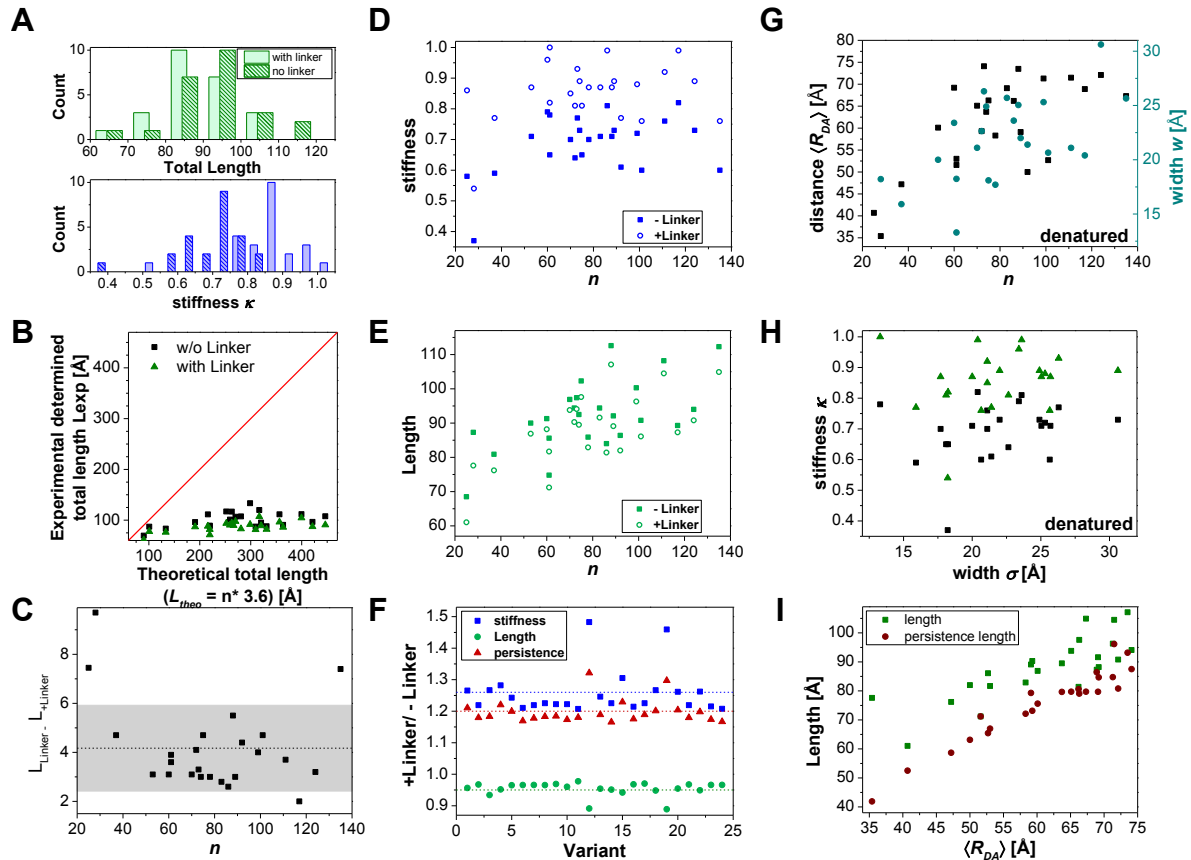


Figure 5.3.3 Additional eTCSPC fit results. (A) Distribution of WLC fit parameters stiffness and length from fit results as histograms. (B) The obtained total length L_{exp} is shorter than would be expected for an ideal chain with L_{theo} being equivalent to 3.6 Å per residue between the labeling sites. (C) The offset due to the linkers in L is 4.2 ± 1.8 Å. (D, E) For the Worm-like chain model, both fit parameters total length and stiffness κ increase with increasing interresidue separation. (F) Adding the linker-correction introduces a constant shift in the fit parameters of the WLC model. (G) Also, for the Gaussian distribution model, and increasing interresidue separation leads to an increased mean distance $\langle R_{DA} \rangle$ and width w_{DA} . (H) The obtained stiffness from the WLC model is not correlated with the width obtained from Gaussian distribution fit. (I) L and l_p increase with increasing $\langle R_{DA} \rangle$.

The distribution of the obtained fit parameter for the WLC, stiffness κ and total chain length L , is summarized in Figure 5.3.3A. Figure 5.3.3B compares the experimentally obtained total chain length L_{exp} with the theoretically expected one for an ideal chain, L_{theo} . L_{theo} assumes a length of 3.6 Å per amino acid residue between the two labeling sites. Figure 5.3.3D-E show that increasing number of residues between the labeling, n , leads to rising

values of both stiffness κ and L . The same is true for the fit parameters obtained in the Gaussian distribution fit, $\langle R_{DA} \rangle$ and width w_{DA} (Figure 5.3.3G). However, Figure 5.3.3H shows that the obtained stiffness from WLC and width of the distribution model are not correlated. Comparing the fit results for the WLC model shows a constant shift with the linker corrected model leads to a constant ratio and an offset in L by $4.2 \pm 1.8 \text{ \AA}$ (Figure 5.3.3C).

5.3.2.2. Tables with fit results of eTCSPC data

Table 5.3.1A Fit results of double labeled samples for the Gaussian distribution and Gaussian chain model. For 0 M urea, GC is not applicable and the Gaussian distribution model is only an approximation. Uncertainties given are standard errors of the mean.

Variant	Distribution fit, 0 M				Distribution fit, 7.5 M				Gaussian chain, 7.5 M		
	$\langle R_{DA} \rangle$ [Å]	w_{DA} [Å]	x_{D0}	χ^2_r	$\langle R_{DA} \rangle$ [Å]	w_{DA} [Å]	x_{D0}	χ^2_r	n	x_{D0}	χ^2_r
8-69	35.2 ± 0.5	8.4 ± 3.5	0.17 ± 0.01	1.16	59.3 ± 0.7	22.6 ± 1.8	0.03 ± 0.01	1.06	323.0	0.00	1.25
8-86	41.6 ± 0.2	9.3 ± 0.2	0.18 ± 0.01	1.11	59.1 ± 1.8	22.0 ± 2.1	0.20 ± 0.16	1.17	461.8	0.00	1.19
8-132	37.9 ± 0.3	10.2 ± 0.2	0.05 ± 0.01	1.15	67.3 ± 2.4	25.7 ± 3.4	0.20 ± 0.05	1.02	552.4	0.12	1.02
19-69 ¹					51.6 ± 0.7	13.3 ± 2.0	0.72 ± 0.02	1.02	276.8	0.84	1.04
19-86	49.5 ± 0.1	9.4 ± 0.2	0.05 ± 0.01	1.40	58.3 ± 0.3	17.7 ± 0.3	0.10 ± 0.01	1.20	312.1	0.00	1.83
19-119	49.4 ± 0.2	8.3 ± 0.4	0.08 ± 0.01	1.42	71.5 ± 0.1	21.1 ± 0.3	0 ± 0.0	1.11	478.4	0.00	2.20
19-132	42.6 ± 0.6	11.1 ± 0.7	0.06 ± 0.01	1.24	72.1 ± 8.7	30.6 ± 12.8	0.58 ± 0.12	1.06	521.8	0.82	1.04
44-86	49.0 ± 0.1	8.6 ± 0.3	0.23 ± 0.01	1.19	60.1 ± 0.6	20.0 ± 1.3	0.02 ± 0.01	1.05	313.1	0.00	1.97
44-119	52.4 ± 0.1	10.1 ± 0.3	0.15 ± 0.01	1.29	66.2 ± 2.1	23.6 ± 3.0	0.18 ± 0.06	1.11	524.3	0.00	1.16
44-132	51.3 ± 0.3	9.4 ± 0.6	0.26 ± 0.02	1.10	71.3 ± 0.2	25.3 ± 0.5	0 ± 0.0	1.00	487.6	0.00	1.35
44-150	51.6 ± 0.1	9.8 ± 0.2	0.03 ± 0.01	1.16	68.9 ± 0.1	20.4 ± 0.4	0.10 ± 0.01	1.11	484.0	0.00	1.80
55-69	37.1 ± 0.2	6.5 ± 0.2	0.25 ± 0.01	1.27	40.7 ± 0.4	11.2 ± 0.3	0.28 ± 0.01	1.04	115.3	0.43	1.42
55-119	60.4 ± 0.4	11.7 ± 0.5	0.14 ± 0.01	1.17	66.3 ± 3.8	18.1 ± 3.4	0.37 ± 0.18	1.35	656.8	0.00	1.41
55-132	49.6 ± 0.1	8.9 ± 0.2	0.04 ± 0.01	1.21	73.5 ± 0.1	25.0 ± 0.5	0 ± 0.0	1.01	521.6	0.00	1.41
60-86	47.2 ± 0.1	7.8 ± 0.5	0.39 ± 0.01	1.21	47.2 ± 0.5	15.9 ± 0.9	0.33 ± 0.01	1.04	206.9	0.53	1.10
60-119	49.9 ± 0.1	9.2 ± 0.2	0.05 ± 0.01	1.22	65.1 ± 0.4	21.1 ± 0.6	0.05 ± 0.01	1.23	393.2	0.00	2.01
60-132	43.7 ± 0.2	9.6 ± 0.2	0.05 ± 0.01	1.17	69.1 ± 3.2	25.7 ± 3.9	0.06 ± 0.01	1.10	459.8	0.03	1.21
60-150	41.6 ± 0.2	7.9 ± 0.2	0.23 ± 0.01	1.13	52.7 ± 1.2	20.7 ± 3.9	0.56 ± 0.05	1.01	299.0	0.77	1.02
69-86	39.6 ± 0.3	5.5 ± 0.3	0.52 ± 0.01	1.17	35.4 ± 1.7	18.2 ± 1.1	0.20 ± 0.01	1.09	133.9	0.44	1.11
69-119	41.2	7.3	0.33	1.15	53.0	18.2	0.37	1.03	311.5	0.52	1.05

THE DENATURED STATE OF T4L

	±0.2	±0.2	±0.01		±0.5	±1.6	±0.02				
69-132	40.4	9.7	0.21	1.21	63.7	24.9	0.20	1.13	533.8	0.00	1.13
	±0.3	±0.5	±0.01		±2.6	±3.6	±0.14				
69-150	45.5	8.6	0.44	1.15	50.0	21.4	0.61	1.09	216.5	0.84	1.12
	±0.2	±0.4	±0.01		±1.6	±3.2	±0.03				
70-119	35.8	9.5	0.10	1.12	69.2	23.4	0.47	1.03	713.0	0.46	1.03
	±0.8	±0.3	±0.01		±9.5	±7.1	±0.33				
70-132	33.7	11.2	0.11	1.10	74.1	26.3	0.25	1.03	760.5	0.00	1.04
	±0.6	±0.4	±0.01		±9.6	±6.4	±0.18				

¹Fitting 0 M data with the Gaussian distribution model was not successful.

Table 5.3.1B Fitresults of double labeled samples in 7.5 M urea for the WLC model.

Variant	WLC fit					WLC _{Linker} fit				
	κ	L [Å]	x_{D0}	χ^2_r	$\langle E_{DA} \rangle^I$	κ	L [Å]	x_{D0}	χ^2_r	$\langle E_{DA} \rangle^I$
8-69	0.64	94.4	0.13	1.15	0.49	0.81	90.3	0.13	1.15	0.47
	±0.02	±1.9	±0.02			±0.02	±1.8	±0.02		
8-86	0.73	92.1	0.37	1.21	0.41	0.89	89.1	0.36	1.21	0.41
	±0.04	±3.7	±0.03			±0.04	±3.6	±0.03		
8-132	0.60	112.3	0.23	1.09	0.42	0.76	104.9	0.24	1.09	0.44
	±0.02	±6.7	±0.06			±0.02	±5.2	±0.04		
19-69²	0.79	74.8	0.72	1.15	0.56	1.00	71.2	0.72	1.14	0.52
	±0.08	±3.7	±0.01			±0.12	±3.6	±0.01		
19-86	0.70	85.9	0.14	1.11	0.51	0.87	82.9	0.13	1.11	0.48
	±0.03	±2.6	±0.02			±0.03	±1.7	±0.02		
19-119	0.76	108.2	0	1.10	0.28	0.92	104.5	0	1.10	0.27
	±0.02	±4.3				±0.02	±3.1			
19-132	0.73	94.0	0.71	1.07	0.40	0.89	90.8	0.71	1.07	0.38
	±0.07	±9.4	±0.04			±0.08	±8.2	±0.04		
44-86	0.71	90.0	0.13	1.06	0.45	0.87	86.9	0.11	1.06	0.43
	±0.02	±1.8	±0.02			±0.02	±1.7	±0.02		
44-119	0.81	84.0	0.42	1.27	0.44	0.99	81.4	0.42	1.26	0.43
	±0.04	±2.5	±0.02			±0.05	±2.4	±0.02		
44-132	0.72	100.3	0.20	1.03	0.37	0.88	96.3	0.20	1.04	0.36
	±0.02	±4.0	±0.002			±0.03	±2.9	±0.03		
44-150	0.82	89.3	0.31	1.08	0.38	0.99	87.3	0.29	1.08	0.36
	±0.04	±3.6	±0.04			±0.04	±3.5	±0.04		
55-69	0.58	68.5	0.24	1.12	0.80	0.86	61.1	0.26	1.03	0.76
	±0.02	±0.7	±0.01			±0.03	±0.6	±0.01		
55-119²	0.65	102.3	0.36	1.62	0.42	0.81	97.6	0.36	1.62	0.52
	±0.03	±6.1	±0.05			±0.03	±4.9	±0.04		
55-132	0.71	112.6	0.06	1.06	0.30	0.87	107.1	0.08	1.07	0.3
	±0.02	±5.6	±0.08			±0.02	±4.3	±0.07		
60-86	0.59	80.9	0.32	1.08	0.67	0.77	76.2	0.32	1.07	0.64
	±0.02	±1.6	±0.01			±0.02	±1.5	±0.01		
60-119	0.70	96.9	0.16	1.28	0.42	0.85	93.8	0.15	1.28	0.42
	±0.02	±2.9	±0.03			±0.03	±2.8	±0.03		
60-132	0.71	94.4	0.30	1.18	0.42	0.87	91.6	0.29	1.18	0.42
	±0.04	±3.8	±0.03			±0.03	±3.7	±0.03		
60-150	0.60	90.8	0.56	1.06	0.57	0.76	86.1	0.56	1.06	0.56
	±0.04	±5.4	±0.03			±0.05	±4.3	±0.03		
69-86	0.37	87.3	0.16	1.17	0.83	0.54	77.9	0.17	1.13	0.82
	±0.03	±2.6	±0.01			±0.02	±3.1	±0.01		
69-119	0.65	85.6	0.40	1.11	0.57	0.82	81.7	0.39	1.11	0.54
	±0.03	±2.6	±0.02			±0.03	±1.6	±0.02		
69-132	0.73	92.5	0.35	1.22	0.42	0.89	89.5	0.35	1.22	0.41
	±0.04	±4.6	±0.04			±0.04	±4.5	±0.04		
69-150	0.61	86.4	0.62	1.00	0.60	0.77	81.9	0.62	1.00	0.57
	±0.05	±4.3	±0.02			±0.05	±4.1	±0.02		

70-119	0.79 ±0.06	91.3 ±6.4	0.59 ±0.04	1.09	0.37	0.96 ±0.07	88.2 ±5.3	0.59 ±0.04	1.09	0.34
70-132	0.77 ±0.05	97.4 ±6.8	0.45 ±0.05	1.17	0.34	0.93 ±0.05	94.1 ±5.6	0.44 ±0.05	1.17	0.32

¹ $\langle E_{DA} \rangle$ is the energy transfer of the double labeled fraction only. For 0 M data, the more accurate fit results from Chapter 3 were considered for calculating $\langle E_{DA} \rangle$.

Table 5.3.2A Fitting results of fluorescence lifetime of donor only labeled variants. $\langle \tau \rangle_x$ is the species weighted fluorescence lifetime, the fluorescence quantum yield Φ_x is calculated using a reference with known fluorescence quantum yield (Donor: $\langle \tau \rangle_{A488} = 4.0$ ns, $\Phi_{A488} = 0.8$; Acceptor: $\langle \tau \rangle_{Cy5} = 1.17$ ns, $\Phi_{Cy5} = 0.32$)

Variant	τ_1	x_1	τ_2	x_2	τ_3	x_3	$\langle \tau \rangle_x$	$\Phi_{D(0)}$	χ^2_r
8-69	0 M	4.09	0.9	1.52	0.1		3.83	0.77	1.27
	7.5 M	3.75	0.82	1.65	0.09	0.21	3.25	0.63	1.05
8-86	0 M	4.09	0.9	1.52	0.1		3.97	0.80	1.13
	7.5 M	3.82	0.87	2.39	0.08	0.66	3.56	0.71	1.08
8-132	0 M	4.09	0.9	1.52	0.1		3.97	0.80	1.13
	7.5 M	3.81	0.83	1.82	0.09	0.26	3.35	0.65	1.00
19-69	0 M	4.18	0.85	1.26	0.15		3.74	0.75	1.33
	7.5 M	3.72	0.77	1.68	0.12	0.26	3.09	0.6	1.04
19-86	0 M	3.81	0.92	1.35	0.08		3.61	0.72	1.29
	7.5 M	3.75	0.81	2.14	0.10	0.50	3.29	0.66	1.11
19-119	0 M	3.77	0.91	0.99	0.09		3.52	0.70	1.18
	7.5 M	3.86	0.83	1.87	0.10	0.35	3.42	0.67	1.03
19-132	0 M	3.83	0.93	1.71	0.07		3.69	0.74	1.21
	7.5 M	3.64	0.77	1.80	0.13	0.32	3.07	0.61	1.07
44-86	0 M	4.31	0.95	1.63	0.05		4.17	0.83	1.16
	7.5 M	3.67	0.79	1.71	0.12	0.30	3.13	0.61	1.01
44-119	0 M	4.29	0.96	1.23	0.04		4.16	0.83	1.20
	7.5 M	3.90	0.96	0.56	0.04		3.78	0.76	1.10
44-132	0 M	4.29	0.96	1.23	0.04		4.16	0.83	1.20
	7.5 M	3.91	0.88	1.87	0.06	0.23	3.58	0.70	1.05
44-150	0 M	4.29	0.96	1.23	0.04		4.16	0.83	1.20
	7.5 M	3.90	0.96	0.56	0.04		3.78	0.76	1.10
55-69	0 M	4.13	0.93	1.38	0.07		3.93	0.79	1.05
	7.5 M	3.83	0.88	2.08	0.06	0.29	3.54	0.69	1.02
55-119	0 M	4.38	0.91	1.59	0.09		3.92	0.83	1.28
	7.5 M	3.87	0.88	2.19	0.09	0.38	3.60	0.72	1.25
55-132	0 M	4.29	0.94	1.46	0.06		4.09	0.82	1.09
	7.5 M	3.88	0.87	1.88	0.07	0.25	3.53	0.69	1.06
60-86	0 M	4.12	0.95	1.95	0.05		4.00	0.80	1.31
	7.5 M	3.82	0.86	2.02	0.06	0.16	3.41	0.68	1.04
60-119	0 M						4.03		
	7.5 M	3.89	0.95	1.06	0.05		3.76	0.75	1.12
60-132	0 M						4.03		
	7.5 M	3.89	0.95	1.06	0.05		3.76	0.75	1.12
60-150	0 M	4.08	0.89	1.69	0.11		3.82	0.76	1.17
	7.5 M	3.77	0.85	2.45	0.09	0.21	3.42	0.68	1.01
69-86	0 M	4.07	0.90	1.62	0.10		3.76	0.77	1.10
	7.5 M	3.76	0.84	1.67	0.08	0.17	3.29	0.66	1.05
69-119	0 M	4.05	0.89	1.56	0.11		3.81	0.76	1.15
	7.5 M	3.78	0.85	2.27	0.08	0.23	3.43	0.69	1.07
69-132	0 M	4.04	0.89	1.41	0.11		4.05	0.75	1.10
	7.5 M	3.92	0.89	2.14	0.07	0.43	3.64	0.73	1.09
69-150	0 M	4.13	0.88	1.36	0.12		3.81	0.76	1.15
	7.5 M	3.93	0.61	2.92	0.24	0.10	3.11	0.62	1.18
70-119	0 M	3.99	0.82	1.47	0.18		3.53	0.71	1.07
	7.5 M	3.74	0.72	1.85	0.16	0.27	3.03	0.61	1.05

70-132	0 M	4.09	0.88	1.17	0.12			3.74	0.75	1.06
	7.5 M	3.74	0.72	1.85	0.16	0.27	0.12	3.03	0.61	1.05

Table 5.3.2B Fitting results of fluorescence lifetime of direct acceptor excitation.

Variant		τ_1	x_1	τ_2	x_2	τ_3	x_3	$\langle\tau\rangle_x$	$\Phi_{A(0)}$	χ^2_r
8-69	0 M	2.13	0.09	1.37	0.62	0.79	0.30	1.26	0.35	1.35
	7.5 M	1.69	0.61	1.00	0.20	0.09	0.19	1.25	0.34	1.13
8-86	0 M	1.75	0.28	1.05	0.72			1.25	0.34	1.32
	7.5 M	2.08	0.13	1.51	0.78	0.59	0.09	1.50	0.41	1.01
8-132	0 M	1.79	0.53	1.06	0.37	0.17	0.10	1.36	0.37	1.08
	7.5 M	1.67	0.58	0.98	0.26	0.11	0.17	1.23	0.34	1.08
19-69	0 M	2.00	0.14	1.29	0.58	0.71	0.28	1.23	0.34	1.07
	7.5 M	1.54	0.43	1.06	0.43	0.18	0.14	1.14	0.31	1.06
19-86	0 M	2.37	0.02	1.59	0.35	1.00	0.64	1.22	0.33	1.48
	7.5 M	1.89	0.28	1.42	0.64	0.41	0.08	1.47	0.40	1.15
19-119	0 M	1.96	0.13	1.30	0.63	0.73	0.24	1.25	0.35	1.12
	7.5 M	1.68	0.52	1.14	0.36	0.22	0.12	1.30	0.36	1.11
19-132	0 M	2.18	0.15	1.54	0.66	0.73	0.20	1.48	0.37	1.16
	7.5 M	1.60	0.54	0.97	0.30	0.13	0.16	1.17	0.32	1.05
44-86	0 M	2.39	0.03	1.40	0.58	0.81	0.39	1.20	0.34	1.13
	7.5 M	1.78	0.50	1.24	0.40	0.17	0.10	1.41	0.38	1.11
44-119	0 M	2.54	0.03	1.45	0.64	0.77	0.33	1.25	0.33	1.27
	7.5 M	2.12	0.07	1.55	0.82	0.77	0.11	1.50	0.41	1.11
44-132	0 M	2.68	0.02	1.56	0.64	0.84	0.34	1.34	0.35	1.11
	7.5 M	1.74	0.56	1.17	0.34	0.21	0.10	1.39	0.38	1.06
44-150	0 M	2.05	0.19	1.39	0.65	0.71	0.16	1.40	0.37	1.13
	7.5 M	1.68	0.57	1.14	0.36	0.27	0.07	1.38	0.38	1.15
55-69	0 M	2.12	0.09	1.37	0.60	0.76	0.30	1.26	0.34	1.27
	7.5 M	1.70	0.54	1.16	0.36	0.16	0.10	1.35	0.37	1.11
55-119	0 M	1.85	0.22	1.26	0.62	0.47	0.16	1.26	0.34	1.14
	7.5 M	1.67	0.69	1.06	0.24	0.18	0.07	1.42	0.39	1.05
55-132	0 M	1.93	0.20	1.27	0.65	0.60	0.15	1.30	0.34	1.19
	7.5 M	1.63	0.57	1.01	0.27	0.13	0.16	1.23	0.34	1.11
60-86	0 M	2.16	0.06	1.37	0.61	0.86	0.34	1.25	0.34	1.28
	7.5 M	1.75	0.53	1.20	0.37	0.17	0.10	1.38	0.38	1.05
60-119	0 M	1.99	0.15	1.26	0.71	0.60	0.14	1.27	0.34	1.20
	7.5 M	1.65	0.68	1.04	0.26	0.22	0.07	1.40	0.38	0.95
60-132	0 M	1.87	0.28	1.26	0.57	0.70	0.15	1.35	0.37	1.22
	7.5 M	1.64	0.68	0.95	0.25	0.21	0.08	1.36	0.37	0.97
60-150	0 M	1.94	0.14	1.28	0.59	0.60	0.27	1.18	0.35	1.29
	7.5 M	1.60	0.58	1.03	0.26	0.11	0.16	1.21	0.33	1.03
69-86	0 M	1.96	0.10	1.36	0.47	0.93	0.43	1.24	0.33	1.32
	7.5 M	1.66	0.55	1.08	0.32	0.12	0.14	1.28	0.35	0.99
69-119	0 M	1.77	0.31	1.05	0.69			1.28	0.34	1.44
	7.5 M	1.70	0.47	1.16	0.40	0.17	0.12	1.29	0.35	1.07
69-132	0 M	1.74	0.50	0.99	0.50			1.36	0.36	1.15
	7.5 M	1.95	0.12	1.52	0.74	0.69	0.14	1.45	0.40	1.13
69-150	0 M	2.50	0.03	1.61	0.56	0.89	0.41	1.33	0.35	1.25
	7.5 M	1.64	0.49	1.16	0.39	0.17	0.12	1.28	0.35	1.17
70-119	0 M	2.57	0.02	1.65	0.55	0.97	0.43	1.37	0.34	1.42
	7.5 M	1.61	0.48	1.07	0.37	0.20	0.15	1.20	0.33	1.11
70-132	0 M	2.98	0.01	1.69	0.54	0.99	0.46	1.37	0.34	1.31
	7.5 M	1.64	0.48	1.14	0.41	0.18	0.11	1.27	0.35	1.10

Table 5.3.2C Fitting results of fluorescence anisotropy of donor only labeled variants.

Variant		ρ_1	b_1	ρ_2	b_2	ρ_3	b_3	$\chi^2_{r,sum}$	$\chi^2_{r,dif}$	r_{ss}
8-69	0 M	0.18	0.08	1.47	0.06	9.76	0.23	1.05	1.04	0.18

	7.5 M	0.25	0.11	0.83	0.16	3.71	0.11	1.08	1.00	0.09
8-86	0 M	0.19	0.08	1.24	0.07	9.09	0.23	1.10	1.08	0.18
	7.5 M	0.04	0.04	0.61	0.20	3.42	0.14	1.10	1.08	0.10
8-132	0 M	0.18	0.08	1.47	0.06	9.76	0.23	1.05	1.04	0.19
	7.5 M	0.16	0.07	0.72	0.18	3.54	0.12	1.17	1.00	0.10
19-69	0 M	0.26	0.12	1.84	0.12	8.47	0.14	1.08	1.02	0.14
	7.5 M	0.38	0.19	1.64	0.13	10.79	0.06	0.98	0.98	0.03
19-86	0 M	0.26	0.12	1.84	0.12	8.47	0.14	1.08	1.02	0.14
	7.5 M	0.18	0.07	1.01	0.21	7.70	0.10	1.21	1.19	0.12
19-119	0 M	0.26	0.12	1.84	0.12	8.47	0.14	1.08	1.02	0.14
	7.5 M	0.21	0.10	0.84	0.17	4.02	0.11	1.07	1.19	0.01
19-132	0 M	0.16	0.08	1.15	0.06	10.39	0.24	1.07	0.89	0.19
	7.5 M	0.34	0.16	1.23	0.14	5.61	0.07	1.17	0.97	0.10
44-86	0 M	0.19	0.11	1.34	0.11	8.03	0.15	1.05	1.09	0.13
	7.5 M	0.42	0.17	1.81	0.14	11.53	0.07	1.19	1.06	0.08
44-119	0 M	0.19	0.11	1.34	0.11	8.03	0.15	1.05	1.09	0.13
	7.5 M	0.07	0.05	0.58	0.18	3.12	0.15	1.02	1.19	0.09
44-132	0 M	0.19	0.11	1.34	0.11	8.03	0.15	1.05	1.09	0.13
	7.5 M	0.18	0.07	0.64	0.17	2.98	0.14	1.05	1.03	0.09
44-150	0 M	0.19	0.11	1.27	0.07	8.60	0.20	1.16	1.05	0.15
	7.5 M	0.07	0.05	0.58	0.18	3.12	0.15	1.02	1.08	0.09
55-69	0 M	0.20	0.14	1.22	0.08	9.52	0.15	1.10	1.02	0.13
	7.5 M	0.35	0.24	1.28	0.10	12.13	0.03	1.10	1.11	0.01
55-119	0 M	0.17	0.13	1.27	0.08	10.89	0.16	1.02	1.00	0.14
	7.5 M	0.05	0.07	0.59	0.19	3.50	0.11	1.13	1.08	0.08
55-132	0 M	0.20	0.14	1.22	0.08	9.52	0.15	1.10	1.02	0.13
	7.5 M	0.46	0.19	2.19	0.15	16.36	0.04	1.06	1.14	0.04
60-86	0 M	0.18	0.10	1.24	0.08	8.24	0.20	1.17	0.97	0.16
	7.5 M	0.10	0.05	0.40	0.27	1.70	0.06	1.07	1.00	0.05
60-119	0 M	0.16	0.08	1.17	0.07	9.00	0.23	1.09	1.02	0.18
	7.5 M	0.06	0.05	0.68	0.21	4.32	0.12	1.04	1.04	0.10
60-132	0 M	0.18	0.10	1.24	0.08	8.24	0.20	1.17	0.97	0.16
	7.5 M	0.06	0.05	0.68	0.21	4.32	0.12	1.04	1.04	0.10
60-150	0 M	0.18	0.10	1.24	0.08	8.24	0.20	1.17	0.97	0.16
	7.5 M	0.05	0.08	0.41	0.27	2.71	0.02	1.01	1.08	0.05
69-86	0 M	0.18	0.12	1.07	0.09	8.26	0.17	1.08	1.03	0.14
	7.5 M	0.37	0.24	1.51	0.11	14.00	0.03	1.08	1.09	0.06
69-119	0 M	0.18	0.12	1.07	0.09	8.26	0.17	1.08	1.03	0.14
	7.5 M	0.07	0.08	0.44	0.27	3.48	0.03	1.04	1.06	0.04
69-132	0 M	0.18	0.12	1.07	0.09	8.26	0.17	1.08	1.03	0.14
	7.5 M	0.29	0.12	1.31	0.17	7.42	0.09	1.17	1.07	0.11
69-150	0 M	0.18	0.12	1.07	0.09	8.26	0.17	1.08	1.03	0.14
	7.5 M	0.17	0.14	0.79	0.13	3.12	0.10	1.11	1.05	0.07
70-119	0 M	0.23	0.11	1.38	0.09	8.42	0.17	1.08	1.03	0.15
	7.5 M	0.36	0.18	1.71	0.14	15.36	0.05	1.12	1.04	0.09
70-132	0 M	0.23	0.11	1.38	0.09	8.42	0.17	1.08	1.03	0.15
	7.5 M	0.36	0.18	1.71	0.14	15.36	0.05	1.12	1.04	0.09

Table 5.3.2D Fitting results of fluorescence anisotropy of direct acceptor excitation.

Variant	ρ_1	b_1	ρ_2	b_2	ρ_3	b_3	$\chi^2_{r,sum}$	$\chi^2_{r,dif}$	r_{ss}	
8-69	0 M	0.17	0.08	0.89	0.12	11.84	0.20	1.12	1.03	0.25
	7.5 M	0.02	0.06	0.68	0.09	2.90	0.25	1.17	1.00	0.18
8-86	0 M	0.10	0.06	0.79	0.15	11.26	0.18	1.15	1.04	0.24
	7.5 M	0.12	0.04	1.46	0.20	12.61	0.15	1.11	1.01	0.24
8-132	0 M	0.15	0.05	1.68	0.08	13.83	0.26	1.13	1.06	0.29
	7.5 M	0.04	0.08	0.34	0.04	1.92	0.27	1.09	0.172	0.18
19-69	0 M	0.08	0.06	0.57	0.12	13.96	0.21	1.21	1.00	0.26

	7.5 M	0.01	0.09	0.96	0.30			1.04	1.39	0.18
19-86	0 M	0.04	0.07	0.71	0.16	12.58	0.17	1.16	0.98	0.24
	7.5 M	0.10	0.03	1.30	0.19	9.46	0.17	1.21	1.05	0.24
19-119	0 M	0.03	0.03	0.80	0.15	17-19	0.21	1.09	0.93	0.24
	7.5 M	0.01	0.07	0.56	0.07	2.60	0.25	1.23	1.03	0.18
19-132	0 M	0.16	0.04	0.96	0.08	10.37	0.28	1.18	1.05	0.27
	7.5 M	0.04	0.09	0.72	0.06	2.29	0.24	1.18	2.05	0.17
44-86	0 M	0.09	0.06	0.78	0.17	19.93	0.16	1.18	1.14	0.23
	7.5 M	0.01	0.07	0.67	0.08	3.05	0.23	1.14	1.07	0.18
44-119	0 M	0.02	0.09	0.99	0.15	14.80	0.15	1.16	1.12	0.26
	7.5 M	0.14	0.04	1.47	0.21	10.50	0.14	1.10	1.01	0.23
44-132	0 M	0.12	0.07	0.67	0.06	10.45	0.27	1.06	1.16	0.28
	7.5 M	0.02	0.06	0.69	0.09	2.92	0.24	2.24	1.03	0.18
44-150	0 M	0.21	0.06	1.05	0.15	13.87	0.18	1.17	0.95	0.23
	7.5 M	0.21	0.01	1.02	0.24	13.42	0.15	1.12	1.01	0.23
55-69	0 M	0.08	0.04	0.66	0.13	13.12	0.22	1.40	1.00	0.26
	7.5 M	0.04	0.07	0.65	0.10	2.73	0.23	0.97	0.97	0.18
55-119	0 M	0.16	0.04	0.90	0.14	14.11	0.21	1.14	0.95	0.20
	7.5 M	0.14	0.02	1.33	0.19	8.06	0.18	1.06	1.01	0.24
55-132	0 M	0.12	0.10	0.78	0.08	11.38	0.22	1.25	1.07	0.24
	7.5 M	0.04	0.07	0.42	0.04	2.29	0.27	1.20	1.13	0.19
60-86	0 M	0.06	0.03	0.89	0.17	17.79	0.19	1.18	1.01	0.26
	7.5 M	0.01	0.08	0.55	0.07	2.68	0.24	1.28	1.11	0.17
60-119	0 M	0.17	0.03	0.82	0.13	11.47	0.23	1.76	1.29	0.25
	7.5 M	0.10	0.02	1.13	0.18	6.48	0.18	0.98	1.08	0.23
60-132	0 M	0.02	0.05	0.53	0.07	8.61	0.29	1.53	1.56	0.27
	7.5 M	0.04	0.03	0.99	0.16	4.89	0.21	1.10	1.14	0.22
60-150	0 M	0.10	0.04	0.70	0.09	11.08	0.26	1.13	1.07	0.27
	7.5 M	0.09	0.03	1.21	0.21	7.09	0.15	1.13	1.01	0.23
69-86	0 M	0.05	0.06	0.91	0.15	14.09	0.18	1.10	1.03	0.26
	7.5 M	0.03	0.10	0.39	0.06	1.95	0.23	1.23	1.49	0.15
69-119	0 M	0.35	0.11	1.72	0.10	19.97	0.19	1.10	1.01	0.26
	7.5 M	0.06	0.02	1.01	0.17	6.13	0.20	1.11	0.90	0.23
69-132	0 M	0.23	0.11	2.86	0.06	11.06	0.22	1.46	1.11	0.29
	7.5 M	0.09	0.03	1.38	0.19	8.28	0.19	1.17	0.97	0.25
69-150	0 M	0.14	0.05	0.69	0.08	8.66	0.26	1.17	1.02	0.27
	7.5 M	0.03	0.04	0.87	0.16	4.04	0.19	1.22	1.11	0.23
70-119	0 M	0.06	0.05	0.73	0.07	10.13	0.27	1.06	0.97	0.29
	7.5 M	0.06	0.03	1.03	0.18	5.96	0.18	1.18	1.05	0.23
70-132	0 M	0.07	0.05	0.75	0.07	9.80	0.27	1.06	0.98	0.28
	7.5 M	0.18	0.03	1.12	0.18	6.52	0.18	1.08	1.07	0.23

5.3.2.3. Time-resolved anisotropy

The time-resolved anisotropy data was fitted as described in the methods section. The global fits of the time-resolved “VV” and “VH” fluorescence intensity decay curves (Figure 5.3.4A) result in the parameter summarized in Table 5.3.2C-D. Figure 5.3.4B shows the fluorescence anisotropy decay of single donor labeled K60pAcF (green) and single acceptor labeled R119C (red) in the absence (dashed lines) and presence (solid) of urea. Figure 5.3.4D-G displays the (average) fraction and rotational correlation times of the time-resolved anisotropy measurements grouped dye-position wise in the absence and presence of urea. In concord with a decreased steady-anisotropy (Figure 5.3.4C), r_{ss} , the amplitude b_3 and/ or the

longest rotational correlation time ρ_3 are decreasing in 7.5 M urea (Figure 5.3.4F,I). However, a decrease in the overall protein rotation seems at first pass contradictory as it is generally known that unfolded protein have larger radius of gyration than their folded counterparts and should thus rotate slower (leading to an increased rotational correlation time).

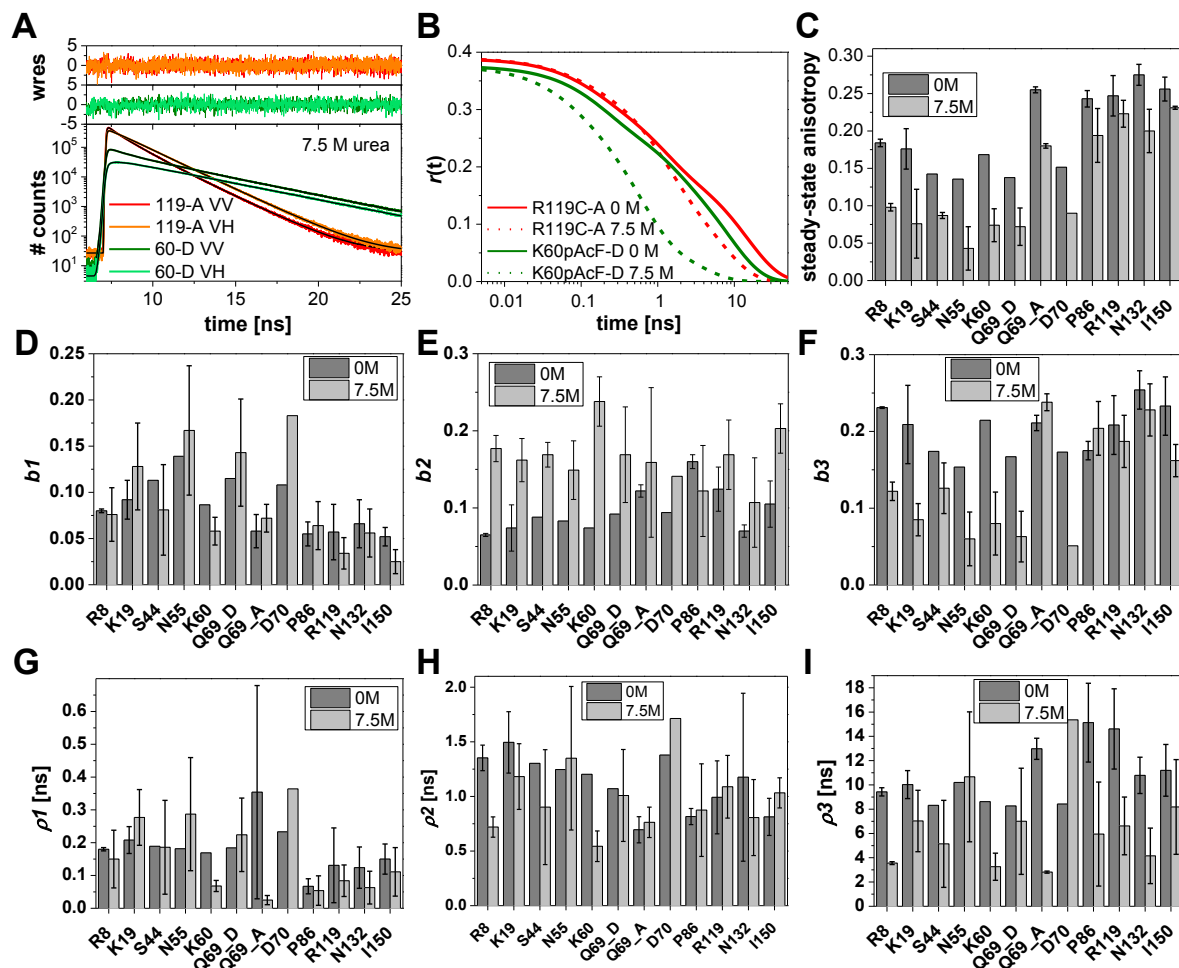


Figure 5.3.4 Time-resolved anisotropy of single labeled variants. (A) Example of time-resolved fluorescence decay measured under vertical excitation and vertical (VV) and horizontal (VH) emission detection, respectively. Weighted residuals are shown on top. (B) Time-resolved anisotropy decay calculated from a global three exponential fit of eTCSPC data measured under VV and VH conditions. One can clearly distinguish the three different time regimes: dye rotation (~ 0.2 ns), linker motion (~ 1 ns) and global protein motion (> 5 ns). (C) Steady-state anisotropy of single labeled variants in the absence or presence of urea. (D-I) The single labeled donor or acceptor labeled variants were fit as described in the methods section and if more than three variants sharing the same position labeled with the same dye were measured, the mean and standard deviation of all fits are shown.

This effect might be explained due to the increased mobility of the dyes; they are less coupled to the protein and also the unfolded protein itself gets much more flexible due to the lack of stable tertiary structure elements and thus the dyes might be able to only probe the motion of larger protein segments but to a much lesser extent the overall tumbling of the protein. The linker movement (Figure 5.3.4E) and its amplitude (Figure 5.3.4B) decrease and increase, respectively, for the positions 8 to 19 whereas for the other positions the changes are only

marginal. For the fastest movement, the dye rotation (Figure 5.3.4A, D), the changes are highly position depending and rather inconclusive.

5.3.2.4. *Stiffness matrix*

To resolve the stiffness of each protein segment probed, we modeled the stiffness values obtained from fitting our eTCSPC data with the linker-corrected WLC model as “spring constants” and made use of the relationship known from physics that for a series of individual springs the resulting spring constant k equals the sum of the reciprocal values of the individual spring constants k_i (152):

$$\frac{1}{K} = \sum_i \frac{1}{k_i} \quad 5.3.32$$

Thus, to gain the stiffness of each of our protein segments we set up the following matrix and solved with Python:

$$\begin{bmatrix} 1.23 \\ 1.12 \\ 1.32 \\ 1.00 \\ 1.15 \\ 1.09 \\ 1.12 \\ 1.15 \\ 1.01 \\ 1.14 \\ 1.01 \\ 1.16 \\ 1.23 \\ 1.15 \\ 1.30 \\ 1.18 \\ 1.15 \\ 1.32 \\ 1.85 \\ 1.22 \\ 1.12 \\ 1.30 \\ 1.04 \\ 1.08 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1/S_1 \\ 1/S_2 \\ 1/S_3 \\ 1/S_4 \\ 1/S_5 \\ 1/S_6 \\ 1/S_7 \\ 1/S_8 \\ 1/S_9 \\ 1/S_{10} \end{bmatrix}$$

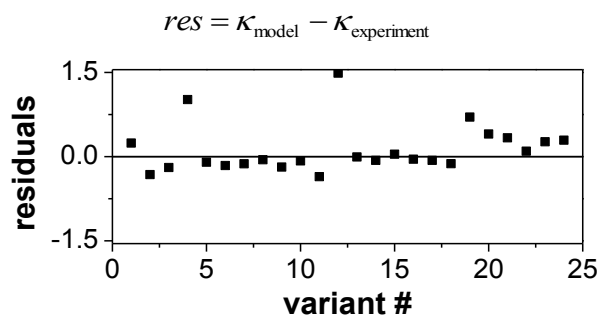
The lhs of the equation are the reciprocals of the κ values summarized in Table 5.3.1B; the column vector on the rhs is stiffness of our ten protein segments and the linkers, respectively.

The matrix in the middle contains a “1”, if the respective segment is enclosed between the two dyes, and a “0”, if not. The definition of the protein segments and the fit result is summarized in Table 5.3.3.

Table 5.3.3 Stiffness of protein segments

Segment	Residues	Stiffness k_i
----	1-8	----
S_1	8-19	0.57
S_2	19-44	0.09
S_3	44-55	0
S_4	55-60	0
S_5	60-69	0.53
S_6	69-70	0
S_7	70-86	1.00
S_8	86-119	0.02
S_9	119-132	0
S_{10}	132-150	0.42
----	150-164	----

To evaluate the fit result, the residuals were calculated (Figure 5.3.5):

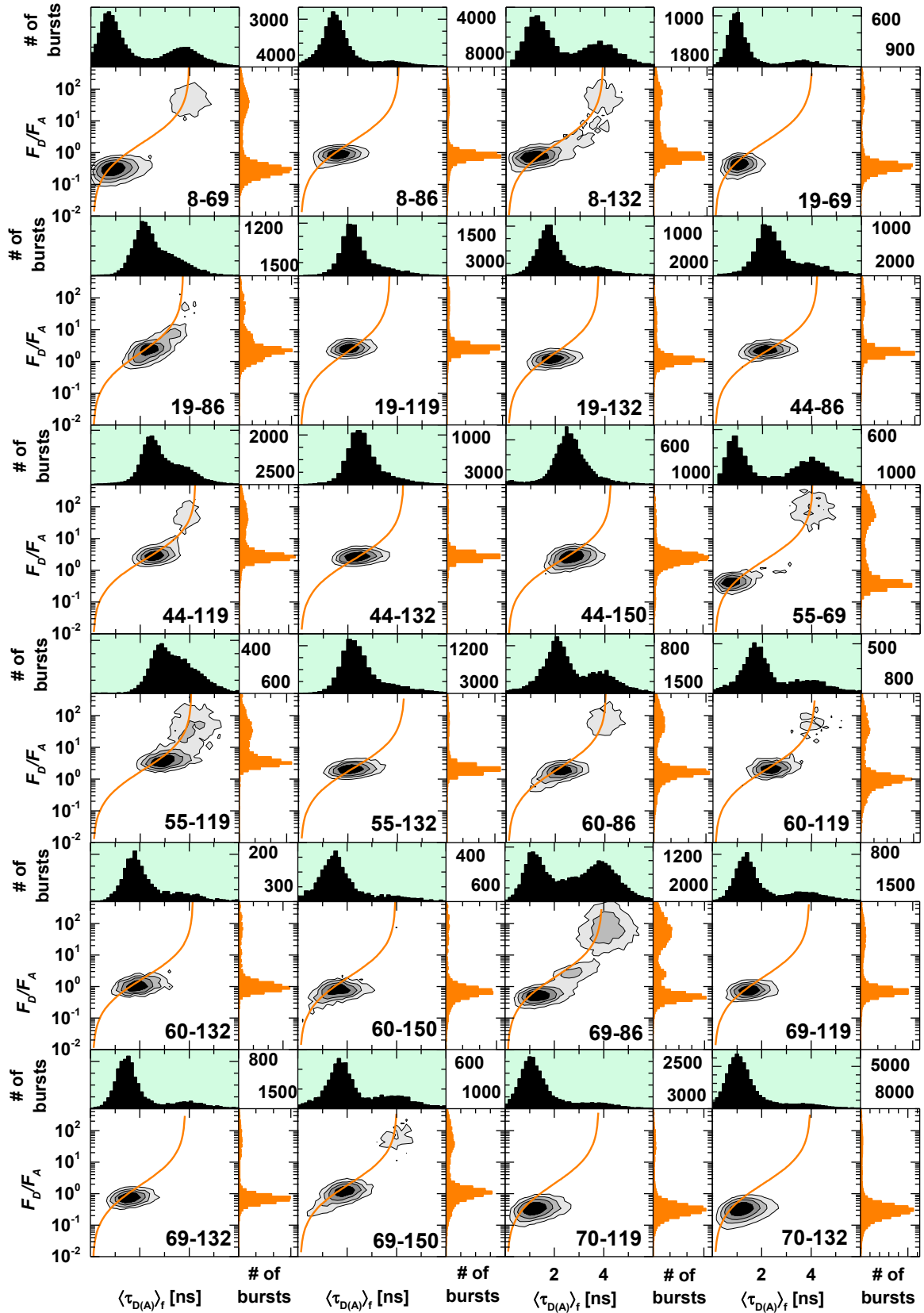


5.3.33

Figure 5.3.5 Residuals of the solved matrix equation.

5.3.2.5. Additional single-molecule MFD histograms

A



B

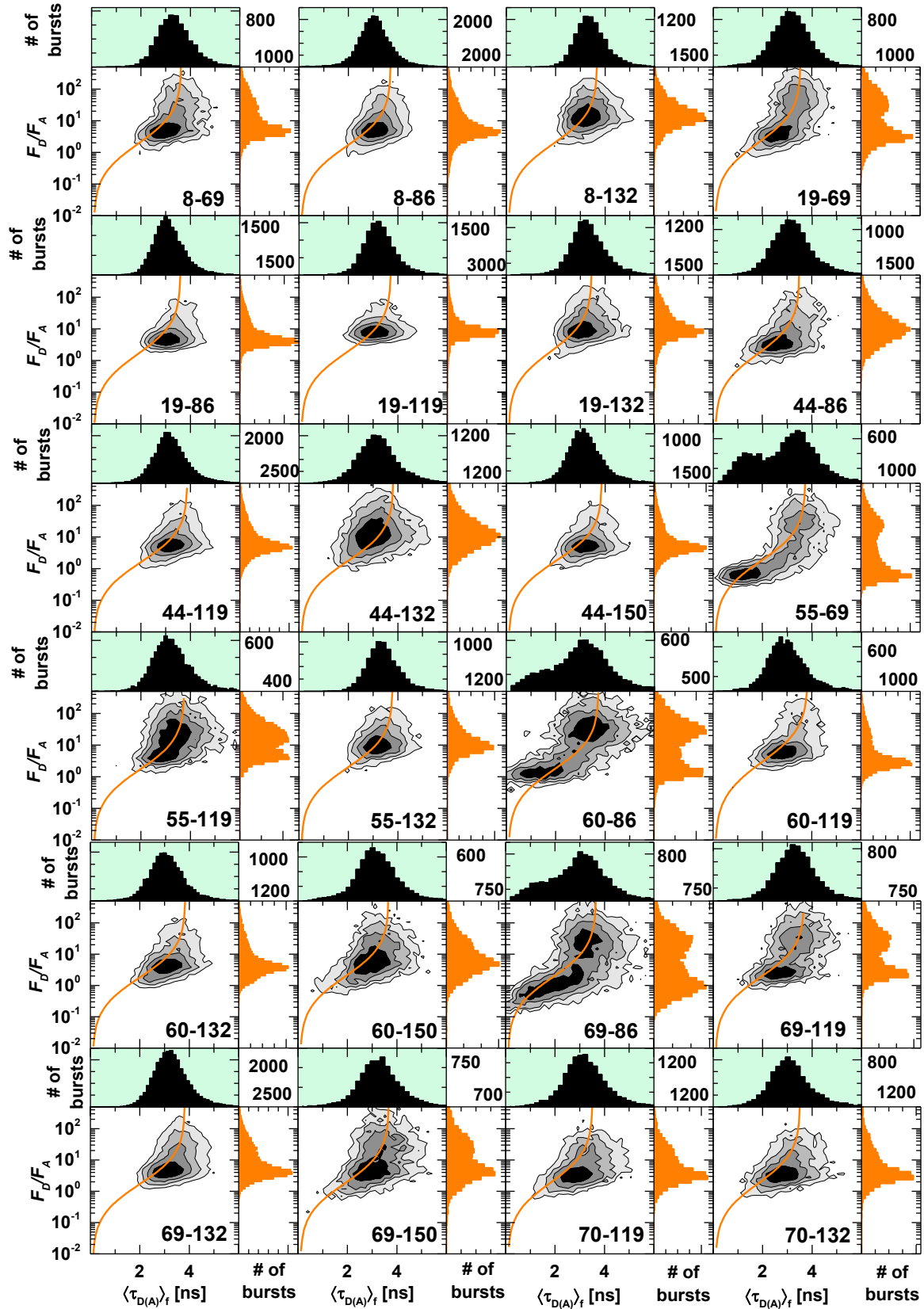


Figure 5.3.6 2D MFD Histograms for all 24 variants in the absence (A) and presence (B) of urea. Correction factors and static FRET lines are summarized in Table 5.3.4.

Table 5.3.4 Correction factors and static FRET lines for MFD experiments

#	1			R8pAcPhe Q69C			
0 M	$(0.7668/0.3500)/((3.9891/((-0.0401*x^3)+(0.2769*x^2)+0.5443*x+-0.0501))-1)$						
	det eff	0.75	crosstalk	1.7 %	bggreen	0.939 kHz	bgred 0.808 kHz
7.5 M	$(0.6504/0.3400)/((3.6360/((-0.0500*x^3)+(0.2927*x^2)+0.6106*x+-0.0545))-1)$						
	det eff	0.65	crosstalk	1.7 %	bggreen	3.086 kHz	bgred 1.282 kHz
#	2			R8pAcPhe P86C			
0 M	$(0.7958/0.3400)/((4.0756/((-0.0377*x^3)+(0.2728*x^2)+0.5253*x+-0.0479))-1)$						
	det eff	0.75	crosstalk	1.7 %	bggreen	1.211 kHz	bgred 0.954 kHz
7.5 M	$(0.7118/0.4100)/((3.7200/((-0.0464*x^3)+(0.2999*x^2)+0.5385*x+-0.0476))-1)$						
	det eff	0.77	crosstalk	2.6 %	bggreen	2.628 kHz	bgred 1.071 kHz
#	3			R8pAcPhe N132C			
0 M	$(0.7431/0.3700)/((3.9334/((-0.0418*x^3)+(0.2793*x^2)+0.5599*x+-0.0511))-1)$						
	det eff	0.75	crosstalk	2.4 %	bggreen	1.257 kHz	bgred 0.776 kHz
7.5 M	$(0.6701/0.3400)/((3.6838/((-0.0485*x^3)+(0.2924*x^2)+0.5940*x+-0.0531))-1)$						
	det eff	0.65	crosstalk	1.7 %	bggreen	3.086 kHz	bgred 1.282 kHz
#	4			K19pAcPhe Q69C			
0 M	$(0.7472/0.3400)/((4.0291/((-0.0400*x^3)+(0.2644*x^2)+0.5970*x+-0.0571))-1)$						
	det eff	0.75	crosstalk	1.7 %	bggreen	1.768 kHz	bgred 1.227kHz
7.5 M	$(0.6184/0.3100)/((3.5574/((-0.0529*x^3)+(0.2907*x^2)+0.6506*x+-0.0585))-1)$						
	det eff	0.65	crosstalk	1.7 %	bggreen	3.086 kHz	bgred 1.282kHz
#	5			K19pAcPhe P86C			
0 M	$(0.7216/0.3800)/((3.7346/((-0.0459*x^3)+(0.3006*x^2)+0.5291*x+-0.0470))-1)$						
	det eff	0.70	crosstalk	2.2 %	bggreen	0.789 kHz	bgred 0.653 kHz
7.5 M	$(0.6584/0.4000)/((3.6012/((-0.0510*x^3)+(0.3002*x^2)+0.5936*x+-0.0532))-1)$						
	det eff	0.70	crosstalk	2.2 %	bggreen	2.065 kHz	bgred 0.946 kHz
#	6			K19pAcPhe R119C			
0 M	$(0.7036/0.3500)/((3.6976/((-0.0469*x^3)+(0.2988*x^2)+0.5488*x+-0.0490))-1)$						
	det eff	0.75	crosstalk	1.7 %	bggreen	1.022 kHz	bgred 0.678 kHz
7.5 M	$(0.6833/0.3600)/((3.7234/((-0.0474*x^3)+(0.2905*x^2)+0.5887*x+-0.0532))-1)$						
	det eff	0.65	crosstalk	1.7 %	bggreen	3.504 kHz	bgred 1.593 kHz
#	7			K19pAcPhe N132C			
0 M	$(0.7375/0.3700)/((3.7660/((-0.0446*x^3)+(0.3006*x^2)+0.5118*x+-0.0449))-1)$						
	det eff	0.75	crosstalk	1.7 %	bggreen	1.346 kHz	bgred 0.980 kHz
7.5 M	$(0.6138/0.3200)/((3.4651/((-0.0562*x^3)+(0.3067*x^2)+0.6270*x+-0.0558))-1)$						
	det eff	0.70	crosstalk	1.7 %	bggreen	3.880 kHz	bgred 1.814 kHz
#	8			S44pAcPhe P86C			
0 M	$(0.8345/0.3400)/((4.2537/((-0.0338*x^3)+(0.2579*x^2)+0.5251*x+-0.0485))-1)$						
	det eff	0.75	crosstalk	1.7 %	bggreen	1.215 kHz	bgred 0.862kHz
7.5 M	$(0.6263/0.3800)/((3.5139/((-0.0543*x^3)+(0.3033*x^2)+0.6197*x+-0.0553))-1)$						
	det eff	0.65	crosstalk	1.7 %	bggreen	3.504 kHz	bgred 1.593kHz
#	9			S44pAcPhe R119C			
0 M	$(0.8302/0.3300)/((4.2502/((-0.0337*x^3)+(0.2562*x^2)+0.5307*x+-0.0490))-1)$						
	det eff	0.90	crosstalk	2.0 %	bggreen	1.674 kHz	bgred 0.769 kHz
7.5 M	$(0.7570/0.4100)/((3.8846/((-0.0408*x^3)+(0.2848*x^2)+0.5204*x+-0.0459))-1)$						
	det eff	0.75	crosstalk	2.6 %	bggreen	3.457 kHz	bgred 1.349 kHz
#	10			S44pAcPhe N132C			
0 M	$(0.8292/0.3500)/((4.2547/((-0.0335*x^3)+(0.2548*x^2)+0.5330*x+-0.0492))-1)$						
	det eff	0.385	crosstalk	1.8 %	bggreen	0.405 kHz	bgred 0.403 kHz
7.5 M	$(0.7151/0.3800)/((3.8285/((-0.0438*x^3)+(0.2849*x^2)+0.5639*x+-0.0504))-1)$						
	det eff	0.65	crosstalk	1.7 %	bggreen	3.504 kHz	bgred 1.593 kHz
#	11			S44pAcPhe I150C			
0 M	$(0.8286/0.3700)/((4.2516/((-0.0336*x^3)+(0.2550*x^2)+0.5329*x+-0.0491))-1)$						
	det eff	0.77	crosstalk	1.8 %	bggreen	0.757 kHz	bgred 0.574 kHz
7.5 M	$(0.7570/0.3800)/((3.8846/((-0.0408*x^3)+(0.2848*x^2)+0.5204*x+-0.0459))-1)$						
	det eff	0.75	crosstalk	1.7 %	bggreen	3.337 kHz	bgred 1.272 kHz

THE DENATURED STATE OF T4L

#	12			N55pAcPhe Q69C			
0 M	(0.7852/0.3400)/((4.0605/((-0.0380*x^3)+(0.2707*x^2)+0.5385*x+-0.0495))-1)						
	det eff	0.75	crosstalk	1.7 %	bggreen	1.196 kHz	bgred 0.839 kHz
7.5 M	(0.7076/0.3700)/((3.7529/((-0.0455*x^3)+(0.2936*x^2)+0.5516*x+-0.0487))-1)						
	det eff	0.70	crosstalk	1.7 %	bggreen	3.880 kHz	bgred 1.814 kHz
#	13			N55pAcPhe R119C			
0 M	(0.7836/0.3700)/((4.0828/((-0.0386*x^3)+(0.2710*x^2)+0.5482*x+-0.0511))-1)						
	det eff	0.80	crosstalk	2.4 %	bggreen	0.888 kHz	bgred 0.919 kHz
7.5 M	(0.7211/0.3900)/((3.7602/((-0.0454*x^3)+(0.2978*x^2)+0.5343*x+-0.0472))-1)						
	det eff	0.80	crosstalk	2.1 %	bggreen	1.876 kHz	bgred 0.945 kHz
#	14			N55pAcPhe N132C			
0 M	(0.8268/0.3400)/((4.2857/((-0.0340*x^3)+(0.2537*x^2)+0.5481*x+-0.0518))-1)						
	det eff	0.75	crosstalk	1.4 %	bggreen	1.046 kHz	bgred 0.453 kHz
7.5 M	(0.7049/0.3400)/((3.7937/((-0.0448*x^3)+(0.2867*x^2)+0.5697*x+-0.0509))-1)						
	det eff	0.65	crosstalk	1.7 %	bggreen	3.086 kHz	bgred 1.282 kHz
#	15			K60pAcPhe P86C			
0 M	(0.7957/0.3400)/((4.0611/((-0.0375*x^3)+(0.2727*x^2)+0.5211*x+-0.0472))-1)						
	det eff	0.75	crosstalk	1.7 %	bggreen	0.966 kHz	bgred 0.677 kHz
7.5 M	(0.6818/0.3800)/((3.7384/((-0.0465*x^3)+(0.2891*x^2)+0.5813*x+-0.0514))-1)						
	det eff	0.80	crosstalk	1.7 %	bggreen	2.458 kHz	bgred 1.210kHz
#	16			K60pAcPhe R119C			
0 M	(0.8035/0.3700)/((4.1544/((-0.0365*x^3)+(0.2651*x^2)+0.5404*x+-0.0503))-1)						
	det eff	0.75	crosstalk	1.6 %	bggreen	1.242 kHz	bgred 0.783 kHz
7.5 M	(0.7512/0.3800)/((3.8493/((-0.0420*x^3)+(0.2894*x^2)+0.5197*x+-0.0460))-1)						
	det eff	0.90	crosstalk	1.7 %	bggreen	3.544 kHz	bgred 1.390 kHz
#	17			K60pAcPhe N132C			
0 M	(0.8067/0.3900)/((4.1617/((-0.0363*x^3)+(0.2647*x^2)+0.5377*x+-0.0499))-1)						
	det eff	0.74	crosstalk	1.6 %	bggreen	1.242 kHz	bgred 0.783 kHz
7.5 M	(0.7512/0.3700)/((3.8493/((-0.0420*x^3)+(0.2894*x^2)+0.5197*x+-0.0460))-1)						
	det eff	1.0	crosstalk	2.4 %	bggreen	4.737 kHz	bgred 1.901 kHz
#	18			K60pAcPhe I150C			
0 M	(0.7633/0.3500)/((3.9644/((-0.0408*x^3)+(0.2805*x^2)+0.5407*x+-0.0496))-1)						
	det eff	0.75	crosstalk	1.7 %	bggreen	1.157 kHz	bgred 1.368 kHz
7.5 M	(0.6846/0.3300)/((3.6759/((-0.0479*x^3)+(0.2996*x^2)+0.5585*x+-0.0487))-1)						
	det eff	0.95	crosstalk	1.7 %	bggreen	4.332 kHz	bgred 1.501 kHz
#	19			Q69pAcPhe P86C			
0 M	(0.7878/0.3300)/((4.0383/((-0.0381*x^3)+(0.2741*x^2)+0.5262*x+-0.0478))-1)						
	det eff	0.75	crosstalk	1.7 %	bggreen	1.239 kHz	bgred 0.861 kHz
7.5 M	(0.6573/0.3500)/((3.6617/((-0.0491*x^3)+(0.2922*x^2)+0.6018*x+-0.0535))-1)						
	det eff	0.90	crosstalk	1.7 %	bggreen	4.508 kHz	bgred 1.501 kHz
#	20			Q69pAcPhe R119C			
0 M	(0.7564/0.3400)/((3.9385/((-0.0414*x^3)+(0.2818*x^2)+0.5436*x+-0.0498))-1)						
	det eff	0.75	crosstalk	1.7 %	bggreen	0.584 kHz	bgred 0.405 kHz
7.5 M	(0.6866/0.3500)/((3.6846/((-0.0478*x^3)+(0.2988*x^2)+0.5602*x+-0.0491))-1)						
	det eff	0.90	crosstalk	1.7 %	bggreen	4.508 kHz	bgred 1.501 kHz
#	21			Q69pAcPhe N132C			
0 M	(0.7486/0.3600)/((3.9279/((-0.0417*x^3)+(0.2803*x^2)+0.5543*x+-0.0511))-1)						
	det eff	0.75	crosstalk	1.7 %	bggreen	1.239 kHz	bgred 0.861 kHz
7.5 M	(0.7274/0.4000)/((3.8264/((-0.0436*x^3)+(0.2881*x^2)+0.5476*x+-0.0489))-1)						
	det eff	0.75	crosstalk	1.7 %	bggreen	3.337 kHz	bgred 1.272 kHz
#	22			Q69pAcPhe I150C			
0 M	(0.7612/0.3500)/((4.0172/((-0.0398*x^3)+(0.2709*x^2)+0.5653*x+-0.0529))-1)						
	det eff	0.75	crosstalk	1.7 %	bggreen	1.239 kHz	bgred 0.861 kHz
7.5 M	(0.6223/0.3500)/((3.6826/((-0.0495*x^3)+(0.2800*x^2)+0.6546*x+-0.0583))-1)						
	det eff	0.95	crosstalk	1.7 %	bggreen	3.000 kHz	bgred 1.229 kHz
#	23			D70nAcPhe R119C			

0 M	$(0.7065/0.3400)/((3.7983/((-0.0464*x^3)+(0.2888*x^2)+0.5853*x+-0.0548))-1)$							
	det eff	0.75	crosstalk	1.7 %	bggreen	1.768 kHz	bgred	1.227 kHz
7.5 M	$(0.6052/0.3300)/((3.5223/((-0.0547*x^3)+(0.2916*x^2)+0.6669*x+-0.0602))-1)$							
	det eff	0.95	crosstalk	1.7 %	bggreen	3.000 kHz	bgred	1.229 kHz
#	24			D70pAcPhe N132C				
0 M	$(0.7480/0.3400)/((3.9834/((-0.0404*x^3)+(0.2704*x^2)+0.5771*x+-0.0541))-1)$							
	det eff	0.75	crosstalk	1.7 %	bggreen	1.768 kHz	bgred	1.227 kHz
7.5 M	$(0.6050/0.3500)/((3.5231/((-0.0546*x^3)+(0.2912*x^2)+0.6680*x+-0.0603))-1)$							
	det eff	0.95	crosstalk	1.7 %	bggreen	3.000 kHz	bgred	1.229 kHz

5.3.2.6. Additional results PDA

The results of fitting the MFD-data with a static, 1 ms TW model (sPDA) or with a model assuming dynamic exchange between two distances (dPDA) by globally analyzing three TW's (2 ms, 3 ms and 4 ms) are summarized in Table 5.3.5. The half-widths hw_{DA} of the 1 ms TW sPDA fit of the distributions and the relaxation time constants t_{rel} of the dPDA fit are not correlated with the stiffness from the eTCSPC WLC fit results (Figure 5.3.7A-B).

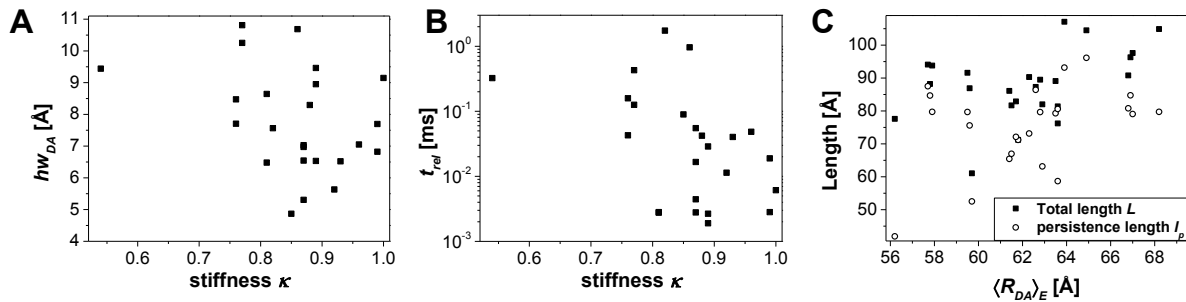


Figure 5.3.7 Results of sPDA and dPDA. (A, B) The stiffness of the variants obtained with the WLC model in eTCSPC is not correlated with the half-width hw_{DA} from sPDA or the relaxation time t_{rel} from dPDA. (C) An increasing mean distance $\langle R_{DA} \rangle_E$ slightly correlates with an increasing total chain length L and persistence length l_p from WLC model.

Increasing $\langle R_{DA} \rangle_E$ are mostly reflected in larger values for total chain length L and persistence length l_p (Figure 5.3.7C).

Table 5.3.5A Static PDA, 1 ms time-window. Uncertainties given are standard errors of the mean.

Variant	0 M			7.5 M		
	$\langle R_{DA} \rangle_E$ [Å]	hw_{DA} [Å]	x_{DOnly} [%]	$\langle R_{DA} \rangle_E$ [Å]	hw_{DA} [Å]	x_{DOnly} [%]
8-69	37.3±0.1	3.8±0.01	55.5±0.2	62.3±0.2	6.5±0.02	15.5±0.04
8-86	43.9±0.1	3.1±0.006	37.6±0.07	63.5±0.4	9.5±0.06	15.8±0.09
8-132	42.1±0.3	2.7±0.02	41.3±0.3	68.2±0.1	7.7±0.01	0.3±0.001
19-69	40.1±0.2	2.6±0.02	64.7±0.4	61.8±0.4	9.1±0.06	27.3±0.2
19-86	51.6±0.3	5.9±0.04	30.2±0.2	61.7±0.3	5.3±0.03	20.4±0.1
19-119	53.0±0.2	2.9±0.01	30.6±0.1	64.9±0.3	5.6±0.02	7.2±0.03
19-132	46.4±0.3	3.6±0.02	50.0±0.3	66.8±0.3	9.0±0.04	5.3±0.03
44-86	51.8±0.2	3.5±0.01	35.8±0.1	59.6±0.4	7.0±0.05	21.4±0.2
44-119	52.3±0.2	3.4±0.01	51.7±0.2	63.6±0.3	7.7±0.03	17.1±0.07
44-132	53.4±0.2	2.7±0.008	17.2±0.05	66.9±0.3	8.3±0.03	8.3±0.03
44-150	56.3±0.4	6.7±0.05	28.6±0.2	62.6±0.4	6.8±0.05	20.9±0.1
55-69	39.4±0.2	2.5±0.01	58.6±0.2	59.7±0.3	10.7±0.05	14.2±0.06
55-119	57.9±0.7	3.9±0.05	52.5±0.7	67.0±1.2	8.6±0.16	56.2±1.0
55-132	51.0±0.1	2.7±0.005	14.0±0.02	63.9±0.2	7.0±0.02	5.0±0.01

60-86	49.5±0.3	6.2±0.04	64.5±0.4	63.6±0.5	10.8±0.09	29.9±0.2
60-119	50.5±0.5	3.2±0.03	52.2±0.6	57.9±0.4	4.9±0.04	40.5±0.3
60-132	46.3±0.4	2.8±0.03	44.3±0.4	59.5±0.3	6.5±0.03	14.0±0.07
60-150	42.6±0.3	4.3±0.03	49.4±0.4	61.4±0.3	8.5±0.04	13.2±0.06
69-86				56.2±0.2	9.4±0.03	13.9±0.04
69-119	42.9±0.2	2.7±0.01	55.4±0.2	61.5±0.4	7.6±0.05	31.2±0.2
69-132	43.4±0.2	2.6±0.1	42.7±0.2	62.8±0.2	6.5±0.02	23.6±0.08
69-150	45.6±0.4	5.7±0.05	61.2±0.5	62.9±0.6	10.3±0.09	27±0.2
70-119	38.4±0.1	3.4±0.01	29.5±0.09	57.8±0.3	7.1±0.04	15.6±0.09
70-132	38.2±0.1	3.2±0.006	32.4±0.06	57.7±0.2	6.5±0.03	18.2±0.07

Table 5.3.5B Dynamic PDA, global fitting of 2 ms, 3 ms and 4 ms time-windows with the model described in Section 5.3.1.7

Variant		$\langle R_{DA} \rangle_{E,1}$ [Å] (fraction [%])	$\langle R_{DA} \rangle_{E,2}$ [Å] (fraction [%])	$\langle R_{DA} \rangle_{E,3}$ [Å] (fraction [%])	hw_{DA} [%]	A_{dyn} [%]	x_{DOnly} [%]	k_{12} [ms ⁻¹]	k_{21} [ms ⁻¹]	t_{rel} [ms]
8-69	0 M	37.3 (26.2)	88.1 (20.4)		5.8	40.5	12.8	0.54	0.14	1.47
	7.5 M	46.2 (1.4)	77.9 (35.0)	61.0 (38.0)	8.6	25.6	0.0	223	132	2.8e-3
8-86	0 M	35.3 (8.7)	57.5 (4.0)	86.9 (26.8)	8.7	54.2	6.2	38.5	53.6	0.011
	7.5 M	45.4 (6.2)	83.3 (33.6)	61.1 (46.7)	8.9	13.5	0.0	251	125	2.7e-3
8-132	0 M	32.2 (2.7)	56.2 (4.5)	90 (18.8)	11.4	57.3	9.8	12.1	20.3	0.031
	7.5 M	56.5 (11.4)	73.3 (22.9)		9.2	65.8	0.0	4.95	1.36	0.16
19-69	0 M	39.9 (30.9)	75 (3.0)	101 (31.13)	4.2	10.1	24.8	0.64	0.48	0.89
	7.5 M	53.5 (19.4)	74.6 (55.1)		11.0	25.1	0.3	54.1	109	6.1e-3
19-86	0 M	46.1 (25.9)	61.4 (16.5)	90 (18.8)	6.7	31.9	6.8	16	19	0.029
	7.5 M	46.0 (2.5)	80.3 (33.2)		11.1	62.7	1.6	39.6	20.6	0.017
19-119	0 M	42.3 (1.8)	66.1 (6.2)	93.3 (18.9)	7.7	66.3	6.8	30.5	23.2	0.019
	7.5 M	53.6 (5.5)	80.2 (21.3)	42.8 (0.8)	7.1	72.4	0.0	53.5	34.4	0.011
19-132	0 M	39.2 (5.4)	59.7 (6.3)	89.4 (27.9)	9.1	43.4	16.9	10.8	15.9	0.037
	7.5 M	53.2 (8.5)	77.4 (39.3)	42.4 (1.9)	7.7	50.3	0.0	21.9	12.8	0.029
44-86	0 M	39.1 (4.2)	66.5 (8.1)	94.6 (28.0)	9.8	58.4	1.3	32.9	24.1	0.018
	7.5 M	49.9 (9.3)	80.4 (32.4)	62.7 (24.4)	7.1	33.9	0.0	84.4	141	4.4e-3
44-119	0 M	40.7 (2.6)	66.8 (5.2)	90.2 (14.1)	9.2	44.7	33.5	32.3	20.5	0.019
	7.5 M	55.5 (18.7)	81.3 (34.3)	43.5 (0.8)	8.5	45.2	0.9	24.7	28.2	0.019
44-132	0 M	44.3 (2.9)	65.8 (4.3)	90.1 (11.1)	6.2	79.1	2.7	39.2	32.9	0.014
	7.5 M	57.7 (21.9)	79.6 (35.3)	44.8 (0.7)	7.4	42.1	0.0	13.4	10.3	0.042
44-	0 M	46.9	79.4	33.2	9.3	49.0	13.4	10.4	9.88	0.049

THE DENATURED STATE OF T4L

150		(7.1)	(14.5)	(16.0)						
	7.5 M	54.7 (17.4)	83.6 (32.7)		8.2	15.4	1.3	163	192	2.8e-3
55-69	0 M	39.1 (15.6)	69.3 (1.7)	89.2 (36.9)	5.3	36.0	9.9	0.110	0.286	2.53
	7.5 M	44.7 (0.8)	62.9 (14.6)	74.8 (35.3)	6.5	49.3	0.0	0.300	0.738	0.96
55-119	0 M	43.9 (1.0)	67.8 (5.1)	92.8 (36.7)	8.1	48.1	9.1	61.9	27.1	0.011
	7.5 M	56.5 (9.1)	88.8 (63.8)		8.6	19.0	8.1	113	249	2.8e-3
55-132	0 M	41.9 (2.1)	62.7 (7.0)	85.4 (11.6)	7.1	79.3	0.0	31.0	28.7	0.017
	7.5 M	53.7 (13.0)	73.5 (30.4)		7.6	56.6	0.0	11.5	6.65	0.055
60-86	0 M	39.1 (6.5)	67.0 (7.2)	97.7 (58.1)	11.4	28.2	0.0	14.2	13.8	0.036
	7.5 M	50.2 (22.2)	82.0 (35.7)	40.6 (0.8)	5.5	41.0	0.3	1.83	0.503	0.43
60-119	0 M	42.2 (2.0)	61.1 (3.6)	89.9 (32.8)	6.7	50.4	11.3	22.5	21.0	0.023
	7.5 M	53.7 (16.9)	75.0 (12.5)		7.2	38.6	32.0	3.92	7.24	0.090
60-132	0 M	35.6 (0.9)	58.8 (3.4)	95.9 (29)	10.4	60.9	5.9	16.0	17.9	0.029
	7.5 M	54.1 (33.8)	72.4 (38.1)		7.3	28.2	0.0	148	211	2.8e-3
60-150	0 M	34.5 (7.8)	55.4 (7.0)	91.3 (36.1)	11.0	41.1	7.9	9.54	14.6	0.041
	7.5 M	53.5 (32.0)	73.7 (42.4)		10.0	25.6	0.0	13.4	9.99	0.043
69-86	0 M	25.8 (1.7)	52.4 (13.4)	91.5 (61.8)	13.0	18.7	4.4	30.1	42.4	0.014
	7.5 M	63.2 (25.5)	44.2 (25.7)	74.6 (21.9)	6.3	26.9	0.0	1.29	1.80	0.32
69-119	0 M	33.1 (2.0)	57.2 (5.6)	90.9 (40.6)	11.1	41.8	10.0	19.3	29.6	0.020
	7.5 M	52.9 (10.1)	76.1 (52.1)		7.8	37.2	0.5	0.231	0.344	1.74
69-132	0 M	33.4 (1.6)	58.1 (5.2)	90.5 (31.1)	11.0	55.8	6.0	16.4	25.5	0.024
	7.5 M	59.1 (36.5)	80.2 (40.3)		8.3	21.2	2.0	68.5	458	1.9e-3
69-150	0 M	38.6 (10.7)	62.7 (5.5)	94.5 (49.7)	11.4	27.9	6.1	9.46	13.8	0.043
	7.5 M	49.8 (13.3)	80.5 (55.3)		11.7	31.5	0.0	4.40	3.53	0.13
70-119	0 M	38.4 (61.5)	69.9 (4.3)	90.8 (18.9)	6.1	13.0	2.3	0.713	0.684	0.72
	7.5 M	51.4 (25.7)	71.1 (42.5)		9.3	32.0	0.0	7.43	13.2	0.048
70-132	0 M	38.2 (61.1)	65.5 (2.7)	88.2 (19.0)	5.8	9.3	7.7	0.792	0.850	0.61
	7.5 M	53.0 (29.1)	74.0 (34.5)		9.5	35.0	1.4	7.59	17.2	0.040

THE DENATURED STATE OF T4L

5.3.2.7. Additional fullFCS results

Table 5.3.6 Fitting results for fullFCS measurements. Gray shaded cells indicate where values have been fixed for fitting. Uncertainties given are standard errors of the mean.

R8pAcF P86C (7.5 M)																	
Variable	DOnly		DA		AOnly		DA		Variable	DA							
	g _s -g _p	g _p -g _s	g _s -g _p	g _p -g _s	r _s -r _p	r _p -r _s	r _s -r _p	r _p -r _s		g _p -r _p	g _p -r _s	g _s -r _p	g _s -r _s	r _p -g _p	r _p -g _s	r _s -g _p	r _s -g _s
<i>b</i>	1				1.07		1.0		<i>b</i>	1							
<i>N</i>	62.7		71.3		16.7		42.9		<i>N</i>	4.11							
<i>t_d</i> [ms]	57.4				7.80		1.54		<i>t_d</i> [ms]	0.24							
<i>z0/ω0</i>	1.10		0.14		7				<i>z0/ω0</i>	3.5							
<i>x_{ab}</i>	0.74		0.85		0.97	0.81	1.0	0.97	<i>x_{ab}</i>	1							
<i>t_{ab}</i> [ms]	(3.00±0.97)*10 ⁻⁶				(2.9±0.69)*10 ⁻⁶		(9.56±4.89)*10 ⁻⁶		<i>t_{ab}</i> [ms]	3.0*10 ⁻⁶							
<i>x_{rc1}</i>	0.38	0.39	0.43	0.42	1.05	0.1	8.62	5.76	<i>x_{rc}</i>	0.49	0.50	0.50	0.51	0.49	0.50	0.49	0.51
<i>t_{rc1}</i> [ms]	(1.67±0.56)*10 ⁻⁵				(6.82±0.18)*10 ⁻⁶				<i>t_{rc}</i> [ms]	1.70*10 ⁻⁵							
<i>x_{rc2}</i>					0.03	0.02	0.19	0.23	<i>f</i>	0.96							
<i>t_{rc2}</i> [ms]					(1.72±1.04)*10 ⁻⁴				<i>x_{cab1}</i>	0.02±0.005							
									<i>t_{cab1}</i> [ms]	(1.15±0.40)*10 ⁻⁴							
<i>x_{b1}</i>			0.05±0.002		0.24±0.01		0.16±0.100		<i>x_{cab2}</i>	0.10±0.001							
<i>t_{b1}</i> [ms]	(2.33±2.05)*10 ⁻⁴				(3.59±1.43)*10 ⁻³				<i>t_{cab2}</i> [ms]	1.00±0.01							
<i>x_{b2}</i>	0.26±0.004		0.22±0.003		0.13±0.01		0.45±0.01		<i>x_{cab3}</i>	0.88±0.11							
<i>t_{b2}</i> [ms]	0.113±0.005				0.016±0.011		0.087±4*10 ⁻⁴		<i>t_{cab3}</i> [ms]	20.1±3.62							
<i>x_{b3}</i>	0.50±0.004		0.30±0.002		0.28±0.01		0.40±0.02		<i>x_{cab4}</i>								
<i>t_{b3}</i> [ms]	0.826±0.012				0.388±0.002		0.887±0.030		<i>t_{cab4}</i> [ms]								
<i>x_{b4}</i>	0.23±0.004		0.42±0.005		0.35±0.004				<i>x_{cab5}</i>								
<i>t_{b4}</i> [ms]	3.73±0.004				2.06±0.07				<i>t_{cab5}</i> [ms]								
K19pAcF P86C (7.5 M)																	
<i>b</i>	1				1.02		1.0		<i>b</i>	1							
<i>N</i>	53.4		73.2		14.4		77.2		<i>N</i>	5.58	5.10	8.00	8.46	6.85	8.23	7.03	7.55
<i>t_d</i> [ms]	1.57				1.25		1.70		<i>t_d</i> [ms]	0.24							
<i>z0/ω0</i>	7		46.41		7				<i>z0/ω0</i>	3.5							
<i>x_{ab}</i>	0.89		0.62		0.94	0.76	1.38	0.91	<i>x_{ab}</i>	1							
<i>t_{ab}</i> [ms]	(3.39±1.81)*10 ⁻⁶		(3.53±1.68)*10 ⁻⁶		(2.01±0.19)*10 ⁻⁶		(2.00±0.40)*10 ⁻⁶		<i>t_{ab}</i> [ms]	3.0*10 ⁻⁶							
<i>x_{rc1}</i>	0.11	0.11	0.21	0.23	0.53	0.18	2.41	0.76	<i>x_{rc}</i>	0.45	0.51	0.32	0.29	0.32	0.29	0.31	0.37

THE DENATURED STATE OF T4L

t_{rc1} [ms]	(6.64±2.29)*10 ⁻⁵				(8.59±0.06)*10 ⁻⁶				t_{rc} [ms]	6.60*10 ⁻⁵									
x_{rc2}					0.02	0.03	0.37	0.43	f	0.95									
t_{rc2} [ms]					(1.48±0.18)*10 ⁻⁴				x_{cab1}	0.03±0.004									
									t_{cab1} [ms]	(2.81±0.73)*10 ⁻⁴									
x_{b1}	0.34±0.003				0.44±0.005		0.08±0.02		x_{cab2}	0.11±0.007									
t_{b1} [ms]	0.095±6.7*10 ⁻⁴				(3.18±0.1)*10 ⁻³				t_{cab2} [ms]	1.00±0.05									
x_{b2}			0.34±0.002		0.24±0.01		0.52±0.01		x_{cab3}	0.87±0.09									
t_{b2} [ms]	0.107±7.5*10 ⁻⁴				0.013±5*10 ⁻⁴		0.090±0.003		t_{cab3} [ms]	18.7±2.95									
x_{b3}	0.66±0.009		0.34±0.003		0.32±0.005		0.40±0.006		x_{cab4}										
t_{b3} [ms]	0.492±0.007				0.428±4*10 ⁻⁴		0.905±9*10 ⁻⁴		t_{cab4} [ms]										
x_{b4}			0.33±0.002						x_{cab5}										
t_{b4} [ms]	1.70±0.03								t_{cab5} [ms]										
S44pAcF R119C (7.5 M)																			
b	1				1.01		1.0		b	1									
N	51.6		51.2		29.0		39.8		N	4.65	3.97	5.85	5.62	4.74	5.00	5.07	6.34		
t_d [ms]	0.53		0.80		2.56		1.52		t_d [ms]	0.21									
$z0/\omega0$	40.1				7				$z0/\omega0$	3.5									
x_{ab}	0.84		0.78		1.0	0.79	1.20	1.00	x_{ab}	1									
t_{ab} [ms]	(2.03±0.82)*10 ⁻⁶		(2.37±0.73)*10 ⁻⁶		(2.13±0.18)*10 ⁻⁶		(3.00±0.27)*10 ⁻⁶		t_{ab} [ms]	3.0*10 ⁻⁶									
x_{rc1}	0.17	0.27	0.01	0.08	0.73	0.35	1.96	0.97	x_{rc}	0.39	0.49	0.34	0.37	0.38	0.43	0.34	0.29		
t_{rc1} [ms]	(3.81±1.42)*10 ⁻⁵				(9.04±0.26)*10 ⁻⁶				t_{rc} [ms]	3.8*10 ⁻⁵									
x_{rc2}					0.18	0.18	0.04	0.05	f	0.95									
t_{rc2} [ms]					(3.12±0.13)*10 ⁻³				x_{cab1}	0.03±0.01									
									t_{cab1} [ms]	(1.15±0.66)*10 ⁻⁴									
x_{b1}			0.31±0.03		0.27±0.004		0.45±0.003		x_{cab2}	0.11±5*10 ⁻⁴									
t_{b1} [ms]	(3.54±1.44)*10 ⁻⁴				2.20±0.004		(1.06±0.34)*10 ⁻⁴		t_{cab2} [ms]	0.86±9*10 ⁻⁴									
x_{b2}	0.91±0.008		0.69±0.008		0.30±0.004		0.04±0.004		x_{cab3}	0.87±0.17									
t_{b2} [ms]	(6.15±1.94)*10 ⁻³				(8.68±0.35)*10 ⁻³				t_{cab3} [ms]	14.3±4.07									
x_{b3}	0.09±0.007		0.01±0.005		0.08±0.01		0.29±0.005		x_{cab4}										
t_{b3} [ms]	0.129±0.002				0.091±0.003				t_{cab4} [ms]										
x_{b4}					0.35±0.02		0.22±0.003		x_{cab5}										
t_{b4} [ms]					0.507±0.002				0.790±0.002									t_{cab5} [ms]	
S44pAcF I150C (7.5 M)																			
b	1				1				b	1									
N	30.6		51.5		15.9		46.5		N	3.96	3.13	4.33	5.54	5.73	5.64	3.65	4.27		

THE DENATURED STATE OF T4L

t_d [ms]	1.32				1.48		1.75		t_d [ms]	0.26							
$z0/\omega0$	7				7				$z0/\omega0$	3.5							
x_{ab}	1.14		0.90		1.08	0.82	1.39	1.10	x_{ab}	1							
t_{ab} [ms]	(1.32±0.78)*10 ⁻⁶		(2.00±0.76)*10 ⁻⁶		(1.45±0.18)*10 ⁻⁶		(2.00±0.29)*10 ⁻⁶		t_{ab} [ms]	3.0*10 ⁻⁶							
x_{rc1}	0.06	0.05	0.01	0.02	0.51	0.14	2.74	1.85	x_{rc}	0.49	0.60	0.51	0.38	0.25	0.35	0.53	0.52
t_{rc1} [ms]	(9.75±3.56)*10 ⁻⁴				(1.16±0.70)*10 ⁻⁵				t_{rc} [ms]	1.0*10 ⁻⁵							
x_{rc2}					0.01	0.01	0.21	0.22	f	0.93							
t_{rc2} [ms]					(2.43±0.25)*10 ⁻⁴				x_{cab1}	0.02±0.005							
									t_{cab1} [ms]	(1.18±0.48)*10 ⁻⁴							
x_{b1}			0.53±0.006		0.33±0.002		0.12±0.010		x_{cab2}	0.10±0.001							
t_{b1} [ms]	(6.99±4.33)*10 ⁻⁵				(6.63±0.24)*10 ⁻³					0.88±0.01							
x_{b2}	0.64±0.01		0.35±0.004		0.25±0.002		0.36±0.008		x_{cab3}	0.88±0.12							
t_{b2} [ms]	0.018±0.002				0.022±0.001		0.086±0.004			17.8±3.74							
x_{b3}	0.36±0.004		0.13±0.001		0.19±0.004		0.14±0.01		x_{cab4}								
t_{b3} [ms]	0.729±0.004				0.275±0.009												
x_{b4}					0.23±0.002		0.38±0.007		x_{cab5}								
t_{b4} [ms]					1.09±0.004				t_{cab5} [ms]								
K60pAcF N132C (0 M)																	
b	1				1				b	1							
N	4.42		1.75		1.96		2.60		N	0.58	0.56	0.88	0.86	0.43	0.60	0.45	0.60
t_d [ms]	0.20		0.26		0.29		0.24		t_d [ms]	0.13							
$z0/\omega0$	5.63		23.36		6.53		13.84		$z0/\omega0$	3.50							
x_{ab}	0.93		0.84		1.00	0.81	1.00	0.89	x_{ab}	0.55							
t_{ab} [ms]	(3.85±0.12)*10 ⁻⁶		(2.94±0.11)*10 ⁻⁶		(1.73±0.04)*10 ⁻⁶		(2.07±0.05)*10 ⁻⁶		t_{ab} [ms]	(5.37±3.58)*10 ⁻⁶							
x_{rc1}	0.16	0.16	0.16	0.16	0.40	0.21	0.29	0.27	x_{rc}	0.48	0.50	0.29	0.32	0.52	0.39	0.51	0.40
t_{rc1} [ms]	(2.66±0.03)*10 ⁻³				(1.14±0.02)*10 ⁻⁵				t_{rc} [ms]	1.0*10 ⁻⁵							
x_{rc2}	0.11	0.33	0.10	0.21	0.04	0.04	0.11	0.11	f	0.89							
t_{rc2} [ms]	(2.09±0.09)*10 ⁻⁵				(2.31±0.04)*10 ⁻⁴				x_{cab1}	0.06	0.06	0.06	0.06	0.02	0.02	0.02	0.02
									t_{cab1} [ms]	(1.97±0.75)*10 ⁻⁶							
x_{b1}			0.21±0.006		0.16±0.002		0.29±0.003		x_{cab2}	0.016±0.003							
t_{b1} [ms]	0.314±0.002				0.104±0.002				t_{cab2} [ms]	(4.10±0.80)*10 ⁻⁵							
x_{b2}	0.60±0.009		0.42±0.003		0.10±0.003		0.19±0.003		x_{cab3}	0.003±1.5*10 ⁻⁴							
t_{b2} [ms]	0.083±0.001				0.019±5*10 ⁻⁴				t_{cab3} [ms]	(3.32±0.36)*10 ⁻³							
x_{b3}	0.40±0.010		0.23±0.003		0.38±0.003		0.29±0.001		x_{cab4}	0.07±5*10 ⁻⁴							
t_{b3} [ms]	0.020±2*10 ⁻⁴				(4.16±0.04)*10 ⁻³				t_{cab4} [ms]	0.417±0.004							

THE DENATURED STATE OF T4L

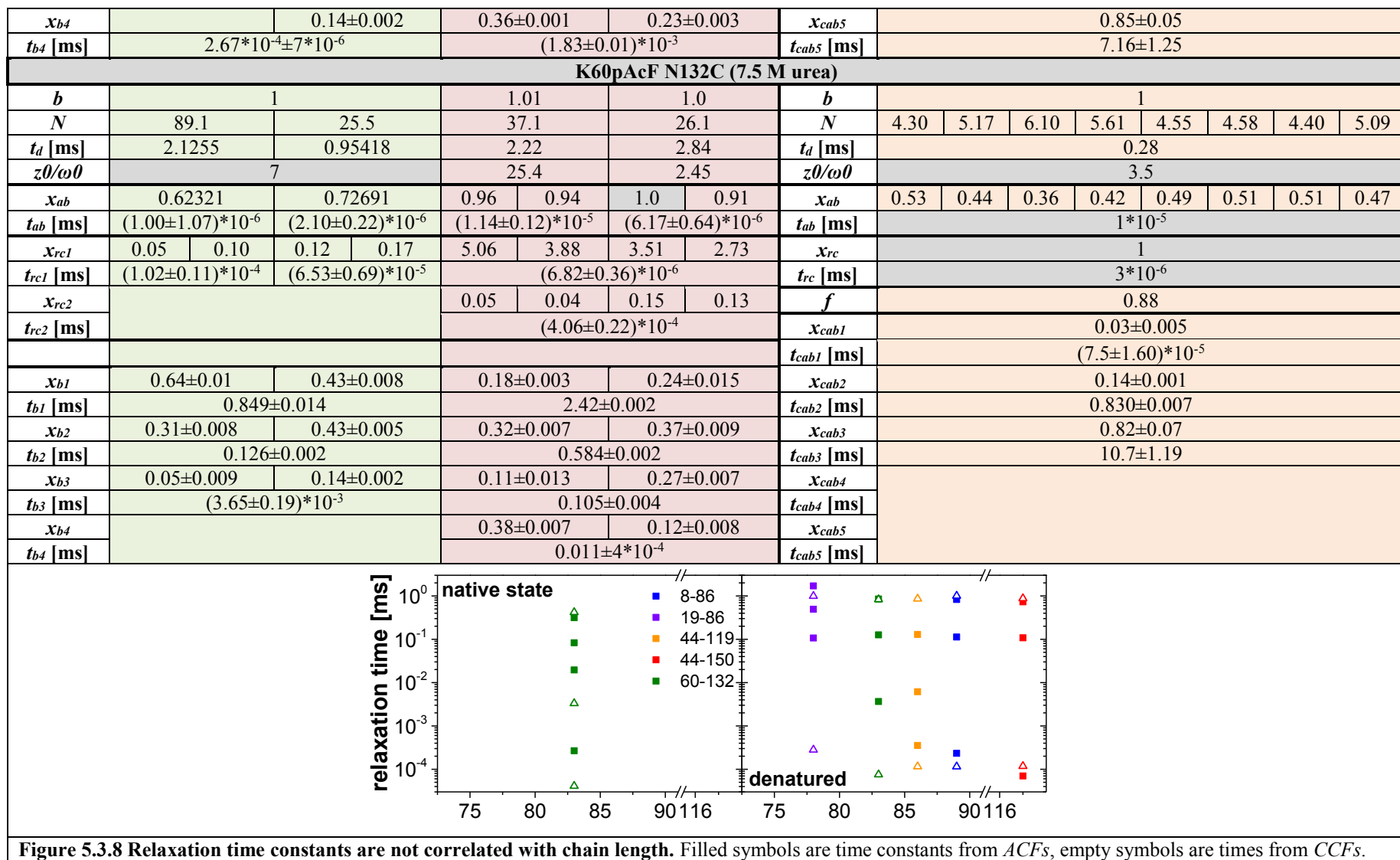
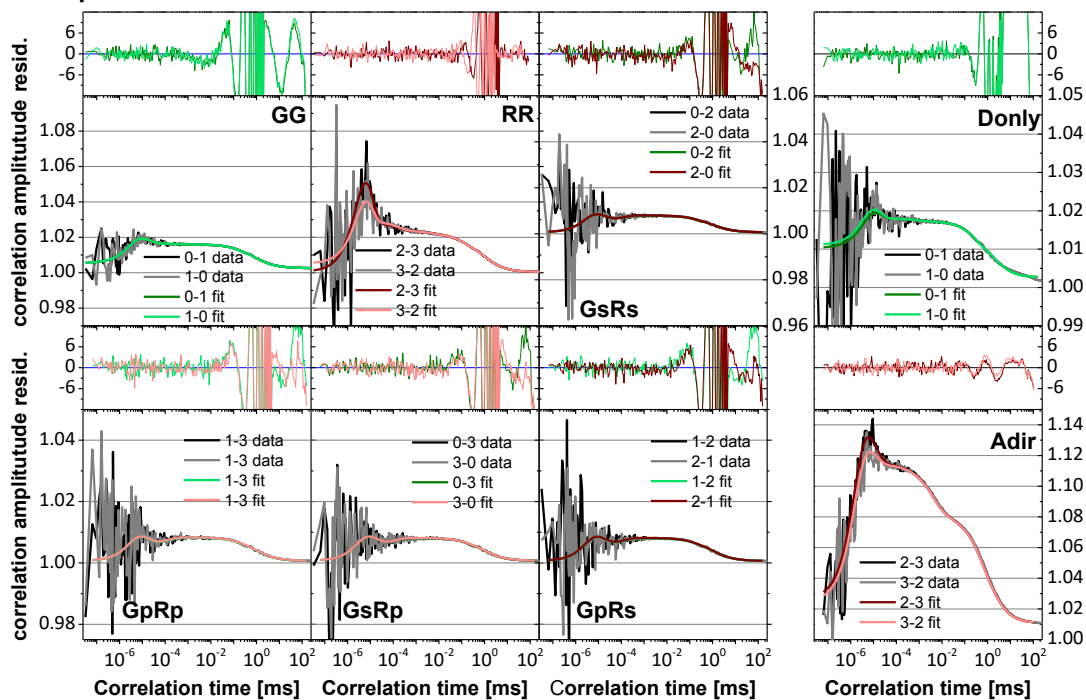
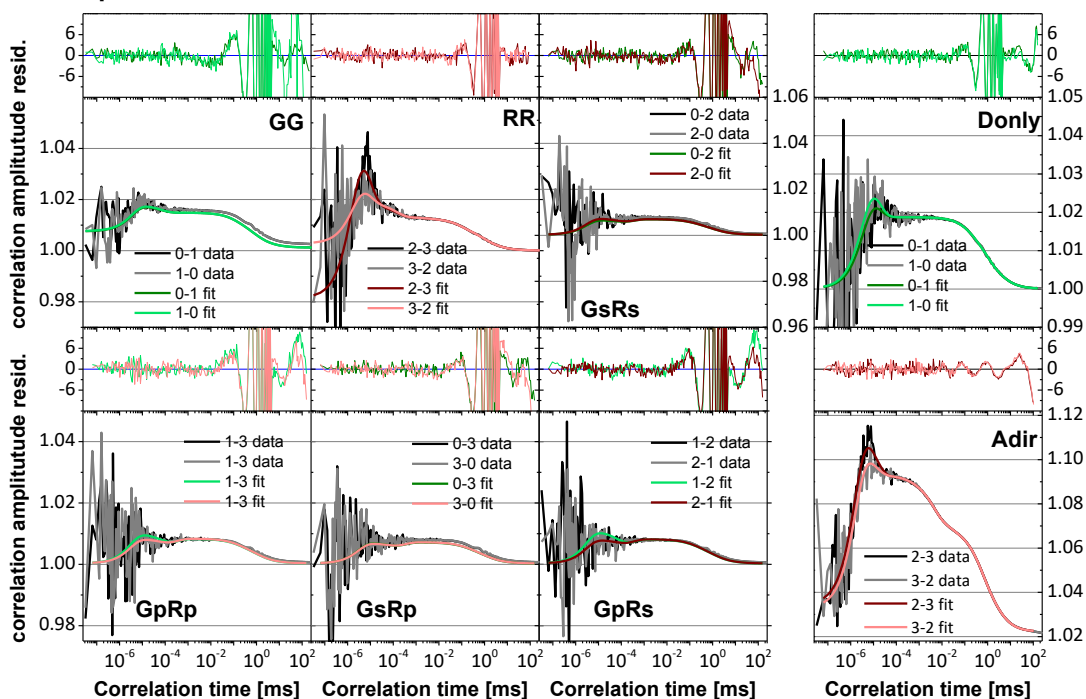


Figure 5.3.8 Relaxation time constants are not correlated with chain length. Filled symbols are time constants from ACFs, empty symbols are times from CCFs.

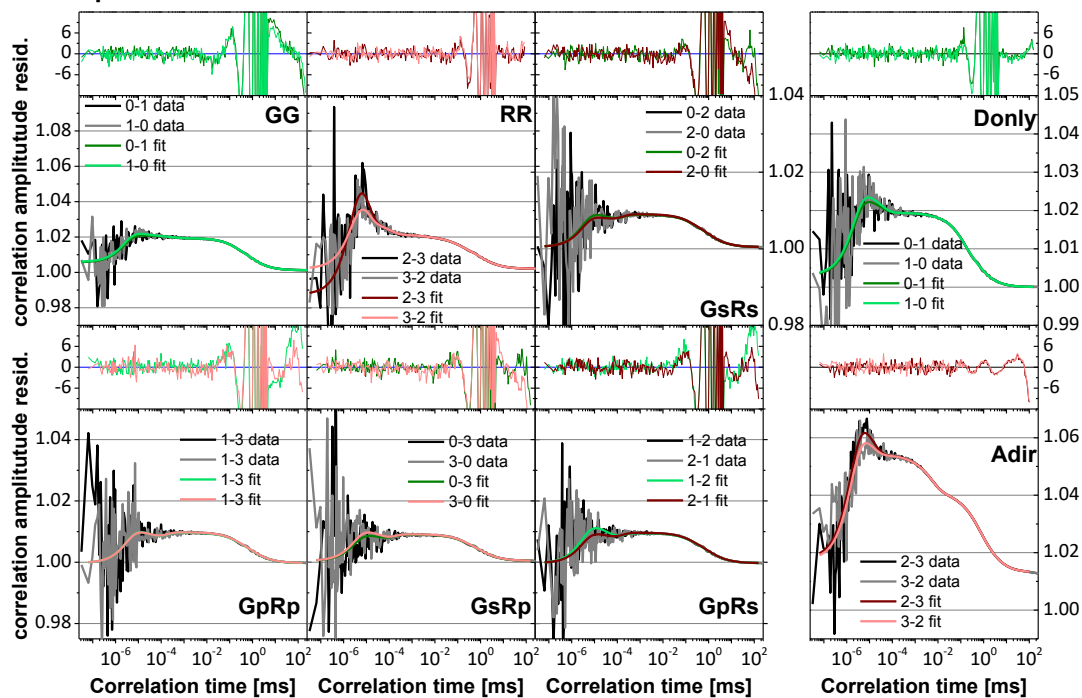
A R8pAcF N132C, 7.5 M urea



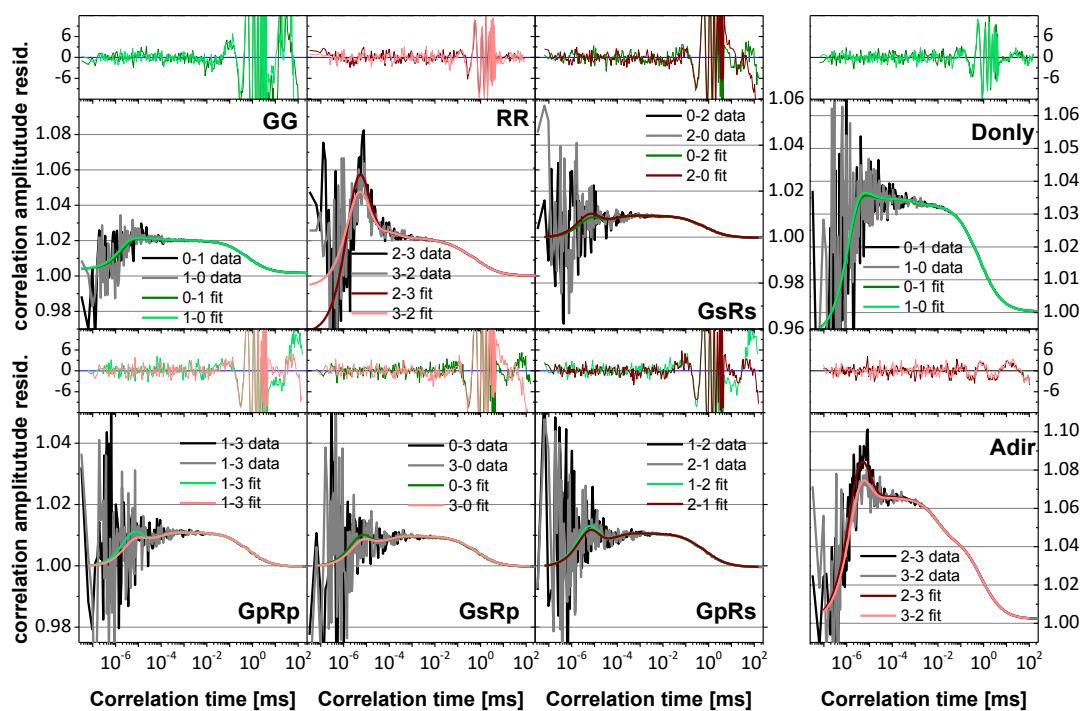
B K19pAcF N132C, 7.5 M urea



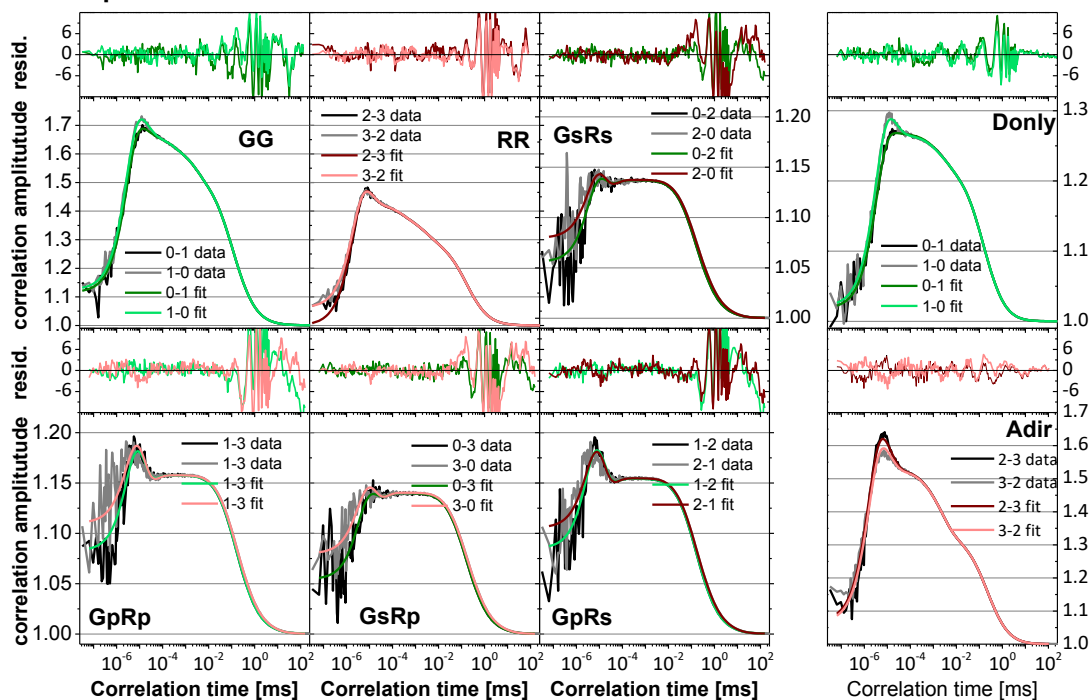
C S44pAcF R119C, 7.5 M urea



D S44pAcF I150C, 7.5 M urea



E K60pAcF N132C, 0 M urea



F K60pAcF N132C, 7.5 M urea

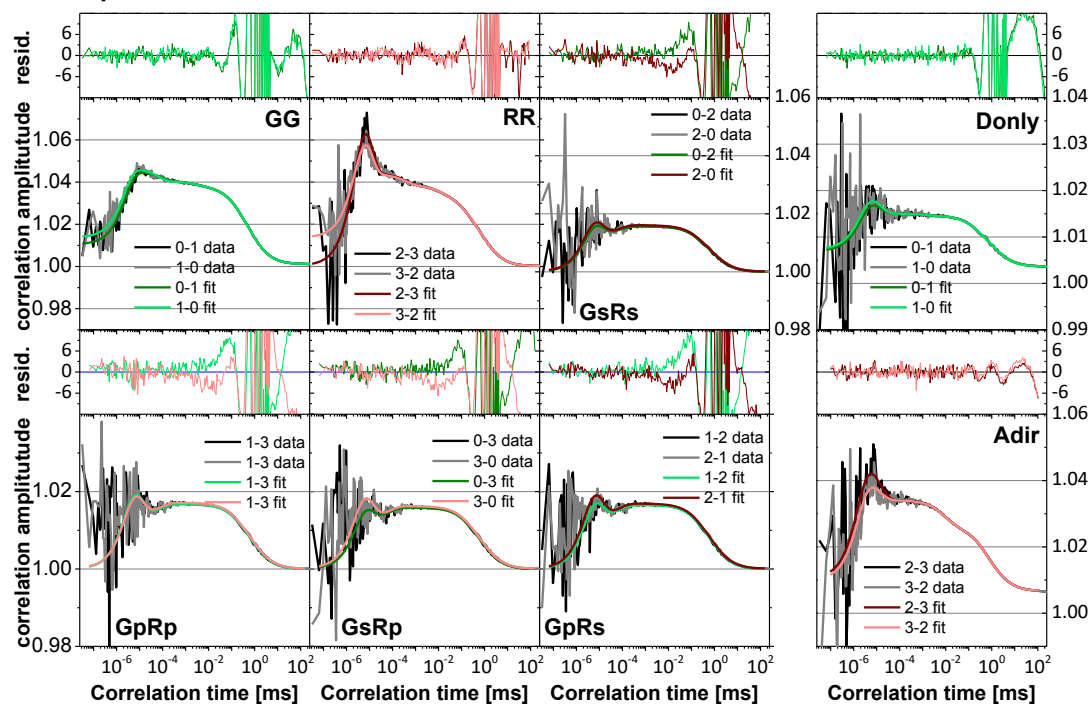
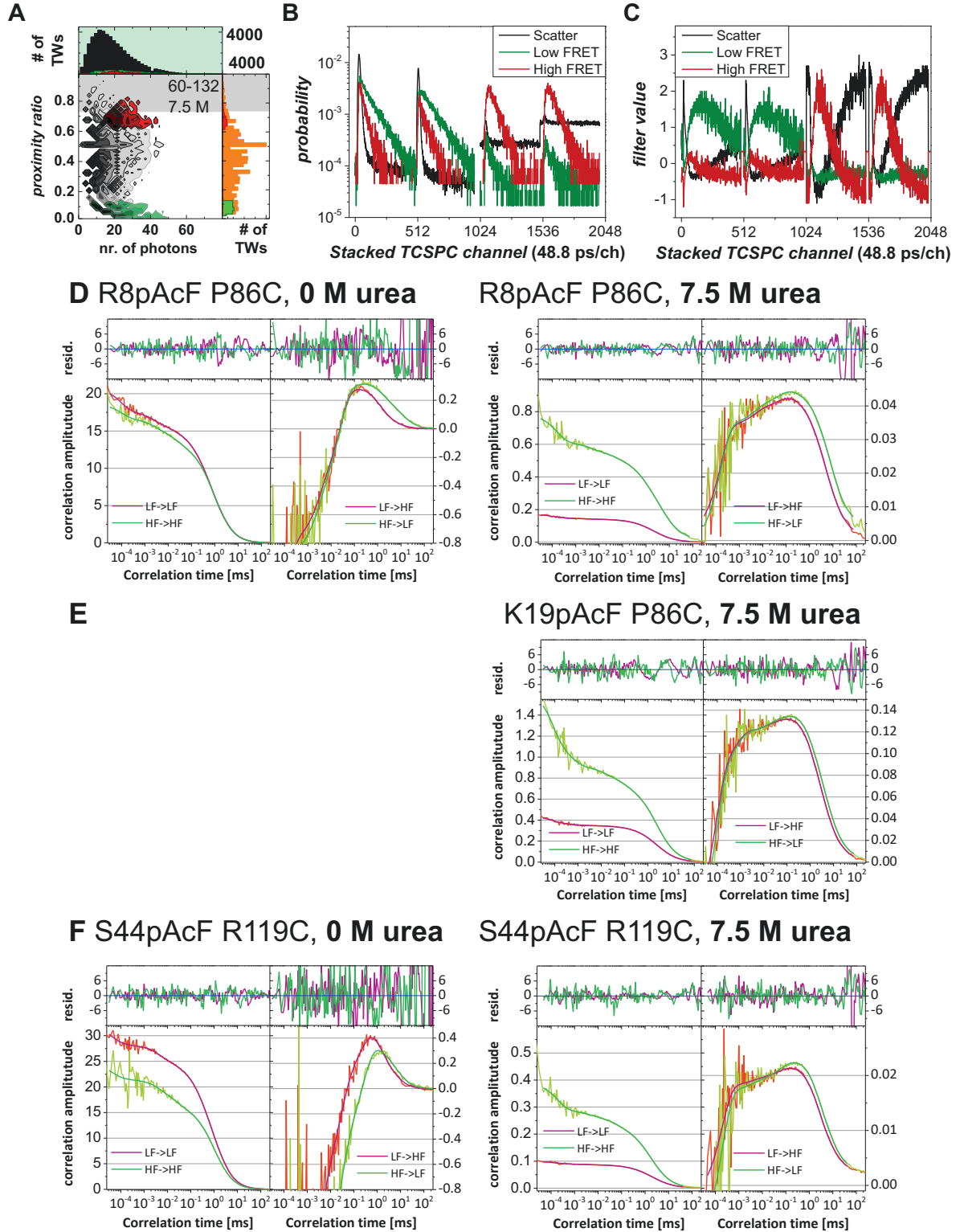


Figure 5.3.9 Full fluorescence Correlation functions.

5.3.2.8. Additional filteredFCS results



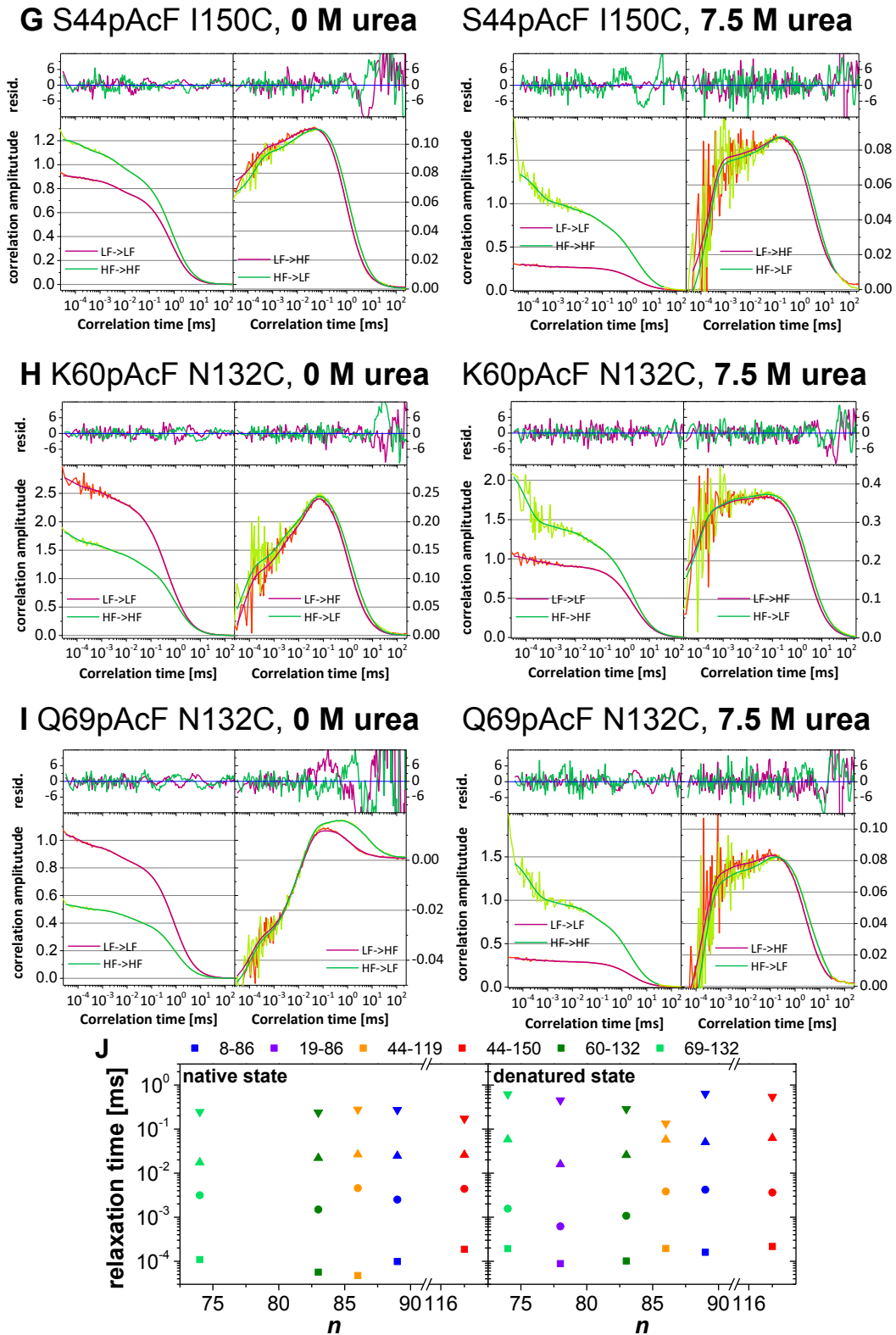


Figure 5.3.10 Example of filter generation for filteredFCS and filteredFCS curves. (A) Time-window ID's of low- (green) and high- (red) FRET species were selected in a 2D frequency histogram of number of photons vs. proximity ratio. (B) Generated decays for the two pseudo-species low FRET and high FRET in addition to the scatter profile. (C) Filters were calculated according to Eq. 3.3.24 in Chapter 3 using the decays from graph (B). (D-I) filteredFCS curves. Fit residuals are shown on top. (F) Relaxation times vs. number of residues between labeling sites n .

THE DENATURED STATE OF T4L

Table 5.3.7 Fitting results for filteredFCS measurements. Uncertainties given are standard errors of the mean. For a_4 , the average uncertainty of a_1 - a_3 are given (21.2 %).

[M urea]		b	N	t_d	$z0/w0$	f	t_{rel1} [ms]	a_1	t_{rel2} [ms]	a_2	t_{rel3} [ms]	a_3	t_{rel4} [ms]	a_4	a_{long}	t_{long} [ms]
R8pAcF P86C																
0	LF-HF	1.00	2.58	0.97 ±0.007	6.94	5.89	(9.84 ±2.11) E-05	0.47	(2.50 ±0.39) E-03	0.15	(2.46 ±0.08) E-02	0.33	0.273 ±0.008	0.05	0.01	7.8 ±0.08
	HF-LF	1.00	0.58			6.06		±0.03		±0.02		±0.02		±0.004	0.77	
	LF-LF	1.00	0.05					0.15		0.08		0.05		0.10	0.17	
	HF-HF	1.00	0.05					0.11		0.08		0.11		0.02	0.17	
7.5	LF-HF	1.00	21.12	4.07 ±0.02	15.1	0.99	(1.60 ±0.44) E-04	0.77	(4.20 ±1.98) E-03	0.12	(5.02 ±0.50) E-02	0.11	0.634 ±0.179	0.00	0.10	4.2 ±0.004
	HF-LF	1.00	16.92			1.16		±0.05		±0.03		±0.02		0.12	0.30	
	LF-LF	1.00	5.85					0.15		0.03		0.01		0.12	0.28	
	HF-HF	1.01	1.29					0.23		0.06		0.11		0.00		
K19pAcF P86C																
7.5	LF-HF	1.00	6.65	2.46 ±0.03	20.0	1.43	(8.82 ±1.15) E-05	0.71	(6.19 ±0.91) E-04	0.17	(1.58 ±0.06) E-02	0.05	0.134 ±0.007	0.07	0.05	3.0 ±0.003
	HF-LF	1.01	5.52			1.60		±0.01		±0.02		±0.002		±0.005	0.21	
	LF-LF	1.00	2.27					0.16		0.05		0.00		0.04	0.01	
	HF-HF	1.01	0.60					0.33		0.12		0.05		0.06	0.03	
S44pAcF R119C																
0	LF-HF	1.02	1.08	1.11 ±0.02	4.9	9.00	(4.74 ±0.49) E-05	0.65	(4.56 ±0.72) E-03	0.10	(2.65 ±0.21) E-02	0.13	0.277 ±0.007	0.11	0.27	3.9 ±0.09
	HF-LF	1.02	0.71			20.43		±0.01		±0.01		±0.01		±0.007	0.72	
	LF-LF	1.01	0.03					0.12		0.08		0.04		0.11	0.18	
	HF-HF	1.04	0.04					0.13		0.09		0.12		0.02	0.16	
7.5	LF-HF	1.01	39.44	2.41 ±0.005	11.0	1.21	(1.94 ±0.15) E-04	0.76	(3.83 ±0.54) E-03	0.05	(5.76 ±0.61) E-02	0.09	0.474 ±0.002	0.11	0.11	4.3 ±0.09
	HF-LF	1.01	31.50			1.59		±0.003		±0.002		±0.01		±0.006	0.24	
	LF-LF	1.00	10.13					0.11		0.02		0.02		0.07	0.05	
	HF-HF	1.01	2.54					0.26		0.05		0.09		0.06	0.17	
S44pAcF I150C																
0	LF-HF	1.00	7.82	0.85 ±0.009	3.7	0.45	(1.86 ±0.17) E-04	0.46	(4.40 ±0.14) E-03	0.14	(2.60 ±0.25) E-02	0.17	0.173 ±0.010	0.23	0.00	1.9 ±0.08
	HF-LF	1.00	5.64			0.52		±0.03		±0.02		±0.01		±0.02	0.27	
	LF-LF	1.00	1.10					0.03		0.11		0.03		0.05	0.09	
	HF-HF	1.00	0.81					0.08		0.11		0.06		0.03	0.00	
7.5	LF-HF	1.00	9.45	2.57 ±0.06	61.0	1.07	(2.17 ±0.12) E-04	0.76	(3.63 ±0.49) E-03	0.03	(6.23 ±0.42) E-02	0.12	0.541 ±0.035	0.10	0.06	5.2 ±0.005
	HF-LF	1.00	7.08			1.17		±0.008		±0.004		±0.007		0.27		
	LF-LF	1.00	3.27					0.10		0.03		0.02		0.08	0.06	
	HF-HF	1.04	0.73					0.27		0.07		0.10		0.07	0.00	

THE DENATURED STATE OF T4L

K60pAcF N132C																
0	LF-HF	1.03	3.62	0.94 ±0.05	4.95	1.17	(5.63 ±2.07) E-05	0.46	(1.49 ±0.18) E-03	0.18	(2.20 ±1.06) E-02	0.30	0.237 ±0.101	0.06	0.00	2.14 ±0.09
	HF-LF	1.03	2.64			1.04		±0.08		±0.004		±0.13		±0.01	0.27	
	LF-LF	1.02	0.35					0.09		0.07		0.06		0.19	0.25	
	HF-HF	1.04	0.51					0.15		0.08		0.11		0.02	0.05	
6.5	LF-HF	1.00	2.54	2.28 ±0.02	6.99	0.69	(1.01 ±0.34) E-04	0.75	(1.07 ±0.42) E-03	0.11	(2.56 ±0.63) E-02	0.06	0.286 ±0.021	0.08	0.00	4.3 ±0.17
	HF-LF	1.00	2.01			0.76		±0.04		±0.06		±0.02		±0.02	0.19	
	LF-LF	1.00	0.94					0.07		0.07		0.02		0.05	0.00	
	HF-HF	1.00	0.44					0.34		0.04		0.07		0.06	0.09	
Q69pAcF N132C																
0	LF-HF	1.00	65.70	0.92 ±0.02	6.20	4.48	(1.09 ±0.14) E-04	0.29	(3.14 ±0.20) E-03	0.20	(1.75 ±0.04) E-02	0.44	0.246 ±0.012	0.06	0.01	4.9 ±0.03
	HF-LF	1.00	11.34			4.26		±0.007		±0.02		±0.008		±0.003	0.81	
	LF-LF	1.00	0.94					0.09		0.08		0.04		0.04	0.17	
	HF-HF	1.00	1.84					0.07		0.05		0.11		0.02	0.12	
7.5	LF-HF	1.00	9.77	2.22 ±0.002	61.0	1.24	(1.93 ±0.21) E-04	0.75	(1.56 ±0.67) E-03	0.07	(5.82 ±0.23) E-02	0.08	0.619 ±0.001	0.10	0.00	17.6 ±0.18
	HF-LF	0.99	3.33			1.40		±0.003		±0.001		±0.003		±0.002	0.62	
	LF-LF	1.00	2.98					0.06		0.05		0.03		0.01	0.00	
	HF-HF	1.03	0.66					0.33		0.05		0.11		0.03	0.00	

5.3.2.9. Results from modelling of fluorophore distributions with FPS

The possible distribution of the fluorophores under native and denaturing conditions in accord with experimental results was modeled using FPS (5.3.1.10). As FRET restraints the mean distances $\langle R_{DA} \rangle$ obtained from ensemble measurements in Section 5.2.2.4 were used.

We calculated the pairwise distance distribution not only for our FRET-pairs and the structural-links but also for the indirect links (Table 5.3.8, Table 5.3.9), which were not measured or specified directly. For example, we measured the distance between 8-69, 8-132 and 69-119, and the distance 119-132 we linked via WLC in the denatured state. Thus, indirectly, we also put restraints on the connection 8-119 (Figure 5.3.11). Table 5.3.8 and Table 5.3.9 show the individual fluorophore distribution for all three cases for all possible labeling pairs, this makes clear how much information these combined restraints (FRET and simulation-model) deliver.

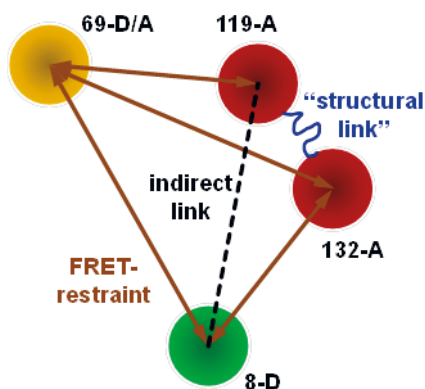


Figure 5.3.11 Different types of restraints build our modeling network.

Table 5.3.8 Residue pairs used to calculate the pairwise distance distributions of the native state shown in Figure 5.2.7A-B

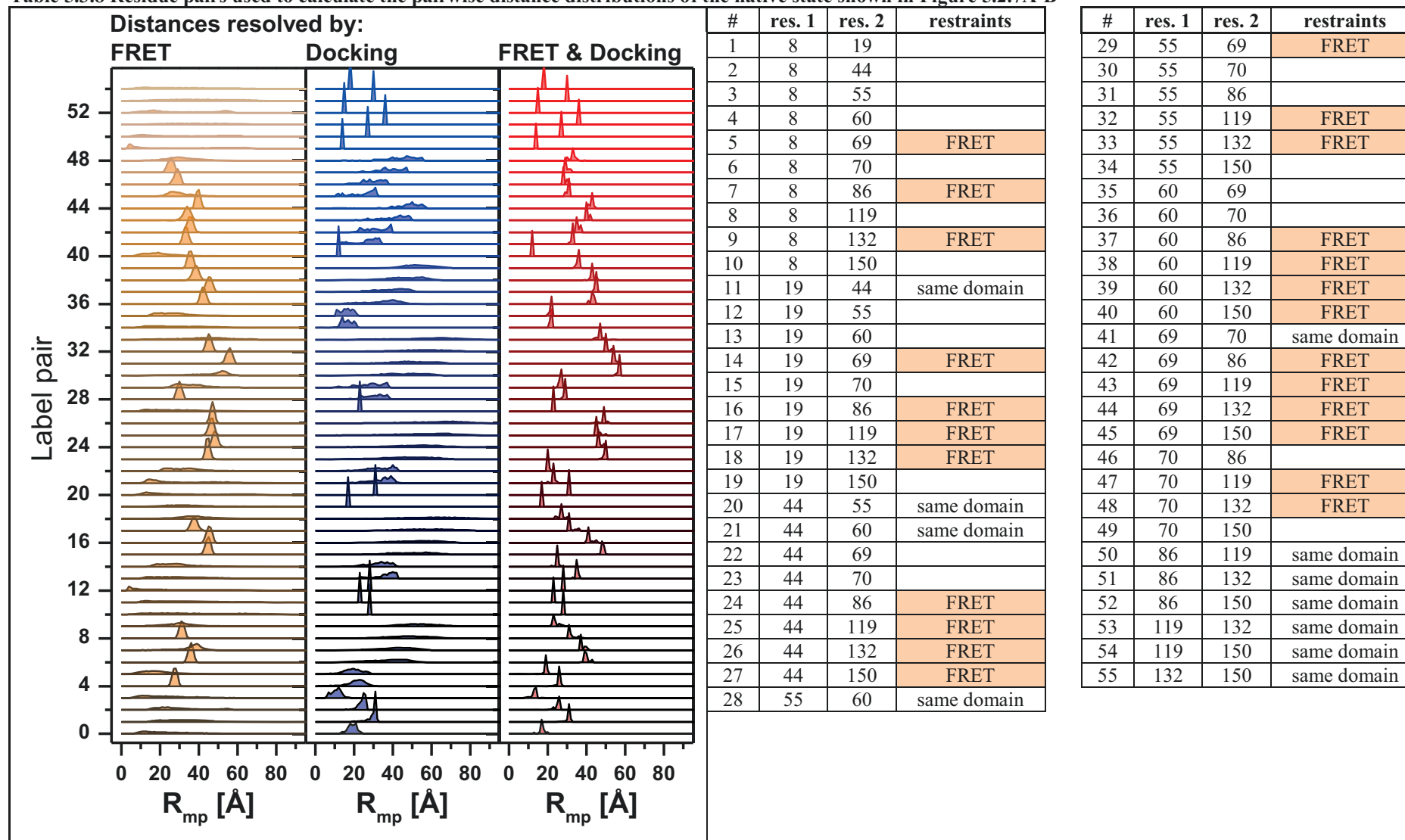
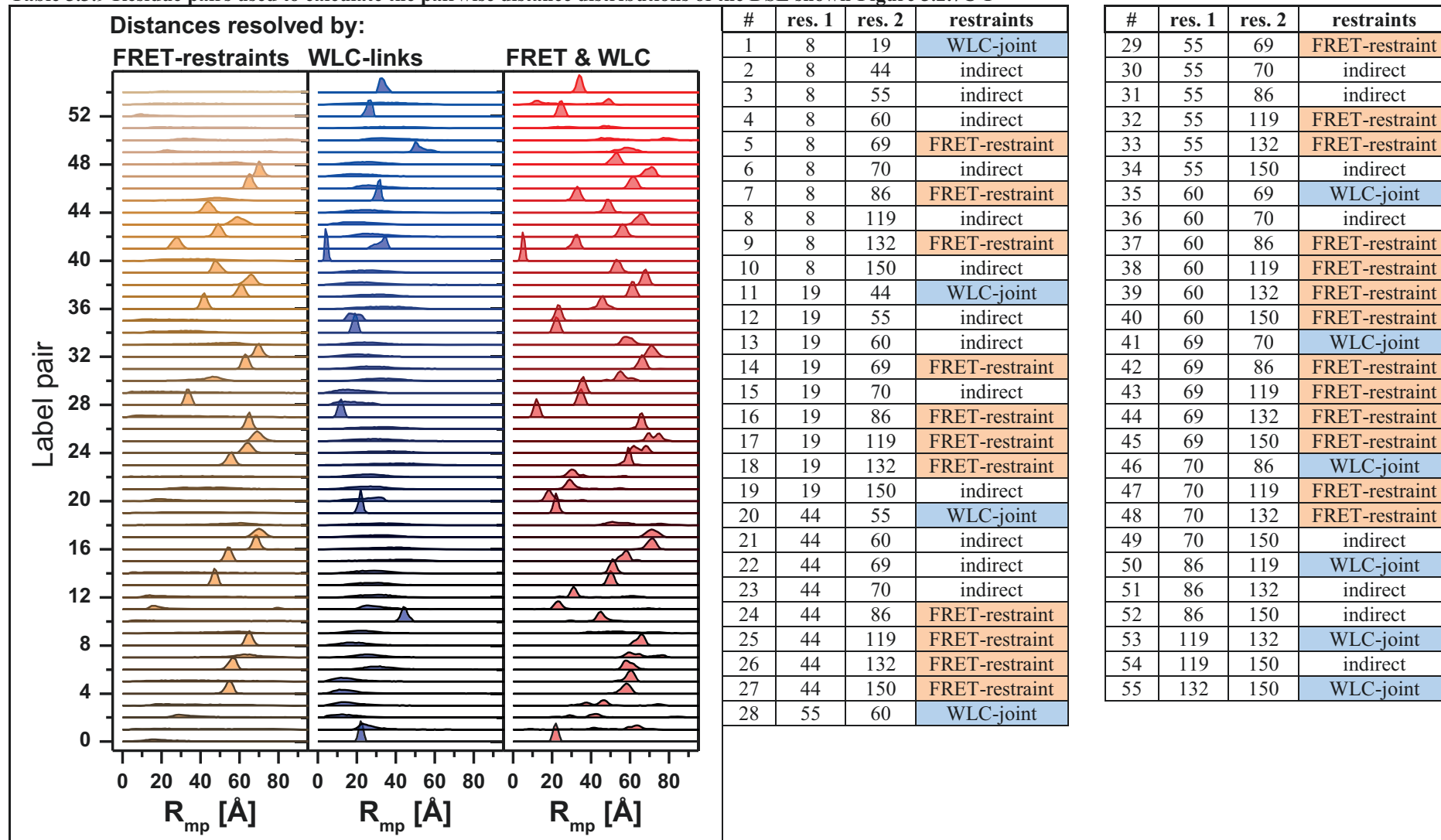


Table 5.3.9 Residue pairs used to calculate the pairwise distance distributions of the DSE shown Figure 5.2.7C-F



6. Summary & Outlook

In the presented work single-molecule and ensemble fluorescence methods were applied to the model enzyme Phage T4 Lysozyme (T4L) to characterize its behavior under native and denaturing conditions as well as while unfolding. The joint approach of (sub-)ensemble time-correlated single photon counting ((s)eTCSPC) for fluorescence lifetime and anisotropy studies, single-molecule multiparameter fluorescence detection (MFD), photon distribution analysis (PDA), full and filtered fluorescence correlation spectroscopy (fullFCS and filteredFCS) – combined with Circular Dichroism (CD) spectroscopy and computer simulations – provided simultaneously information about the protein structure and its conformational fluctuations.

The enzymatic pathway requires conformational changes and thus structural flexibility. We found that flexibility is associated to the N-terminal subdomain. We find that the N-terminal subdomain seems to unfold first. These results are selfconsistent and in line with NMR studies, which found under denaturing conditions an unfolded N-terminal and a folded C-terminal subdomain. Additionally, we find a hierarchically unfolding of T4L, while applying multiple methods that allow disentangling the unfolding process with so far unachieved resolution.

6.1. Structure and function of T4L under native conditions

From T4L it is known that its two subdomains undergo a hinge-bending motion while processing its substrate peptidoglycan. Our goal was to study the dynamics of this motion by a network of FRET-pairs and to relate the observed transitions and conformational states to distinct steps in the enzymatic cleavage cycle. To facilitate the latter, we created additionally T4L mutants, which altered the enzymatic activity.

By globally analyzing the time-resolved fluorescence intensities of the network of 24 FRET variants, three distinct conformations (C_1 , C_2 and C_3) were identified. Screening the obtained distance sets against 380 X-ray structures of T4L deposited in the PDB, we could show that C_1 and C_2 resemble the known “open” and “closed” state of T4L. However, the minor state C_3 is a thermally excited state and the derived FRET-restraints for this state could not be related to any of the PDB structures. C_3 can thus be considered a “hidden” excited state as it elucidated detection before. Using mutants altering the enzymatic activity of T4L, the

conformational states and their fluctuations were assigned to the enzymatic reaction states. Thus, the energy landscape of the catalytic cleavage cycle could be derived.

Here, the importance of describing the function of thermally driven enzymatic reactions by state matrices, which allow assigning functional roles to conformations, was shown. In the view of the Michaelis-Menten kinetics based description, the shown results stress the importance of considering also short-lived functional states in kinetic descriptions, i.e. short lived (excited states) product release states might exist for many other enzymes as well. The next step will be to use the gained distance set for C_3 to derive the structure of this hidden conformation of T4L by computational methods.

6.2. Unfolding pathway of T4L

We showed that the native state of T4L consists of three conformational states in equilibrium and it has been shown that T4L unfolds via at least two intermediates. Hence, we asked how this subdomain composition of T4L and the equilibrium of native states, especially the presence of thermally excited states, might influence the unfolding pathway of proteins. To understand this, we aim to disentangle the stepwise loss of structural elements and the fast folding-unfolding transitions in T4L. This helps us to assess the structural features and the stability of the partially folded intermediate structures.

Our combined approach of CD spectroscopy and fluorescence measurements – supported by *a priori* knowledge from earlier studies – allowed disentangling the individual unfolding steps of T4L via at least two intermediates. Here, using our network of variants we show the sequential, stepwise loss in the structural elements based on the relative stability of secondary structure and tertiary arrangements. By comparing the stabilities of our three native states with the stabilities for the two intermediates and the unfolded states derived by Cellitti *et al.* (20), we propose that thermally excited states, which are already sampled under native conditions, and the presence of a conformational equilibrium of native states might play a critical role and could facilitate the unfolding of (sub-) domain-wise organized proteins like T4L. Here, we identified C_3 as a thermally excited state, which is closest in energy to J . Thus, the unfolding of T4L might lead over C_3 . The structure of C_3 is still unknown, but preliminary results from FRET-restrained MD simulations showed a minor loss of structure within the C-terminal subdomain accompanying a reorientation of the two subdomains. Additionally, the fluctuation analysis was able to fill the existing gaps in the fast kinetic timescales, and thus we could derive a complemented reaction scheme, which covered all potentially relevant

timescales – compared to the earlier proposed scheme proposed by stopped-flow measurements.

However, so far one of the fluorophores was coupled to the N- and the other to the C-terminal subdomain of T4L. Thus, the observed changes in FRET involved the transition over the whole protein. To be able to unequivocally assign also the sequential loss of structure within the more stable C-terminal subdomain, the next step will be to design new FRET pairs with different fluorophores such that the short distances (20 Å) expected within the single subdomains can be measured.

6.3. Characterization of the denatured state of T4L

The characterization of the denatured or unstructured state of proteins is a challenge due to their heterogeneity of the ensemble, which consists of a vast number of quickly interconverting conformations. Transiently formed local structure of protein may serve as nucleation site for folding. To identify potential sites and to study internal dynamics, we use multidimensional fluorescence methods. This allows answering the question whether a denatured state of a protein is ordered or if it behaves like an ideal polymer.

Using the full network of 24 variants generated within the first part of this work, ensemble fluorescence lifetime and anisotropy and single-molecule methods were used to characterize the ensemble of denatured states from T4L. Firstly, we showed that chemically denatured T4L in 7.5 M urea does not behave like an ideal polymer, but it represents more a heterogeneous ensemble of conformations presumably all displaying varying degrees of structural order. This leads to the observed heterogeneity on the single-molecule level and to the physically meaningless short chain length in polymer models. By building a stiffness matrix and evaluating the local side chain mobility with fluorescence anisotropy, we identified regions in T4L, which showed apparent residual order and are thus likely candidates for nucleation sites. Additionally, we could derive a possible distribution of mean fluorophore positions within T4L's DSE and, by comparison to the native state, we found that the individual subdomains, specifically the C-terminal subdomain, likely show a native-like organization, albeit of larger size. To obtain a higher resolution model, more FRET-distances will have to be added to our network, such that all positions can be unequivocally trilaterated and also distance measurements within one subdomain would need to be added. The latter would help to resolve the different characteristics of the two subdomains, also interesting generally with respect to the organization of proteins out of smaller, flexibly connected

structural elements. Finally, correlation analysis of local quenching and global motions provided the rough energy landscape within the DSE.

The combination of methods used here to detect transient structural elements and to describe the dynamic landscape of an unordered protein can be used as a powerful tool to characterize the class of intrinsically disordered proteins. These IDP's have been recognized as important key players within cell cycle or signaling processes. Additionally, the generated network of restraints can be implemented in simulation software to generate or screen for structures potentially describing the ensemble of states.

6.4. Fluorescence as a complementary tool in structural biology

All in all, in this work it could be shown that the combination of several different fluorescence methods (eTCSPC, MFD, PDA, fullFCS and filteredFCS) is well suited to firstly detect and describe fast dynamics (e.g. enzymatic cleavage under native conditions, unfolding dynamics) over more than seven orders of magnitude in time, and secondly to detect so far “hidden” excited, i.e. minuscule populated, states. The structural and dynamic landscape of T4L was explored under native and denaturing conditions while at the same time obtaining distance and dynamics information. In this work still existing gaps in the knowledge of the very well-studied model enzyme T4L could be filled – proving the outstanding potential of fluorescence in the field of structure biology where classical structural biology methods reach their limits.

Additionally, the distance restraints gained for the network of variants under native and denaturing conditions could be implemented in simulations to selectively choose or generate (an ensemble of) structures. So far, this has been applied to restraints generated from different NMR, EPR or SAXS measurements. Furthermore, the fluctuation analysis methods complement the analysis of kinetics on timescales inaccessible for NMR or Stopped-flow experiments.

7. References

1. Querfurth HW & LaFerla FM (2010) Alzheimer's disease. *N Engl J Med* 362:329-344.
2. Kay LE (2005) NMR studies of protein structure and dynamics. *J Magn Reson* 173:193-207.
3. Neudecker P, Lundstrom P, & Kay LE (2009) Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding. *Biophys J* 96:2045-2054.
4. Henzler-Wildman K & Kern D (2007) Dynamic personalities of proteins. *Nature* 450:964-972.
5. Tompa P (2012) On the supertertiary structure of proteins. *Nat Chem Biol* 8:597-600.
6. Hensen U, *et al.* (2012) Exploring protein dynamics space: the dynasome as the missing link between protein structure and function. *PLoS One* 7:e33931.
7. Santoro MM & Bolen DW (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* 27:8063-8068.
8. Kim PS & Baldwin RL (1990) Intermediates in the folding reactions of small proteins. *Annu Rev Biochem* 59:631-660.
9. Bryngelson JD, Onuchic JN, Socci ND, & Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 21:167-195.
10. Dill KA (1999) Polymer principles and protein folding. *Protein Sci* 8:1166-1180.
11. Campos LA, *et al.* (2013) Gradual disordering of the native state on a slow two-state folding protein monitored by single-molecule fluorescence spectroscopy and NMR. *J Phys Chem B* 117:13120-13131.
12. Fleming PJR, George D. (2004) Conformational Properties of Unfolded Proteins. *Protein Folding Handbook*, ed T. BJK (WILEY -VCH Verlag GmbH & Co, KGaA), Vol I, pp 706-732.
13. Flory PJ (1969) *Statistical Mechanics of Chain Molecules* (Wiley).
14. Chan HS & Dill KA (1991) Polymer principles in protein structure and stability. *Annu Rev Biophys Chem* 20:447-490.
15. Chen Y, Wedemeyer WJ, & Lapidus LJ (2010) A general polymer model of unfolded proteins under folding conditions. *J Phys Chem B* 114:15969-15975.
16. O'Brien EP, Morrison G, Brooks BR, & Thirumalai D (2009) How accurate are polymer models in the analysis of Forster resonance energy transfer experiments on proteins? *J Chem Phys* 130:124903.
17. Shortle D & Ackerman MS (2001) Persistence of native-like topology in a denatured protein in 8 M urea. *Science* 293:487-489.
18. Fitzkee NC & Rose GD (2004) Reassessing random-coil statistics in unfolded proteins. *Proc Natl Acad Sci USA* 101:12497-12502.
19. Munoz V (2007) Conformational dynamics and ensembles in protein folding. *Annu Rev Biophys Biomol Struct* 36:395-412.
20. Cellitti J, Bernstein R, & Marqusee S (2007) Exploring subdomain cooperativity in T4 lysozyme II: uncovering the C-terminal subdomain as a hidden intermediate in the kinetic folding pathway. *Protein Sci* 16:852-862.
21. Kato H, Feng H, & Bai Y (2007) The folding pathway of T4 lysozyme: the high-resolution structure and folding of a hidden intermediate. *J Mol Biol* 365:870-880.
22. Vendruscolo M (2007) Determination of conformationally heterogeneous states of proteins. *Curr Opin Struct Biol* 17:15-20.
23. Zeeb M & Balbach J (2005) Millisecond protein folding studied by NMR spectroscopy. *Protein Pept Lett* 12:139-146.
24. Palmer AG, 3rd (2004) NMR characterization of the dynamics of biomacromolecules. *Chem Rev* 104:3623-3640.
25. Sisamakias E, Valeri A, Kalinin S, Rothwell PJ, & Seidel CA (2010) Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods Enzymol* 475:455-514.

REFERENCES

26. Richardson JS & Richardson DC (2014) Biophysical highlights from 54 years of macromolecular crystallography. *Biophys J* 106:510-525.
27. Baase WA, Liu L, Tronrud DE, & Matthews BW (2010) Lessons from the lysozyme of phage T4. *Protein Sci* 19:631-641.
28. Matthews BW (1996) Structural and genetic analysis of the folding and function of T4 lysozyme. *FASEB J* 10:35-41.
29. Matthews BW (1995) Studies on protein stability with T4 lysozyme. *Adv Protein Chem* 46:249-278.
30. Alber T (1989) Mutational effects on protein stability. *Annu Rev Biochem* 58:765-798.
31. Poteete AR & Hardy LW (1994) Genetic analysis of bacteriophage T4 lysozyme structure and function. *J Bacteriol* 176:6783-6788.
32. Tsugita A & Inouye M (1968) Purification of bacteriophage T4 lysozyme. *J Biol Chem* 243:391-397.
33. Matthews BW & Remington SJ (1974) The three dimensional structure of the lysozyme from bacteriophage T4. *Proc Natl Acad Sci U S A* 71:4178-4182.
34. 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.
35. Kuroki R, Weaver LH, & Matthews BW (1993) A covalent enzyme-substrate intermediate with saccharide distortion in a mutant T4 lysozyme. *Science* 262:2030-2033.
36. 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.
37. Yirdaw RB & McHaourab HS (2012) Direct observation of T4 lysozyme hinge-bending motion by fluorescence correlation spectroscopy. *Biophys J* 103:1525-1536.
38. Shoichet BK, Baase WA, Kuroki R, & Matthews BW (1995) A relationship between protein stability and protein function. *Proc Natl Acad Sci U S A* 92:452-456.
39. Elwell M & Schellman J (1975) Phage T4 lysozyme. Physical properties and reversible unfolding. *Biochim Biophys Acta* 386:309-323.
40. Desmadril M & Yon JM (1984) Evidence for intermediates during unfolding and refolding of a two-domain protein, phage T4 lysozyme: equilibrium and kinetic studies. *Biochemistry* 23:11-19.
41. Llinas M & Marqusee S (1998) Subdomain interactions as a determinant in the folding and stability of T4 lysozyme. *Protein Sci* 7:96-104.
42. Llinas M, Gillespie B, Dahlquist FW, & Marqusee S (1999) The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nat Struct Biol* 6:1072-1078.
43. Kato H, Vu ND, Feng H, Zhou Z, & Bai Y (2007) The folding pathway of T4 lysozyme: an on-pathway hidden folding intermediate. *J Mol Biol* 365:881-891.
44. Cellitti J, *et al.* (2007) Exploring subdomain cooperativity in T4 lysozyme I: structural and energetic studies of a circular permutant and protein fragment. *Protein Sci* 16:842-851.
45. Lu J & Dahlquist FW (1992) Detection and characterization of an early folding intermediate of T4 lysozyme using pulsed hydrogen exchange and two-dimensional NMR. *Biochemistry* 31:4749-4756.
46. Kalinin S, *et al.* (2012) A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat Methods* 9:1218-1225.
47. Felekyan S, Kalinin S, Sanabria H, Valeri A, & Seidel CA (2012) Filtered FCS: species auto- and cross-correlation functions highlight binding and dynamics in biomolecules. *Chemphyschem* 13:1036-1053.
48. Kalinin S, Felekyan S, Valeri A, & Seidel CA (2008) Characterizing multiple molecular States in single-molecule multiparameter fluorescence detection by probability distribution analysis. *J Phys Chem B* 112:8361-8374.
49. Kühnemuth R & Seidel CAM (2001) Principles of Single Molecule Multiparameter Fluorescence Spectroscopy. *Single Molecules* 2:251-254.

REFERENCES

50. Tompa P (2010) *Structure and function of intrinsically disordered proteins* (CRC Press, Boca Raton); 1 Ed pp 31-84.
51. Mittag T & Forman-Kay JD (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* 17:3-14.
52. Walla PJ (2009) *Modern biophysical chemistry* (Wiley-VCH, Weinheim); 1 Ed.
53. Sindbert S, *et al.* (2011) 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.
54. Weber G (1952) Polarization of the fluorescence of macromolecules. I. Theory and experimental method. *Biochem J* 51:145-155.
55. Förster T (1948) Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik* 437:55-75.
56. Moerner WE & Fromm DP (2003) Methods of single-molecule fluorescence spectroscopy and microscopy. *Rev Sci Instrum* 74:3597-3619.
57. Widengren J, *et al.* (2006) Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. *Anal Chem* 78:2039-2050.
58. Keller RA, *et al.* (1996) Single-Molecule Fluorescence Analysis in Solution. *Applied Spectroscopy* 50:12A-32A.
59. Eggeling C, Fries JR, Brand L, Gunther R, & Seidel CA (1998) Monitoring conformational dynamics of a single molecule by selective fluorescence spectroscopy. *Proc Natl Acad Sci U S A* 95:1556-1561.
60. Margittai M, *et al.* (2003) Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. *Proc Natl Acad Sci U S A* 100:15516-15521.
61. Rothwell PJ, *et al.* (2003) Multiparameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 reverse transcriptase:primer/template complexes. *Proc Natl Acad Sci U S A* 100:1655-1660.
62. Maus M, *et al.* (2001) An experimental comparison of the maximum likelihood estimation and nonlinear least-squares fluorescence lifetime analysis of single molecules. *Anal Chem* 73:2078-2086.
63. Kalinin S, Sisamakias E, Magennis SW, Felekyan S, & Seidel CA (2010) On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits. *J Phys Chem B* 114:6197-6206.
64. Antonik M, Felekyan S, Gaiduk A, & Seidel CA (2006) Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J Phys Chem B* 110:6970-6978.
65. Kalinin S, Valeri A, Antonik M, Felekyan S, & Seidel CA (2010) 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.
66. Elson EL & Magde D (1974) Fluorescence correlation spectroscopy. I. Conceptual basis and theory. *Biopolymers* 13:1-27.
67. Magde D, Elson E, & Webb WW (1972) Thermodynamic Fluctuations in a Reacting System\char22{}Measurement by Fluorescence Correlation Spectroscopy. *Physical Review Letters* 29:705-708.
68. Schwille P, Meyer-Almes FJ, & Rigler R (1997) Dual-color fluorescence cross-correlation spectroscopy for multicomponent diffusional analysis in solution. *Biophys J* 72:1878-1886.
69. Felekyan S, *et al.* (2005) Full correlation from picoseconds to seconds by time-resolved and time-correlated single photon detection. *Rev Sci Instrum* 76.
70. Felekyan S, Sanabria H, Kalinin S, Kuhnemuth R, & Seidel CA (2013) Analyzing Forster resonance energy transfer with fluctuation algorithms. *Methods Enzymol* 519:39-85.
71. Böhmer M, Wahl M, Rahn HJ, Erdmann R, & Enderlein J (2002) Time-resolved fluorescence correlation spectroscopy. *Chemical Physics Letters* 353:439-445.

REFERENCES

72. Kapusta P, Wahl M, Benda A, Hof M, & Enderlein J (2007) Fluorescence lifetime correlation spectroscopy. *J Fluoresc* 17:43-48.
73. Muschielok A, *et al.* (2008) A nano-positioning system for macromolecular structural analysis. *Nat Methods* 5:965-971.
74. Hammes GG (1964) Mechanism of Enzyme Catalysis. *Nature* 204:342-343.
75. Smock RG & Gierasch LM (2009) Sending signals dynamically. *Science* 324:198-203.
76. Selkoe DJ (2003) Folding proteins in fatal ways. *Nature* 426:900-904.
77. 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. *Proc Natl Acad Sci U S A* 106:11837-11844.
78. Kleckner IR & Foster MP (2011) An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta* 1814:942-968.
79. 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.
80. Woźniak AK, Schröder G, Grubmüller H, Seidel CAM, & Oesterhelt F (2008) Single molecule FRET measures bends and kinks in DNA. *Proc Natl Acad Sci USA* 105:18337-18342.
81. 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.
82. 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. *Journal of Molecular Biology* 308:745-764.
83. Zheng D & Lu HP (2014) Single-molecule enzymatic conformational dynamics: spilling out the product molecules. *J Phys Chem B* 118:9128-9140.
84. Michaelis L, Menten ML, Johnson KA, & Goody RS (2011) The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. *Biochemistry* 50:8264-8269.
85. Kou SC, Cherayil BJ, Min W, English BP, & Xie XS (2005) Single-molecule Michaelis-Menten equations. *Journal of Physical Chemistry B* 109:19068-19081.
86. 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.
87. Lange OF & Grubmüller H (2008) Full correlation analysis of conformational protein dynamics. *Proteins* 70:1294-1312.
88. Zacharias M (2008) Combining elastic network analysis and molecular dynamics simulations by hamiltonian replica exchange. *Journal of Chemical Theory and Computation* 4:477-487.
89. Bouvignies G, *et al.* (2011) Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* 477:111-114.
90. Lu HP (2011) Revealing time bunching effect in single-molecule enzyme conformational dynamics. *Phys Chem Chem Phys* 13:6734-6749.
91. Choi Y, *et al.* (2012) Single-molecule lysozyme dynamics monitored by an electronic circuit. *Science* 335:319-324.
92. Akhterov MV, *et al.* (2015) Observing lysozyme's closing and opening motions by high-resolution single-molecule enzymology. *ACS Chem Biol* 10:1495-1501.
93. Lu M & Lu HP (2014) Probing protein multidimensional conformational fluctuations by single-molecule multiparameter photon stamping spectroscopy. *J Phys Chem B* 118:11943-11955.
94. 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.
95. Fleissner MR, *et al.* (2009) Site-directed spin labeling of a genetically encoded unnatural amino acid. *Proc Natl Acad Sci U S A* 106:21637-21642.
96. Chiang YW, Borbat PP, & Freed JH (2005) The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. *J Magn Reson* 172:279-295.

REFERENCES

97. Chiang YW, Borbat PP, & Freed JH (2005) Maximum entropy: a complement to Tikhonov regularization for determination of pair distance distributions by pulsed ESR. *J Magn Reson* 177:184-196.
98. Klein-Seetharaman J, *et al.* (2002) Long-range interactions within a nonnative protein. *Science* 295:1719-1722.
99. Gassner NC, *et al.* (2003) Multiple methionine substitutions are tolerated in T4 lysozyme and have coupled effects on folding and stability. *Biophys Chem* 100:325-340.
100. Vetter IR, *et al.* (1996) Protein structural plasticity exemplified by insertion and deletion mutants in T4 lysozyme. *Protein Sci* 5:2399-2415.
101. Rennell D, Bouvier SE, Hardy LW, & Poteete AR (1991) Systematic mutation of bacteriophage T4 lysozyme. *J Mol Biol* 222:67-88.
102. Chen Y, Hu DH, Vorpapel ER, & Lu HP (2003) Probing single-molecule T4 lysozyme conformational dynamics by intramolecular fluorescence energy transfer. *Journal of Physical Chemistry B* 107:7947-7956.
103. Fersht A (1998) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W. H. Freeman, New York); First Ed.
104. Ratzke C, Hellenkamp B, & Hugel T (2014) Four-colour FRET reveals directionality in the Hsp90 multicomponent machinery. *Nat Commun* 5:4192.
105. Vale RD & Oosawa F (1990) Protein motors and Maxwell's demons: does mechanochemical transduction involve a thermal ratchet? *Adv Biophys* 26:97-134.
106. Gohlke H, *et al.* (2013) Binding region of alanopine dehydrogenase predicted by unbiased molecular dynamics simulations of ligand diffusion. *Journal of Chemical Information and Modeling* 53:2493-2498.
107. Lemke EA (2011) Site-specific labeling of proteins for single-molecule FRET measurements using genetically encoded ketone functionalities. *Methods Mol Biol* 751:3-15.
108. Becker W (2005) *Advanced Time-Correlated Single Photon Counting Techniques* p 331.
109. Lakowicz JR (2006) *Principles of Fluorescence Spectroscopy* (Springer, New York); Third Ed.
110. Magde D, Elson EL, & Webb WW (1974) Fluorescence correlation spectroscopy. II. An experimental realization. *Biopolymers* 13:29-61.
111. 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.
112. Schwille P (2001) Fluorescence Correlation Spectroscopy. Theory and Applications., ed E.L. Elson RR (Springer, Berlin), pp 360-378.
113. Schaffer J, *et al.* (1999) Identification of single molecules in aqueous solution by time-resolved fluorescence anisotropy. *J Phys Chem A* 103:331-336.
114. Koshioka M, Sasaki K, & Masuhara H (1995) Time-Dependent Fluorescence Depolarization Analysis in Three-Dimensional Microspectroscopy. *Applied Spectroscopy* 49:224-228.
115. Cai Q, *et al.* (2007) Nanometer distance measurements in RNA using site-directed spin labeling. *Biophysical Journal* 93:2110-2117.
116. Enderlein J, Robbins DL, Ambrose WP, Goodwin PM, & Keller RA (1997) Statistics of single-molecule detection. *Journal of Physical Chemistry B* 101:3626-3632.
117. 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.
118. 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.
119. Dix JA, Hom EFY, & Verkman AS (2006) Fluorescence correlation spectroscopy simulations of photophysical phenomena and molecular interactions: A molecular dynamics/Monte Carlo approach. *Journal of Physical Chemistry B* 110:1896-1906.

REFERENCES

120. Dale RE, Eisinger J, & Blumberg WE (1979) The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer. *Biophysical Journal* 26:161-193.
121. van der Meer BW, Cooker G, & Chen SY (1994) *Resonance Energy Transfer: Theory and Data* (VCH Publishers, New York).
122. Maeda H (1980) A new lysozyme assay based on fluorescence polarization or fluorescence intensity utilizing a fluorescent peptidoglycan substrate. *J Biochem* 88:1185-1191.
123. Gopich IV & Szabo A (2012) Theory of the energy transfer efficiency and fluorescence lifetime distribution in single-molecule FRET. *Proc Natl Acad Sci U S A* 109:7747-7752.
124. Englander SW & Mayne L (2014) The nature of protein folding pathways. *Proc Natl Acad Sci U S A* 111:15873-15880.
125. Krishna MM, Hoang L, Lin Y, & Englander SW (2004) Hydrogen exchange methods to study protein folding. *Methods* 34:51-64.
126. Guyett PG, Lisa (2012) Intermediates in Protein Folding. *Comprehensive Biophysics*, ed Egelman EH (Oxford: Academic Press), 1 Ed, pp 43-71.
127. Englander SW, Mayne L, & Krishna MM (2007) Protein folding and misfolding: mechanism and principles. *Q Rev Biophys* 40:287-326.
128. Schuler B & Hofmann H (2013) Single-molecule spectroscopy of protein folding dynamics--expanding scope and timescales. *Curr Opin Struct Biol* 23:36-47.
129. Banerjee PR & Deniz AA (2014) Shedding light on protein folding landscapes by single-molecule fluorescence. *Chem Soc Rev* 43:1172-1188.
130. Nettels D, Hoffmann A, & Schuler B (2008) Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds. *J Phys Chem B* 112:6137-6146.
131. Chen H, Ahsan SS, Santiago-Berrios MB, Abruna HD, & Webb WW (2010) Mechanisms of quenching of Alexa fluorophores by natural amino acids. *J Am Chem Soc* 132:7244-7245.
132. Doose S, Neuweiler H, & Sauer M (2009) Fluorescence quenching by photoinduced electron transfer: a reporter for conformational dynamics of macromolecules. *Chemphyschem* 10:1389-1398.
133. Neubauer H, *et al.* (2007) Orientational and dynamical heterogeneity of rhodamine 6G terminally attached to a DNA helix revealed by NMR and single-molecule fluorescence spectroscopy. *J Am Chem Soc* 129:12746-12755.
134. Yadav R, Sengupta B, & Sen P (2014) Conformational fluctuation dynamics of domain I of human serum albumin in the course of chemically and thermally induced unfolding using fluorescence correlation spectroscopy. *J Phys Chem B* 118:5428-5438.
135. Sherman E & Haran G (2011) Fluorescence correlation spectroscopy of fast chain dynamics within denatured protein L. *Chemphyschem* 12:696-703.
136. Werner JH, Joggerst R, Dyer RB, & Goodwin PM (2006) A two-dimensional view of the folding energy landscape of cytochrome c. *Proc Natl Acad Sci U S A* 103:11130-11135.
137. Chen H, Rhoades E, Butler JS, Loh SN, & Webb WW (2007) Dynamics of equilibrium structural fluctuations of apomyoglobin measured by fluorescence correlation spectroscopy. *Proc Natl Acad Sci U S A* 104:10459-10464.
138. Pabbathi A, Patra S, & Samanta A (2013) Structural transformation of bovine serum albumin induced by dimethyl sulfoxide and probed by fluorescence correlation spectroscopy and additional methods. *Chemphyschem* 14:2441-2449.
139. Becker W (*Advanced Time Correlated Single Photon Counting Techniques* (Springer) p 331.
140. Sindbert S (2012) FRET restrained high-precision structural modeling of biomolecules. Dissertation (Heinrich-Heine Universität, Düsseldorf).
141. Becker NB, Rosa A, & Everaers R (2010) The radial distribution function of worm-like chains. *Eur Phys J E Soft Matter* 32:53-69.
142. Enderlein J, Gregor I, Patra D, Dertinger T, & Kaupp UB (2005) Performance of fluorescence correlation spectroscopy for measuring diffusion and concentration. *Chemphyschem* 6:2324-2336.

REFERENCES

143. Kawahara K & Tanford C (1966) Viscosity and density of aqueous solutions of urea and guanidine hydrochloride. *J Biol Chem* 241:3228-3232.
144. Babu MM, van der Lee R, de Groot NS, & Gsponer J (2011) Intrinsically disordered proteins: regulation and disease. *Curr Opin Struct Biol* 21:432-440.
145. Wright PE & Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* 16:18-29.
146. Brucale M, Schuler B, & Samori B (2014) Single-molecule studies of intrinsically disordered proteins. *Chem Rev* 114:3281-3317.
147. Ferreon AC, Moran CR, Gambin Y, & Deniz AA (2010) Single-molecule fluorescence studies of intrinsically disordered proteins. *Methods Enzymol* 472:179-204.
148. Hofmann H, *et al.* (2012) Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc Natl Acad Sci U S A* 109:16155-16160.
149. Kudryavtsev V, *et al.* (2012) Combining MFD and PIE for accurate single-pair Forster resonance energy transfer measurements. *Chemphyschem* 13:1060-1078.
150. Fersht AR (1997) Nucleation mechanisms in protein folding. *Curr Opin Struc Biol* 7:3-9.
151. Lapidus LJ, Steinbach PJ, Eaton WA, Szabo A, & Hofrichter J (2002) Effects of chain stiffness on the dynamics of loop formation in polypeptides. Appendix: Testing a 1-dimensional diffusion model for peptide dynamics. *Journal of Physical Chemistry B* 106:11628-11640.
152. Lim T-C (2003) Spring Constant Analogy for Estimating Stiffness of a Single Polyethylene Molecule. *Journal of Mathematical Chemistry* 34:151-161.
153. Shortle D (1996) The denatured state (the other half of the folding equation) and its role in protein stability. *Faseb Journal* 10:27-34.
154. McCarney ER, Kohn JE, & Plaxco KW (2005) Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins. *Crit Rev Biochem Mol* 40:181-189.
155. Kohn JE, *et al.* (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proceedings of the National Academy of Sciences of the United States of America* 101:12491-12496.
156. Lapidus LJ (2013) Exploring the top of the protein folding funnel by experiment. *Curr Opin Struc Biol* 23:30-35.
157. Najbar LV, Craik DJ, Wade JD, & McLeish MJ (2000) Identification of initiation sites for T4 lysozyme folding using CD and NMR spectroscopy of peptide fragments. *Biochemistry* 39:5911-5920.
158. Levitus M & Ranjit S (2011) Cyanine dyes in biophysical research: the photophysics of polymethine fluorescent dyes in biomolecular environments. *Q Rev Biophys* 44:123-151.
159. Muschiello A & Michaelis J (2011) Application of the Nano-Positioning System to the Analysis of Fluorescence Resonance Energy Transfer Networks. *Journal of Physical Chemistry B* 115:11927-11937.
160. Zagrovic B, Snow CD, Khaliq S, Shirts MR, & Pande VS (2002) Native-like mean structure in the unfolded ensemble of small proteins. *Journal of Molecular Biology* 323:153-164.
161. Ohnishi S, Lee AL, Edgell MH, & Shortle D (2004) Direct demonstration of structural similarity between native and denatured eglin C. *Biochemistry* 43:4064-4070.
162. Makarov DE (2010) Spatiotemporal correlations in denatured proteins: The dependence of fluorescence resonance energy transfer (FRET)-derived protein reconfiguration times on the location of the FRET probes. *Journal of Chemical Physics* 132.
163. Kalinin S, Sisamakias E, Magennis SW, Felekyan S, & Seidel CAM (2010) On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits. *Journal of Physical Chemistry B* 114:6197-6206.
164. Matsumura M, Becktel WJ, Levitt M, & Matthews BW (1989) Stabilization of phage T4 lysozyme by engineered disulfide bonds. *Proc Natl Acad Sci U S A* 86:6562-6566.

REFERENCES

165. Cellitti SE, *et al.* (2008) In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J Am Chem Soc* 130:9268-9281.
166. Warren JR & Gordon JA (1966) On Refractive Indices of Aqueous Solutions of Urea. *J Phys Chem-Us* 70:297-&.
167. Gasteiger E, *et al.* (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31:3784-3788.
168. Schrödinger L (The PyMOL Molecular Graphics System.).

8. Appendix

8.1. Vector Map and Nucleotide Sequence pET11a-T4L wt*

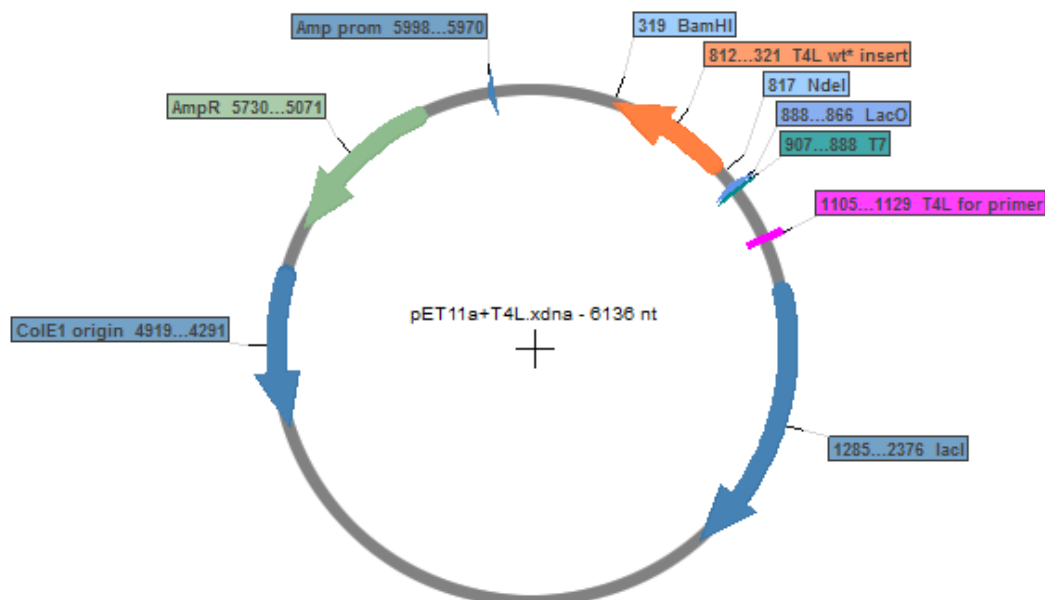


Figure 8.1.1 Vector map of pET11a with the T4L wt* insert (orange) cloned between the NdeI and BamHI restriction sites (light blue).

Table 8.1.1: Nucleotide sequence of T4L, Genbank Identifier: AAD42568.1, underlined triplets have been replaced by ACT (Thr) and GCT (Ala), respectively to create the cysteine free T4L wt* (164)

```

1   ATG AAT ATA TTT GAA ATG TTA CGT ATA GAT GAA CGT CTT AGA CTT AAA ATC TAT
55  AAA GAC ACA GAA GGC TAT TAC ACT ATT GGC ATC GGT CAT TTG CTT ACA AAA AGT
109 CCA TCA CTT AAT GCT GCT AAA TCT GAA TTA GAT AAA GCT ATT GGG CGT AAT TGC
163 AAT GGT GTA ATT ACA AAA GAT GAG GCT GAA AAA CTC TTT AAT CAG GAT GTT GAT
217 GCT GCT GTT CGC GGA ATT CTG AGA AAT GCT AAA TTA AAA CCG GTT TAT GAT TCT
271 CTT GAT GCG GTT CGT CGC TGT GCA TTG ATT AAT ATG GTT TTC CAA ATG GGA GAA
325 ACC GGT GTG GCA GGA TTT ACT AAC TCT TTA CGT ATG CTT CAA CAA AAA CGC TGG
379 GAT GAA GCA GCA GTT AAC TTA GCT AAA AGT ATA TGG TAT AAT CAA ACA CCT AAT
433 CGC GCA AAA CGA GTC ATT ACA ACG TTT AGA ACT GGC ACT TGG GAC GCG TAT AAA
487 AAT CTA TAA
  
```

Table 8.1.2: Amino acid sequence of the cysteine free T4L wt* (164). The molecular weight is 18.64 kDa.

```

1   MNIFEMLRID EGLRLKIYKD TEGYYTIGIG HLLTKSPSLN
41  AAKSELDKAI GRNTNGVITK DEAEKLFNQD VDAAVRGILR
81  NAKLKPVYDS LDAVRRRAALI NMVFQMGETG VAGFTNSLRM
121 LQQKRWDEAA VNLAKSRYWYN QTPNRAKRV I TTFRTGTWDA
161 YKNL
  
```

8.2. Additional Materials & Methods

8.2.1. Materials

The chemicals used in the course of this work were purchased from the following companies: Eurofins MWG Operon, Fermentas, Fluka, GE-Healthcare, Hampton Research, Invitrogen, Merck, Millipore, Qiagen, Riedel de Haen, Roche, Roth, and Sigma-Aldrich. The unnatural amino acid *para*-Acetylphenylalanine (pAcF) was purchased from SynChem OHG (Felsberg, Germany).

8.2.1.1. Enzymes

Table 8.2.1: Enzymes and Standards used during the course of this work

Enzyme	Company
Pfu polymerase (native)	Thermo Scientific
Phusion polymerase	New England Biolabs
T4-DNA ligase	New England Biolabs
Restriction endonucleases: Nde I, BamH I	New England Biolabs
Standards	
GeneRuler 1 kb ladder	Thermo Scientific
PageRuler prestained protein ladder	Thermo Scientific

8.2.1.2. Kits

The following kits were used in the course of this work: QIAprep® Spin Miniprep Kit and QIAprep® Spin HiSpeed Midiprep Kit. Both kits were purchased from Qiagen. Additionally, the NucleoSpin® Gel and PCR Clean-up Kit from Machery-Nagel was used.

8.2.1.3. Vectors

The cysteine-free wildtype wt* and all phage T4 lysozyme mutants were cloned into the pET11a vector (Novagen®, Merck KGaA, Darmstadt, Germany). A detailed vector map is shown in Figure 8.1.1. It carries an Ampicillin resistance as selection marker and the protein production is induced by addition of a final concentration of 1 mM IPTG in the medium. The pEvolv-AcF plasmid needed for incorporation of the unnatural amino acid pAcF carries a chloramphenicol resistance and its production of the tyrRS (tyrosine synthetase) is induced by addition of 0.2 % arabinose.

8.2.1.4. Oligonucleotides

Two different types of oligonucleotides were used during the course of this work, primer for subcloning the whole gene sequence into the backbone and mutation primer for introduction of a one amino acid exchange. All oligonucleotides were purchased by Metabion (Planegg, Germany). For subcloning primer, the restriction sites are in italics. For mutation primer, the mutation site is underlined. All oligonucleotides were solved to 100 µM in ddH₂O.

Table 8.2.2: Primer used for subcloning and mutation in the course of this work.

Primer*	Sequence (5'→3')
T4Lfor	GGAATGGTGCATGCAAGGAGATGG
T4Lend**	GCCGGATCCTTATAGATTTTATACGC
R8Amber for	AATATATTTGAAATGTTATAGATAGATGAACGTCTTAGA
R8Amber rev	TCTAAGACGTTTCATCTATCTATAACATTTCAAATATATT
E11A for	GAAATGTTACGTATAGATGCTGGTCTTAGACTTAAAAATC
E11A rev	GATTTTAAAGTCTAAGACCAGCATCTATACGTAACATTTC
T26E for	GACACAGAAGGCTATTACGAGATTGGCATCGGTCATTG
T26E rev	CAAATGACCGATGCCAATCTCGTAATAGCCTTCTGTGTC
K19Amber for	CTTAGACTTAAATCTATAGGACACAGAAGGCTATTAC
K19Amber rev	GTAATAGCCTTCTGTGTCCTAATAGATTTTAAGTCTAAG
S44Amber for	TCACCTAATGCTGCTAAATAGGAATTAGATAAAGCTATT
S44Amber rev	AATAGCTTTATCTAATTCCTATTTAGCAGCATTAAAGTGA
S44C for	TCACCTAATGCTGCTAAATGTGAATTAGATAAAGCTATT
S44C rev	AATAGCTTTATCTAATTCACATTTAGCAGCATTAAAGTGA
N55Amber for	GCTATTGGGCGTAATACTTAGGGTGAATTACAAAAGAT
N55Amber rev	ATCTTTTGTAATTACACCCTAAGTATTACGCCCAATAGC
K60Amber for	ACTAATGGTGTAATTACATAGGATGAGGCTGAAAACTC
K60Amber rev	GAGTTTTTCAGCCTCATCCTATGTAATTACACCATTAGT
Q69Amber for	GCTGAAAACTCTTTAATTAGGATGTTGATGCTGCTGTT
Q69Amber rev	AACAGCAGCATCAACATCCTAATTAAGAGTTTTTCAGC
Q69C for	GCTGAAAACTCTTTAATTGTGATGTTGATGCTGCTGTT
Q69C rev	AACAGCAGCATCAACATCACAAATTAAGAGTTTTTCAGC
D70Amber for	GAAAACTCTTTAATCAGTAGGTTGATGCTGCTGTTTCGC
D70Amber rev	GCGAACAGCAGCATCAACCTACTGATTAAAGAGTTTTTC
P86C for	AGAAATGCTAAATTAATAATGTGTTTATGATTCTCTGAT
P86C rev	ATCAAGAGAATCATAAACACATTTTAATTTAGCATTCT
R119C for	GGATTTACTAACTCTTTATGTATGCTTCAACAAAAACGC
R119C rev	GCGTTTTTGTTGAAGCATACATAAAGAGTTAGTAAATCC
N132C for	TGGGATGAAGCAGCAGTTTGTGTTAGCTAAAAGTAGATGG
N132C rev	CCATCTACTTTTAGCTAAACAACTGCTGCTTCATCCCA
I150C for	AATCGCGCAAACGAGTCTGTACAACGTTTAGAACTGGC
I150C rev	GCCAGTTCTAAACGTTGTACAGACTCGTTTTGCGCGATT

*The underlined nucleotides mark the mutation side.

**The italic nucleotides mark the restriction enzyme recognition site

8.2.1.5. Plasmids

In the following table all plasmids used in and prepared in this work are listed.

Table 8.2.3: Plasmids used in this work

Plasmid	GOI	Derivation to wt*	Reference
pEvolv			(165)
T4L wt*	T4L	None	Mark Fleissner (UCLA)
T4L R8Amber	T4L	pAcF at aa 8	<i>This work</i>
T4L K19Amber	T4L	pAcF at aa 19	Daniel Rohrbeck (HHU)
T4L S44Amber	T4L	pAcF at aa 44	Hugo Sanabria (HHU)
T4L N55Amber	T4L	pAcF at aa 55	Hugo Sanabria (HHU)
T4L Q69Amber	T4L	pAcF at aa 69	<i>This work</i>
T4L P86C	T4L	Cys at aa 86	<i>This work</i>
T4L R119C	T4L	Cys at aa 119	<i>This work</i>
T4L N132C	T4L	Cys at aa 132	<i>This work</i>
T4L I150C	T4L	Cys at aa 150	<i>This work</i>
R8 Q69C	T4L	pAcF at aa 8, Cys at aa 69	<i>This work</i>
R8 P86C	T4L	pAcF at aa 8, Cys at aa 86	<i>This work</i>
R8 N132C	T4L	pAcF at aa 8, Cys at aa 132	<i>This work</i>
K19 Q69C	T4L	pAcF at aa 19, Cys at aa 69	<i>This work</i>
K19 P86C	T4L	pAcF at aa 19, Cys at aa 86	Daniel Rohrbeck (HHU)

Plasmid	GOI	Derivation to wt*	Reference
K19 R119C	T4L	pAcF at aa 19, Cys at aa 119	Daniel Rohrbeck (HHU)
K19 N132C	T4L	pAcF at aa 19, Cys at aa 132	Daniel Rohrbeck (HHU)
K19 I150C	T4L	pAcF at aa 19, Cys at aa 150	Daniel Rohrbeck (HHU)
S44 Q69C	T4L	pAcF at aa 44, Cys at aa 69	<i>This work</i>
S44 P86C	T4L	pAcF at aa 44, Cys at aa 86	Daniel Rohrbeck (HHU)
S44 R119C	T4L	pAcF at aa 44, Cys at aa 119	Hugo Sanabria (HHU)
S44 N132C	T4L	pAcF at aa 44, Cys at aa 132	Hugo Sanabria (HHU)
S44 I150C	T4L	pAcF at aa 44, Cys at aa 150	Hugo Sanabria (HHU)
S44C I150C E11A	T4L	pAcF at aa 44, Cys at aa 150, Ala at aa 11	<i>This work</i>
S44 I150C T26E	T4L	pAcF at aa 44, Cys at aa 150, Glu at aa 26	Hugo Sanabria (HHU)
N55 Q69C	T4L	pAcF at aa 55, Cys at aa 69	<i>This work</i>
N55 P86C	T4L	pAcF at aa 55, Cys at aa 86	Hugo Sanabria (HHU)
N55 R119C	T4L	pAcF at aa 55, Cys at aa 119	Hugo Sanabria (HHU)
N55 N132C	T4L	pAcF at aa 55, Cys at aa 132	Hugo Sanabria (HHU)
N55 I150C	T4L	pAcF at aa 55, Cys at aa 150	Hugo Sanabria (HHU)
K60 P86C	T4L	pAcF at aa 60, Cys at aa 86	<i>This work</i>
K60 R119C	T4L	pAcF at aa 60, Cys at aa 119	<i>This work</i>
K60 N132C	T4L	pAcF at aa 60, Cys at aa 132	<i>This work</i>
K60 I150C	T4L	pAcF at aa 60, Cys at aa 150	<i>This work</i>
Q69 P86C	T4L	pAcF at aa 69, Cys at aa 86	<i>This work</i>
Q69 R119C	T4L	pAcF at aa 69, Cys at aa 119	<i>This work</i>
Q69 N132C	T4L	pAcF at aa 69, Cys at aa 132	<i>This work</i>
Q69 I150C	T4L	pAcF at aa 69, Cys at aa 150	<i>This work</i>
D70 R119C	T4L	pAcF at aa 70, Cys at aa 119	<i>This work</i>
D70 N132C	T4L	pAcF at aa 70, Cys at aa 132	<i>This work</i>

8.2.1.6. Bacterial Strains

In the following table the bacterial strains used in the course of this work are listed. All strains were used as heat shock competent cells which were stored in 240 µL aliquots at -80 °C until use.

Table 8.2.4: Bacterial strain used in the course of this work

Strain	Genotype	Company
DH5α	F- Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK-, mK+) phoA supE44 λ- thi-1 gyrA96 relA1	Invitrogen
XL2blue	endA1 gyrA96(nal ^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB ⁺ lacI ^q Δ(lacZ)M15 Amy Cm ^R] hsdR17(rK- mK ⁺)	Stratagene
BL21	F - ompT hsdSB(rB - mB -) dcm + dam + Tet R galλ (DE3) endA Hte [argU, ileY, leuW, Cm R]	Stratagene

8.2.1.7. Media and Buffers

In the following table the media and buffers used in the course of this work are listed. The pH of buffers was adjusted with HCl and NaOH, respectively, if not stated otherwise.

Table 8.2.5 Media and buffers used in the course of this work

Media	Composition
Luria Bertani (LB-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
Mg ²⁺ -Mix	500 mM MgCl ₂ , 500 mM MgSO ₄
TMF-buffer	100 mM CaCl ₂ , 50 mM RbCl ₂ , 40 mM MnCl ₂
TMFG-buffer	TMF-buffer, 20 % (v/v) glycerol
Buffers	
10x PBS	1.38 M NaCl, 27 mM KCl, 81 mM NaH ₂ PO ₄ , 15 mM KH ₂ PO ₄ pH 7.5
CD buffer 0 M	50 mM sodium phosphate pH 7.5

CD buffer 8 M	50 mM sodium phosphate pH 7.5, 8 M urea
DNA loading dye (5x)	10 mM Tris-HCl, pH 7.5, 0.2 % (w/v) orange G, 1 mM EDTA, 50 % glycerol
CatIEX buffer A	25 mM HEPES pH 7.5, 0.1 mM EDTA, 5 mM DTT
CatIEX buffer B	25 mM HEPES pH 7.5, 0.1 mM EDTA, 1 M NaCl, 5 mM DTT
Cys buffer A	20 mM Tris-HCl pH 8.5
Cys buffer B	20 mM Tris-HCl pH 8.5, 1 M NaCl
Keto-labeling buffer	50 mM Sodium acetate, 150 mM NaCl, adjusted with acetic acid to pH 4
PBS	50 mM sodium phosphate, 150 mM NaCl pH 7.5
PBS 8 M	50 mM sodium phosphate, 150 mM NaCl pH 7.5, 8 M urea
Denaturation buffer	50 mM sodium phosphate, 150 mM NaCl, 8 M urea, pH 7.5
SDS-PAGE running buffer	25 mM Tris-HCl, 192 mM glycine, 0.1 % (w/v) SDS
SDS-PAGE sample buffer (5x)	10 % SDS, 30 % sucrose, 0.1 % bromophenol blue, 0.25 M Tris pH 6.8, 0.5 M β -mercaptoethanol
SDS-PAGE separating gel (12%)	4 ml acrylamid (30 %, 29:1), 3.75 ml 1 M Tris pH 8.8, 2.09 ml dH ₂ O, 50 μ l 20 % SDS, 100 μ l 10 % APS, 10 μ l TEMED
SDS-PAGE separating gel (15%)	5 ml acrylamid (30 %, 29:1), 3.75 ml 1 M Tris pH 8.8, 1.09 ml dH ₂ O, 50 μ l 20 % SDS, 100 μ l 10 % APS, 10 μ l TEMED
SDS-PAGE stacking gel (5%)	830 μ l acrylamid (30 %, 29:1), 626 μ l 1 M Tris pH 6.8, 3.459 ml dH ₂ O, 25 μ l 20 % SDS, 50 μ l 10 % APS, 10 μ l TEMED
TAE buffer	40 mM Tris-HCl pH 8.2, 20 mM sodium acetate, 1 mM EDTA

All buffer containing urea were examined for their exact concentration using a refractometer and by comparing the index of refraction of buffer with (η_{8M}) and without urea (η_{0M}). The exact urea concentration could then be calculated using the following formula (166) (Eq. 8.2.1):

$$c(Urea[M]) = 117.66 \cdot (\eta_{8M} - \eta_{0M}) + 29.753 \cdot (\eta_{8M} - \eta_{0M})^2 + 185.56 \cdot (\eta_{8M} - \eta_{0M})^3 \quad 8.2.1$$

8.2.2. Molecular biological methods

8.2.2.1. Preparation of chemically competent cells

For the preparation of chemically competent cells 5 μ L of an aliquot of the respective strain was added to 5 mL of LB-medium and incubated while shaking over night at 37 °C. Next day, 100 mL of LB were inoculated with 1/100 of its volume with the overnight culture and 1/50 of its volume with Mg²⁺-Mix. The cells were grown at 37 °C while shaking until an OD_{600} of ~ 0.6 was reached, and then harvested in 2 mL aliquots at 8'000 * g for 3 min. The supernatant was removed, the cell pellet subsequently resolved in 1 mL of TMF-buffer and incubated for 30 min on ice. The cells were harvested a second time at 8'000 * g for 3 min, resuspended in 250 μ L of TMFG- buffer and stored immediately at -80 °C until use.

8.2.2.2. Transformation of plasmids in *E.coli*

The transformation of plasmid DNA was done *via* heat shock transformation. The cells were thawed on ice for 30 min and 30 ng of plasmid was added. After an incubation of 15 min on ice, the cells were placed for 45 sec at 42 °C, followed by a further incubation of 2 min on ice. Finally, the cells were spread on LB-agar plates containing the respective antibiotic.

For the co-transformation of pEvolv, which carries a Chloramphenicol resistance, and one of the Amber-mutant carrying pET11a plasmids, 30 ng of the first and 15 ng of the latter plasmid were used simultaneously. Here, the cells were spun down after the heat shock for 1 min at $13'000 \times g$, resuspended in 250 μ L of LB-medium and grown while shaking for 30 min at 37 °C prior to plating on LB-agar plates containing both antibiotics.

8.2.2.3. Preparation of Plasmid DNA

The preparation of plasmid DNA was done following the instructions of the QIAprep Spin Miniprep Kit using 5 mL of an overnight culture of *E. coli* DH5 α or XL2blue cells. For the preparation of larger amounts of plasmid DNA, this was isolated out of 100 mL of an overnight culture using the QIAprep Hispeed Midi Kit.

8.2.2.4. Agarose gel electrophoresis

For the analysis of DNA fragments and to check for successful amplification of plasmids during PCR, agarose gel electrophoresis was used. Agarose was dissolved to 1 % in 1x TAE buffer and poured into a tray containing 2 μ L of a 1 % (w/v) ethidium bromide solution to stain the DNA. The hardened gel was put into the electrophoresis chamber and flooded with 1x TAE-buffer. The samples containing 2 μ L 5x loading dye were pipetted into the pockets of the gel. As a size marker 5 μ L of the GeneRuler 1kb ladder were used. The electrophoresis was then carried out with a constant voltage of 100 V at 30 min. Subsequently, pictures of the gels were taken under UV-light.

8.2.2.5. DNA purification

All DNA purifications were done by following the protocol in the NucleoSpin® Gel and PCR Clean-up Kit from Machery-Nagel.

8.2.2.6. Restriction enzyme digestion

A restriction digest was done for the original plasmids and amplified inserts to generate sticky ends for subsequent ligation. Per 1 μ g of DNA 10 U of both restriction enzymes Nde I and BamH I was used. The respective buffer and BSA at a final concentration of 1 μ g/ml were added following the recommendation of the manufacturer's website.

All reagents were mixed and the digests were incubated for one hour at 37 °C. The digests were subsequently separated using agarose gel electrophoresis and purified using the NucleoSpin® Gel and PCR Clean-up Kit.

8.2.2.7. *Site-directed mutagenesis*

For insertion of a specific mutation in a protein, the respective base pair triplet in the gene of interest has to be modified. Two variations of the polymerase chain reaction (PCR) can be used here: Site-directed mutagenesis and Overlap-extension mutagenesis. The first method is the quickest one. It requires only one temperature-gradient PCR and one complementary pair of primer carrying the modified base triplet.

Here, a temperature-gradient during the annealing phase is needed because the complementary primer pair tends to self-annealing and one needs to find that temperature where this is minimal and annealing to template strand is maximal. For T4L, this temperature usually was found to be ~ 10 °C below the primer melting temperature.

For each mutation five reaction mixtures for scanning five different annealing temperatures were prepared as follows:

- 5 µl 10x Pfu buffer +MgSO₄
- 1µl dNTP mix (10 mM each)
- 10 ng DNA template
- 0.5 µl forward primer
- 0.5 µl reverse primer
- *ad* 49 µl with ddH₂O
- Add 1 µl of polymerase at last step

The PCR program used was the following:

1. 95 °C 3 min
2. 95 °C 30 s
3. Tgrad 30 s
4. 72 °C 12 min
5. Repeat step 2-4 25x
6. 72 °C 15 min (final extension)
7. 4 °C forever

After completion of the PCR, 0.5 µL of DpnI was added to digest the template DNA for 1 hr at 37 °C. Subsequently, 10 µL of each reaction was mixed with 5x DNA loading dye and loaded onto a 1 % agarose gel.

2 µL of each positive reaction mixture was transformed into *E.coli* DH5α cells.

8.2.2.8. *Overlap- extension mutagenesis*

The second method to introduce a specific mutation in the gene of interest (goi) is the overlap-extension mutagenesis. Here, two subsequent PCR's are needed. The first one produces the downstream and upstream regions of the mutation site in two fragments, the

second PCR hybridizes both fragments to the full-length goi. This method is more time-consuming. However, it avoids the use of self-complementary primer and thus potential problems with self-annealing.

The parameters for the first PCR are as follows:

Table 8.2.6 Parameters for the fragment PCR

Fragment 1	Fragment 2	Program
0.5 μ L template	0.5 μ L template	1. 98 °C 2 min
1 μ L dNTPs (10 mM each)	1 μ L dNTPs (10 mM each)	2. 98 °C 30 sec
1 μ L primer forward	1 μ L primer reverse	3. 67 °C 30 sec*
1 μ L mutagenesis primer reverse	1 μ L mutagenesis primer forward	4. 72 °C 60 sec
0.5 μ L HiFi polymerase	0.5 μ L HiFi polymerase	5. Repeat 2-4 35x
10 μ L 5x buffer	10 μ L 5x buffer	6. 72 °C 10 min
36 μ L ddH ₂ O	36 μ L ddH ₂ O	

*This temperature is equal to the lowest primer melting temperature

After completion of PCR, the reaction mixes were loaded onto a 1 % agarose gel; amplified bands of the right size were cut and purified.

Subsequently, the second PCR was done to hybridize both fragments to the full goi. The used parameters are listed in Table 8.2.7.

During the program break, 1 μ L of each forward and reverse primer have to be added to both tubes. After completion of PCR, both reaction mixtures were applied onto a 1 % agarose gel and bands of the correct size were purified. The goi was digested with the respective restriction enzymes and ligated into the empty backbone.

Table 8.2.7 Parameters for Hybrid-PCR

Tube 1	Tube 2	Program
4 μ L fragment 1	4 μ L fragment 1	1. 98 °C 2 min
1 μ L fragment 2	1 μ L fragment 2	2. 98 °C 30 sec
1 μ L dNTP's	1 μ L dNTP's	3. 67 °C 30 sec
0.5 μ L HiFi polymerase	0.5 μ L HiFi polymerase	4. 72 °C 90 sec
10 μ L 5x buffer	10 μ L 5x buffer	5. repeat 2-4 20x
33.5 μ L ddH ₂ O	33.5 μ L ddH ₂ O	6. 72 °C 10 min
		7. Break
		8. 98 °C 30 sec
		9. 67 °C 30 sec
		10. 72 °C 90 sec
		11. Repeat 8-10 35x
		12. 72 °C 10 min

8.2.2.9. DNA-Ligation

In the DNA-ligation the digested inserts were placed at the complementary ends of the digested vectors to obtain the new, closed constructs with the help of T4-DNA-ligase.

Each reaction mixture consisted of 100 ng of plasmid, 1 μ L of T4-DNA-ligase, and 10x T4-ligase buffer as well as ddH₂O and the respective insert. The amount of insert needed was calculated using the following formula (Eq. 8.2.2).

$$m_I = m_V \cdot \frac{l_I}{l_V} \cdot R \quad 8.2.2$$

Here m_I is the amount of insert, m_V the amount of target vector (100 ng), l_I and l_V are the length of insert and vector in base pairs, respectively. R gives the molar ratio vector: insert. In these experiments the ratio was set to 1:5. The missing volume was filled up with ddH₂O. The ligation mixtures were pipetted together, centrifuged down and incubated at room temperature for one hour.

8.2.2.10. DNA- Sequencing

All prepared constructs were sent for sequencing to GATC Biotech (Konstanz, Germany) in vials à 20 µL.

8.2.2.11. Photometric Determination of DNA and Protein Concentration

The concentration of the plasmid DNA and unlabeled protein was determined photometrically using the spectrometer PEQLab NanoDrop ND-1000 at a wavelength of 260 nm and 280 nm. For each measurement 2µl of sample was used while the respective buffer was used as reference. For proteins, the extinction coefficient for the determination of the concentration was estimated by the ProtParam tool (<http://web.expasy.org/protparam/>, (167)).

8.2.3. Protein production & purification

8.2.3.1. Protein production

In the evening, a colony from a plate of freshly transformed cells was picked and incubated overnight in 100 mL of LB medium containing the respective antibiotics and grown overnight at 37 °C at 150 rpm. Next morning, the main culture was inoculated with 1/50 of its volume of preculture and the respective antibiotics. For the production of protein containing the unnatural amino acid pAcF, 0.4 g/L of this chemical was added additionally. The cells were grown while shaking at 37 °C until an OD_{600} of ~ 0.6 was reached. Then, the protein production was induced by addition of a final concentration of IPTG of 1 mM and, for pAcF containing protein, 4 g/L of arabinose. Prior to induction and then every two hours, a 1 mL sample was taken to monitor the protein production.

To check, whether the protein had been produced, the taken samples were applied onto SDS-PAGE. For a better comparison, samples were normalized to have the same OD_{600} according to the following formula (Eq. 8.2.3):

$$\text{Resuspension volume} = \frac{\text{measured } OD_{600} \cdot 1000}{3} \quad 8.2.3$$

The samples were spun down, the supernatant was removed and the pellet resuspended in the above calculated value. 12 µL of the normalized sample were applied onto the gel.

8.2.3.2. *Cell harvest and lysis*

After 6 hrs of induction, cells were harvested by centrifugation for 15 min at 10'000 * g in a Sorvall EvolutionRC (ThermoScientific) equipped with a SLA3000 rotor at 4 °C and stored at -20°C until purification.

Cells were thawed on ice in 15 mL CatIEX buffer A per 1 OD_{600} per liter of original culture. Subsequently, cell lysis was performed on ice via sonification for three cycles of 30 s on – 60 s off with a Sonoplus HD 2200 (Bandelin) equipped with a MS72 sonotrode (Bandelin) at 40 % power. Insoluble cell fragments were separated from the soluble part by centrifugation for 45 min at 30'000 * g at 4 °C in a Sorvall EvolutionRC (ThermoScientific) equipped with a SS34 rotor.

8.2.3.3. *Ion exchange chromatography (IEX)*

In an ion exchange chromatography (IEX) the proteins are separated by their different interaction with the charged column material. In a cationic IEX as used here, the column material is negatively charged and the protein of interest is in a buffer with a pH value below its pI (positively charged) and a low salt concentration. The protein was loaded on a HiTrap SP Sepharose FF 5/50 column (GE Healthcare), which was preequilibrated with CatIEX buffer A. After the protein was completely loaded, the column was washed with 5 CV of buffer A to remove unbound protein. The elution was started by applying an increasing salt gradient from 0 % to 100 % CatIEX buffer B with a length of 10 CV through mixing of buffer A and buffer B. Fractions were collected in 5 mL size and those containing protein were analyzed on an SDS gel. During the whole purification the flow rate was kept constant at 1.5 mL/min.

8.2.4. *Protein analytical methods*

8.2.4.1. *SDS polyacrylamide gelelectrophoresis (SDS-PAGE)*

SDS PAGE was used to evaluate the purity and yield during the protein production. The gels were composed of two parts, the upper stacking gel (5 % acrylamide) and the lower resolving gel (12 % or 15 % acrylamide). Typically, 12 µL of protein solution together with 3 µL of SDS loading buffer were applied on the gel. Viscous samples, e.g. samples taken from the

growth culture, were heated for 5 min at 95 °C prior to application to the gel. The gel was run at a constant voltage of 200 V until the dye front reached the end of the gel (usually 45 min). After the electrophoresis, the gel was washed with water to remove SDS and then stained with SimplyBlue™ SafeStain dye (Invitrogen) for 1 hr.

8.2.4.2. Protein concentration

All proteins were concentrated using Amicon concentrators (Millipore) with a molecular weight cutoff (MWCO) of maximal half of the size of the target protein. The proteins were concentrated at 5'000 * g (50 mL concentrators) and 4'000 * g at 4 °C, respectively, until the desired protein concentration was reached.

For T4L, the pooled protein fractions from IEX were first filtered through a 50 kDa concentrator to remove impurities of high molecular weight and then concentrated using a 10 kDa filter. For the 50 kDa concentrator it could be shown that impurities larger than 30 kDa did not pass through its membrane.

8.2.4.3. Buffer exchange

Small scale buffer exchange was done using Amicon concentrators. Firstly, the protein was concentrated, then diluted 100fold with the desired buffer and concentrated again. This procedure was repeated three times.

8.2.4.4. Site-specific labeling using orthogonal chemistry

T4L variants containing a cysteine and the unnatural amino acid pAcF were labeled using orthogonal chemistry. The acceptor dye Alexa647-maleimide was coupled to the thiol-group of the cysteine, whereas the keto-group of pAcF was coupled to the donor dye Alexa488-hydroxylamine (94) (Figure 8.2.1).

A fivefold molar excess of Alexa647-maleimide was added to the selected variant and the reaction mixture was incubated in the dark for 2 h at room temperature in PBS. Subsequently, the excess of dye was removed and the buffer exchanged to Keto-labeling buffer using a desalting column (Nap5 or Nap10, GE Healthcare) according to the manufacturer's instructions. Immediately after elution, a fivefold molar excess of Alexa488-hydroxylamine was added to the protein solution. After incubation for 72 h in the dark at 37 °C, the reaction mixture was centrifuged for 30 min at 13'000 * g to remove any precipitated protein. The cleared supernatant was transferred to a new 1.5 mL eppendorf tube and the pellet was solved slowly in 100 – 300 µL of PBS containing additionally 8 M urea. For efficient refolding of T4L, the cleared supernatant solution (containing no urea) was added in several small

fractions to the resolved pellet, each addition was followed by severe mixing, thus, decreasing the urea concentration slowly. Finally, non-reacted dye was removed and the buffer exchanged to PBS using a desalting column (Nap10 or PD10, GE Healthcare) according to the manufacturer's instructions.

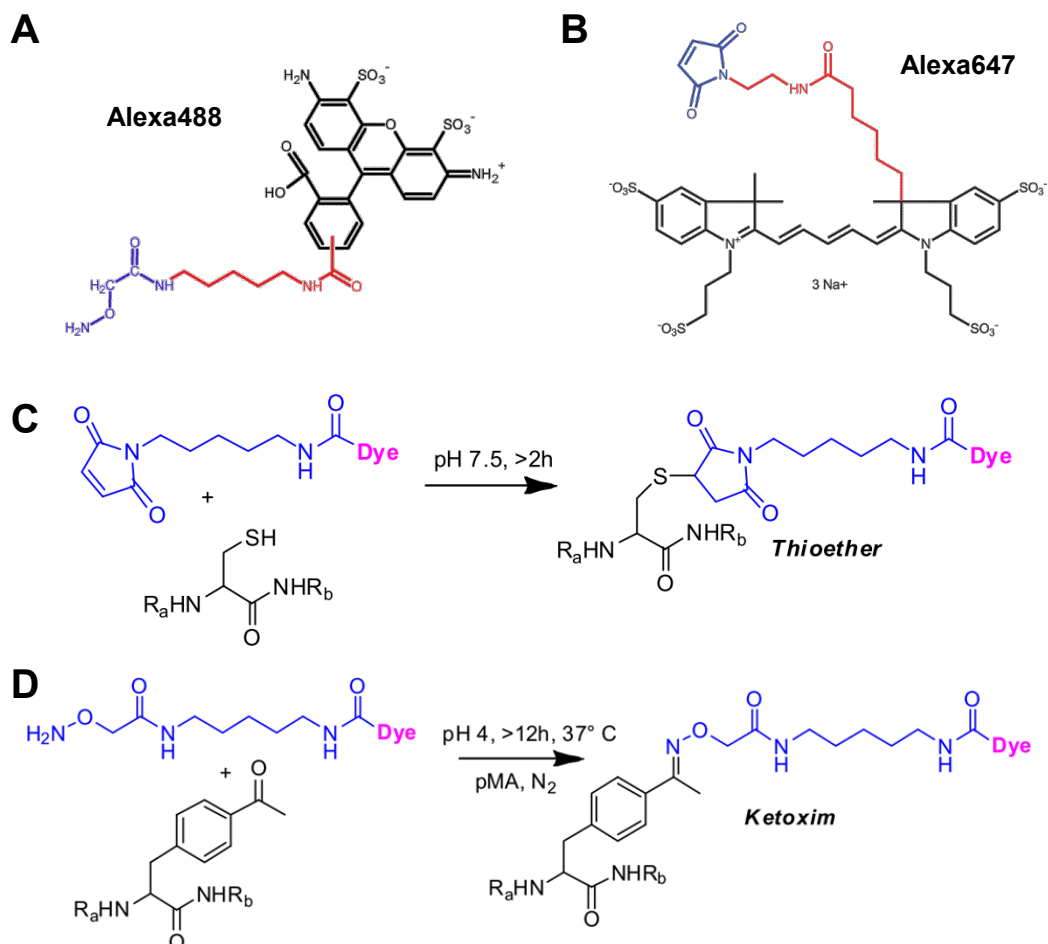


Figure 8.2.1 Structures of the fluorophores used and labeling reactions. (A) Alexa488-hydroxylamine. The fluorophore is shown in black, the linker in red and the reactive group in blue. (B) Alexa647-maleimide. Color code is the same as for (A). (C) Labeling of cysteines with the acceptor dye Alexa647-maleimide occurred via the thiol-group at an ambient pH value at room temperature. (D) pAcF was labeled with the donor dye Alexa488-hydroxylamine for 72 hrs at 37 °C and a pH of 4.

The eluted protein was concentrated using Amicon concentrators (Millipore) with a MWCO of 10 kDa up to a volume of 250 μ L. To ensure complete removal of free dye, the concentrated protein was diluted 15fold with PBS and concentrated again. This process was repeated twice, and then the concentration and labeling degree was determined. The labeled protein was stored until use in aliquots à 20 μ L at -20 °C.

For donor-only labeled samples, the buffer of the variant was exchanged to keto-labeling buffer directly and then proceeded as described above.

8.2.4.5. *Site-selective labeling of double cysteine variants*

T4L variants containing two cysteines were labeled using the two Maleimide-dyes Alexa488 and Alexa647.

The T4L variants solved in PBS were incubated with 0.8 molar equivalent of Alexa488-maleimide for 2 h in the dark at 4 °C. The total reaction volume was around 200 µL. In the meantime a ReSource S column (6 mL, GE Healthcare) was equilibrated with Cys buffer A according to standard procedures. Before loading the reaction mixture on the column, a 10fold volume of Cys buffer A was added to increase the pH to 8.5 and to reduce the salt concentration to such an extent that the protein could bind to the cation exchanger. After loading, the column was washed with 5 CV of Cys buffer A (3 mL/min, 3 mL fractions). For specific elution, a gradient from 0 to 40 % Cys buffer B of a length of 15 CV was applied (1.5 mL/min, 1.5 mL fractions). Molecules, where both cysteines had reacted with a dye molecule, were eluted first, followed by single labeled protein, and finally the unreacted protein was eluted from the column.

From all fractions showing high UV-absorption the concentration and degree of labeling was determined. Fractions containing only single-labeled protein were pooled, one half of it, the respective donor-only sample, was concentrated as described above, and stored at -20 °C until use. To the second half 1/10 of the volume of 1 M Tris-HCl pH 7.4 was added to reduce the pH to an optimal value for maleimide coupling, and a 5 fold molar excess of Alexa647-maleimide was added. After incubation for 2 h in the dark at 4 °C, the excess dye was removed using a desalting column (Nap10 or PD10, GE Healthcare) and the protein was further concentrated, the concentration and labeling degree determined, and stored at -20 °C until use.

8.2.4.6. *Determination of protein concentration and labeling degree*

The concentration of labeled protein c_P and its degree of labeling was determined via absorption spectrometry. The absorption spectrum of the samples was measured in a range from 250 to 800 nm using a Cary 400 absorption spectrometer, a two beam reference machine, with having a cuvette containing the respective buffer in the reference beam.

The data was exported, corrected for background and the concentration of donor-only labeled sample was calculated as follows (Eq. 8.2.4a):

$$c_P = (Abs_{280} - (Abs_{496} \cdot cf_{D,280})) \cdot \frac{dF}{\epsilon_P \cdot d} \quad 8.2.4a$$

$$\text{Labeling degree}_{A488} = \frac{c_{A488}}{c_P} \quad 8.2.4b$$

If a sample yielded for a wavelength an absorption higher than 1, the sample was diluted. To obtain the original protein concentration, thus, the dilution factor dF had to be taken into account. d is the pathlength of the used cuvette, usually 1 cm, and ε_P is the extinction coefficient of the protein as determined with ProtParam (167).

For double-labeled samples, not only the absorption of donor and acceptor dye at 280 nm has to be taken into account, but also the absorption of the acceptor dye in the donor absorption region (Eq. 8.2.5a).

$$c_P = (Abs_{280} - [((Abs_{496} - Abs_{650} \cdot cf_{A,496}) \cdot cf_{D,280}) + Abs_{650} \cdot cf_{A,280}]) \cdot \frac{dF}{\varepsilon_P \cdot d} \quad 8.2.5a$$

$$\text{Labeling degree}_{A488} = \frac{c_{A488}}{c_P} \quad 8.2.5b$$

$$\text{Labeling degree}_{A647} = \frac{c_{A647}}{c_P} \quad 8.2.5c$$

The correction factors for Alexa488 $cf_{D,280}$ and Alexa647 $cf_{A,280}$ and $cf_{A,496}$ are summarized in Table 8.2.8:

Table 8.2.8 Correction factor and extinction coefficients for Alexa488 and Alexa647

Correction factors		Extinction coefficient [M ⁻¹ ·cm ⁻¹]	
$cf_{D,280}$	0.11	ε_{496}	71'000
$cf_{A,280}$	0.03	ε_{650}	239'000
$cf_{A,496}$	0.00		

8.2.5. Steady-state fluorescence measurements

8.2.5.1. Determination of R_0

The Förster Radius R_0 needed for distance determination between a Förster Resonance Energy dye pair depends upon several factors (Eq. 8.2.6): the spectral overlap integral between the donor emission and acceptor excitation J , the orientation factor κ^2 , the fluorescence quantum yield of the donor ϕ_D and the index of refraction of the surrounding buffer η .

$$R_0 = 0.2108 \cdot (\kappa^2 \cdot J \cdot \phi_D \cdot \eta^{-4})^{1/6} \quad 8.2.6$$

As ϕ_D varies from one donor position to another donor position the reduced Förster radius R_{0r} can be defined (Eq. 8.2.7):

$$R_{0r} = \frac{R_0}{\phi_D^{1/6}} = 0.2108 \cdot (\kappa^2 \cdot J \cdot \eta^{-4})^{1/6} \quad 8.2.7$$

The index of refraction η is measured in a refractometer. Steady-state fluorescence excitation and emission spectra of donor and acceptor only labeled variants were measured on a Fluorolog-4 (Horiba) spectrofluorimeter to determine the overlap integral J . The optical density of the samples was adjusted to be 0.05 or lower to avoid non-linear effects due to the inner filter effect by dilution in the respective buffer. The spectra were recorded at wavelengths summarized in Table 8.2.9 with an integration time of 0.5 s with slit width adjusted such that the count rate was between 10^5 to 10^6 cps, usually one to five nm.

Table 8.2.9 Wavelength at which excitation and emission spectra were recorded

	Excitation	Emission
Alexa488 (donor)	400-510 nm	520 nm
	470 nm	480- 800 nm
Alexa647 (acceptor)	400-660 nm	670 nm
	610 nm	620-800 nm

After subtraction of background, the overlap integral was calculated as follows 8.2.8:

$$J = \int_{480nm}^{660nm} \frac{F_{D,em}(\lambda)}{\sum F_{D,em}(\lambda)} \left(\frac{F_{A,ex}(\lambda)}{F_{A,ex}(\lambda_{max})} \right) \epsilon_A(\lambda_{max}) \lambda^4 d\lambda \quad 8.2.8$$

with $F_{D,em}(\lambda)$ being the emission signal of the donor at the respective wavelength, $F_{A,ex}(\lambda)$ the signal at the respective excitation wavelength, $F_{A,ex}(\lambda_{max})$ the signal at the excitation maximum (670 nm) and $\epsilon_A(\lambda_{max})$ the extinction coefficient at the excitation maximum ($239'000 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

8.2.5.2. Titration with urea

The ensemble-averaged fluorescence as a function of urea concentration was measured on a Fluorolog-3 (Horiba) spectrofluorimeter equipped with a home-build temperature controlled sample holder at 25 °C. For donor-only labeled samples, excitation was set to 495 nm and emission was measured at 520 nm. For double labeled samples, fluorescence at three different setting was recorded: donor emission (excitation at 495 nm, emission at 520), direct acceptor emission (excitation at 650, emission at 670) and acceptor-sensitized emission (excitation at 495, emission at 670). Each measurement was repeated until the standard error was < 0.1 %, integration time per data point was 1 s. To avoid inner filter and anisotropy effects, samples were diluted in PBS until their absorption was smaller than 0.05 and measurements were conducted under magic angle conditions (emission polarizer 0°, excitation polarizer 54.7°).

Samples were titrated from 0 M urea to 7.5 M urea, after each addition of urea samples were mixed carefully and after a delay of ~ 1 min measured to ensure complete mixing. Slit sizes were adjusted this way that the initial countrate was $\sim 10^6$ counts per seconds. If the signal decreased too much while dilution, slits were opened more and the respective urea concentration was measured before and after changes in slit size for the generation of a correction factor. From the titration of a donor- only labeled sample and the direct acceptor excitation, the correction for a urea-dependent quantum yield of the dyes was derived. Final acceptor-sensitized fluorescence $F(A_D)$ was calculated according to equation 8.2.9:

$$F(A_D) = F_{AD} \cdot dF \cdot \frac{S_{old}}{S_{new}} \cdot \frac{F_{0M,acc}}{F_{xM,acc}} \cdot \frac{F_{0M,D0}}{F_{xM,D0}} \quad 8.2.9$$

F_{AD} is the measured signal, dF the dilution factor and S_{old}/S_{new} is the slit correction factor (signal before slit change/ signal after slit change). The last two terms correct for the change in quantum yield of the acceptor $F_{0M,acc}/F_{xM,acc}$ and donor dye $F_{0M,D0}/F_{xM,D0}$, respectively, relative to their values in the absence of urea.

8.2.5.3. Determination of Steady-State Anisotropy

To be able to fit time-resolved anisotropy data, the steady-state or residual anisotropy r_∞ is needed. It can be calculated from the steady-state fluorescence signal of a (single) labeled probe, which was measured at four different polarization angles using a spectrofluorimeter (Fluoromax4, Horiba). The fluorescence signal is measured at both vertical (V) and horizontal (H) position of the excitation and emission polarizer with an integration time of 0.5 s and slit width adjusted such that the count rate was between 10^5 to 10^6 cps, usually between one to five nm.

The residual anisotropy r_∞ can then be calculated according to equation 8.2.10:

$$r_\infty = \frac{F_{VV} - G * F_{VH}}{F_{VV} + 2 * G * F_{VH}} \quad 8.2.10$$

with the signal subscripts F_{XX} having the following nomenclature: $F_{(position \text{ of excitation polarizer})(position \text{ of emission polarizer})}$. G is a device-dependent factor (compare eq. 2.1.7).

8.2.6. Bioinformatics & Software

The extinction coefficients as well as other theoretical data about the produced proteins were predicted by the Protparam tool. Figures of the proteins were generated with PyMOL (168).

To be able to compare the obtained distances R_{DA} of the measurements with the X-ray structures from T4L, the dyes were added *in silico* to the PDB structures and the to be

expected FRET determinable distances were computed by modeling the dyes as spheres with flexible, cylindrical linkers (46, 53). In the first modeling step the accessible space or volume (“AV”) of the dyes is explored and in the second step a mean distance of all possible, explored dye position is computed. The used modeling parameters are dye dependent (Table 8.2.10).

Table 8.2.10 AV-modeling parameters

Parameter	Alexa488-hydroxylamine	Alexa647-maleimide
linker length	20 Å	22 Å
linker width	4.5 Å	4.5 Å
dye radius	3.5 Å	3.5 Å

The maximum end-to-end length of the unfolded polypeptide chain of T4L was calculated assuming a length of 3.6 Å per residue. Additionally, the length of the linkers used to attach the dyes Alexa488-hydroxylamine (donor) and Alexa647-maleimide (acceptor) needed to be taken into account. The residue and dye length (from attachment point of the protein to the center of mass of the dye) were determined using ChemDraw and yielded 20 Å for A488 and 22 Å for A647.

Data analysis of eTCSPC, MFD, fullFCS and fFCS data was performed using home-written Matlab-scripts, LabView (Kristine, Paris, Margarita, FitMachine, GMCR, Tatiana) and Python (chi2surf) programs. Accessible volume simulations were performed with FPS or its successor Olga.

8.3. Exemplary results for Protein Production, Purification & Labeling

8.3.1. Site-directed mutagenesis

The selected sites for labeling were mutated to either Amber- (TAG) or cysteine codon (TGT) in the T4L gene cloned into the pET11a vector by using two different site-directed mutagenesis methods: One-step PCR and Overlap-Extension PCR. The advantage of the prior one is its simplicity as it only requires one PCR step in which the whole plasmid with the newly mutated site is amplified (Figure 8.3.1A). However, here it is essential to select the right annealing temperature such that the used primer anneal to the templates and don't form primer dimers (lower band in Figure 8.3.1A).

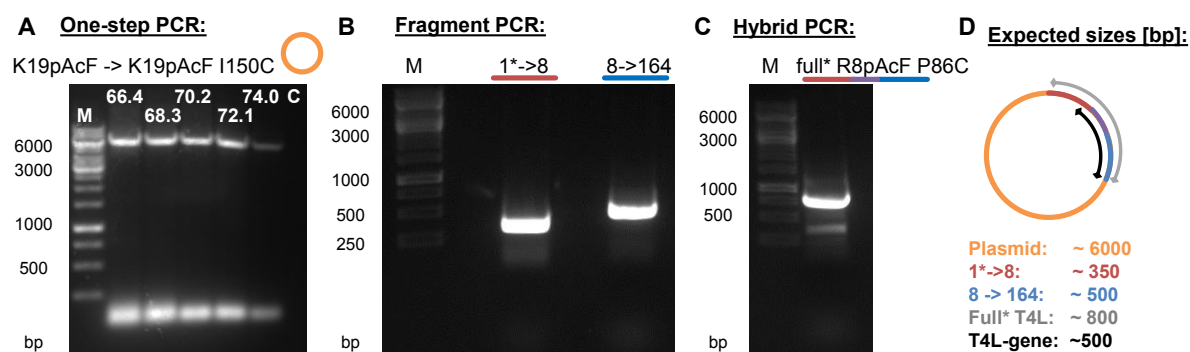


Figure 8.3.1 Agarose-gels showing PCR-results. (A) One step PCR used to generate the variant K19pAcF I150C from the K19pAcF template. (B, C) Overlap-Extension PCR used to generate the variant R8pAcF P86C from the P86C template. (B) The fragment PCR produces the gene fragments in two reactions until and after the mutation site. Note: To ensure a sufficient length of the first fragment, the primer sequence of *T4Lfor* is located in the ampicillin resistance gene. (C) Both produced fragments are fused together in the hybrid PCR. (D) Expected length of PCR products in bp.

The Overlap-Extension PCR instead avoids the use of complementary primers in a single reaction tube but needs two subsequent PCR's (Figure 8.3.1B-C) where first the two gene parts before and after the mutation site are amplified and then those two fragments are hybridized and amplified. After digesting the overhanging ends with the restriction nucleases, the mutated T4L gene is ligated into the empty pET11a vector.

The correct sequence of the mutated T4L was confirmed for all generated plasmids by sequencing.

8.3.2. Protein production & purification

The obtained T4L variants were transformed together with the tRNA-synthetase gene carrying vector pEvolv into *E.coli* BL21 for protein production. After induction of protein production, the generation of T4L was monitored (Figure 8.3.2A). After cell lysis and centrifugation the soluble fraction of the harvested *E. coli* pellet was subjected to a cation-

exchange chromatography column (CatIEX) and following an initial washing step, the protein was eluted with an increasing NaCl salt gradient (Figure 8.3.2B, C). The eluted fractions containing a high amount of T4L were pooled and passed through a 50 kDa MWCO filter to remove higher molecular weight impurities. The resulting flow-through was concentrated with a 10 kDa MWCO filter and the buffer exchanged. For storage at -20 °C, 20 % w/v glycerol was added to the concentrated protein solution and aliquots of ~ 50 - 100 µL were prepared.

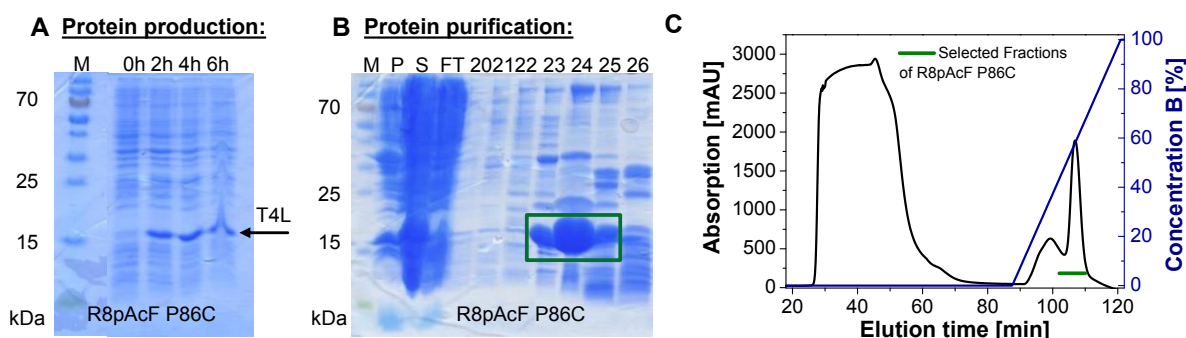


Figure 8.3.2 Protein production and purification of variant R8pAcF P86C. (A) 1 mL samples were taken prior to induction and every two hours after induction of protein production. For comparison purposes, the OD600 of the samples was adjusted to 0.3 before application to the SDS PAGE. (B, C) Protein purification of the 2 L culture shown in (A). (B) The SDS gel shows a sample from the insoluble (P) and soluble (S) fraction after cell lysis and centrifugation, then a sample from the flow-through (F, protein not binding to the CatIEX-column) and finally the elution fraction showing absorption at 280 nm. (C) Chromatogram of the CatIEX purification of (B). Fraction selected for further purification and concentration are marked in green.

8.3.3. Site-specific labeling

For the fluorescence measurements conducted later on in this work, all variants were labeled in at least two different kinds: The donor-only single labeled type (DOnly) and the both dyes containing double-labeled type (DA). An acceptor-only sample is in most cases not needed as by excitation of the acceptor the donor is not excited and also energy transfer cannot take place. The properties of the acceptor (~ position) can thus be assessed by direct acceptor excitation of DA samples (denoted as A_{dir}). As donor dye Alexa488 and as acceptor dye Alexa647 were used. In general, the donor was attached to the unnatural amino acid pAcF inserted in the N-terminus, whereas the acceptor was coupled to the cysteine residue in the C-terminus as described in the Materials and Methods section. A SDS-PAGE of the single and double labeled variant K60pAcF N132C under standard exposure and fluorescence excitation is shown in Figure 8.3.3. The band of the DA sample is shifted slightly upwards because of the additional labeling with the A647-maleimide fluorophore which results in an increase of ~ 1.3 kDa in mass.

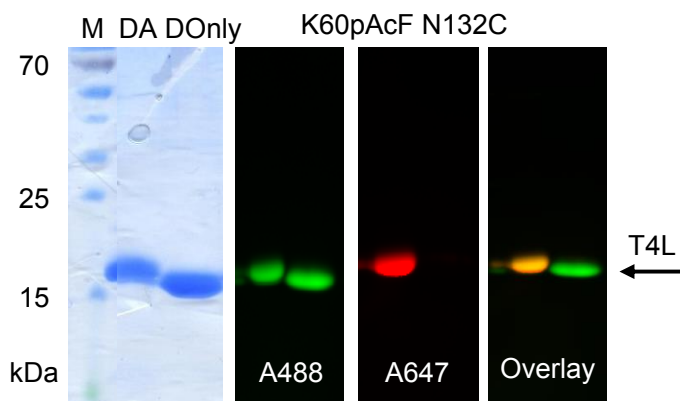


Figure 8.3.3 Site-specific Labeling of variant K60pAcF N132C. (From left to right:) Coomassie stained SDS gel of double-labeled (DA) and donor-only labeled (DOnly) K60pAcF N132C. The same gel when exciting at 480 nm and at 650 nm and observe the emission at 520 nm and 670 nm, respectively. An overlay of the two fluorescence pictures results in an orange signal where both green and red emission are located (DA-sample). The band of the DA-sample is shifted slightly upward due to the additional acceptor fluorophore attached to the protein (1.3 kDa in mass).

After removal of non-reacted dye, the labeled proteins were concentrated, their labeling degree determined and stored in aliquots at -20 °C until needed for experiments.

8.4. Influence of urea on the Förster Radius

Next to the fluorophore-specific characteristics, spectral overlap integral and donor fluorescence quantum yield, also the index of refraction of the surrounding solvent has an influence on the Förster radius R_0 . Thus, when doing FRET experiments in buffers containing ingredients influencing their index of refraction and/or the spectral properties of the used fluorophores, it is advisable to recheck the literature values given.

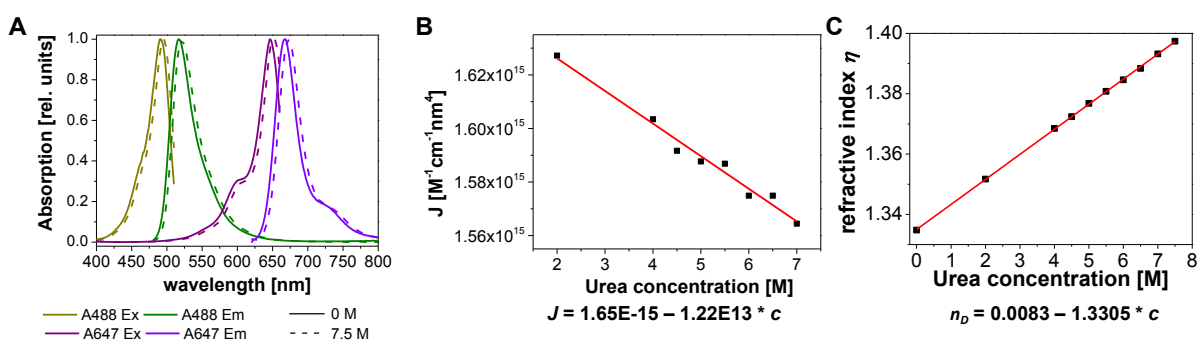


Figure 8.4.1 Influence of urea on the Förster Radius. (A) Excitation and emission spectra of Alexa488 and Alexa647 in PBS and in PBS with 7.5 M urea. The excitation wavelengths were 470 and 610 nm, respectively; the emission wavelengths 520 and 670 nm, respectively. (B) Dependency of the overlap integral J on the urea concentration. With increasing urea concentration the overlap integral decreases linearly. (C) Change of refractive index η with the urea concentration. The target urea concentrations were prepared by mixing different amounts of PBS with PBS containing additionally 8 M urea and the index of refraction n_D was determined. The actual urea concentration was calculated using the following formula: c_{Urea} [mol/L] = $117.66 \cdot n_D + 29.753 \cdot n_D^2 + 185.56 \cdot n_D^3$ (166).

Hence, the spectra from both free donor and acceptor fluorophores were measured in increasing concentrations of urea. Figure 8.4.1A shows the excitation and emission spectra of free fluorophores in PBS and in PBS with 7.5 M urea. All spectra show a small bathochromic shift of ~ 3 nm. The calculated overlap integral decreases continuously with increasing urea concentration (Figure 8.4.1B).

Table 8.4.1 Reduced Förster radius R_{0r} and Förster radius R_0 assuming a donor fluorescence quantum yield Φ_D of 0.8 in increasing urea concentrations for the dye pair Alexa488-Alexa647.

Urea concentration [M]	R_{0r} [Å]	$R_0 (\Phi_D = 0.8)$ [Å]	n_D
0	55.7	53.7	1.3305
2	55.3	53.3	1.3469
3	55.0	53.0	1.3551
4	54.7	52.7	1.3633
4.5	54.6	52.6	1.3674
5	54.4	52.4	1.3715
5.5	54.3	52.3	1.3756
6	54.1	52.2	1.3797
6.5	54.0	52.1	1.3838
7	53.9	51.9	1.3879
7.5	53.9	51.9	1.3920

Additionally, the influence of the changed refractive index of the buffer was evaluated. Here, the relationship between urea concentration (in M) and the change in refractive index has been already derived from Warren and Gordon (166). The original derived equation shows a cubic dependency, however, in our used measurement range, it can be approximated to be linear with high accuracy (Figure 8.4.1C). Table 8.4.1 lists the reduced Förster radius R_{0r} and the Förster radius R_0 assuming a donor fluorescence quantum yield of 0.8 when applying both corrections, the decrease in spectral overlap and the increase in index of refraction. When changing the solvent from PBS to PBS with 7.5 M urea, R_0 decreases by ~ 2 Å.

8.5. Steady-state Fluorescence

The simplest FRET experiment which can be performed to determine whether a factor influences the distance between two fluorophores of a sample, and thus the FRET efficiency, - prior to any more advanced methods - is to measure its ensemble fluorescence signal in a fluorescence spectrometer. Figure 8.5.1A and B show the influence of urea on the two fluorophores. After correction for dilution, theoretically a stable signal would be expected, if the fluorescence of the fluorophores would not be influenced by its surrounding solvent and the proximity of (possible quenchers in) the protein they are coupled to.

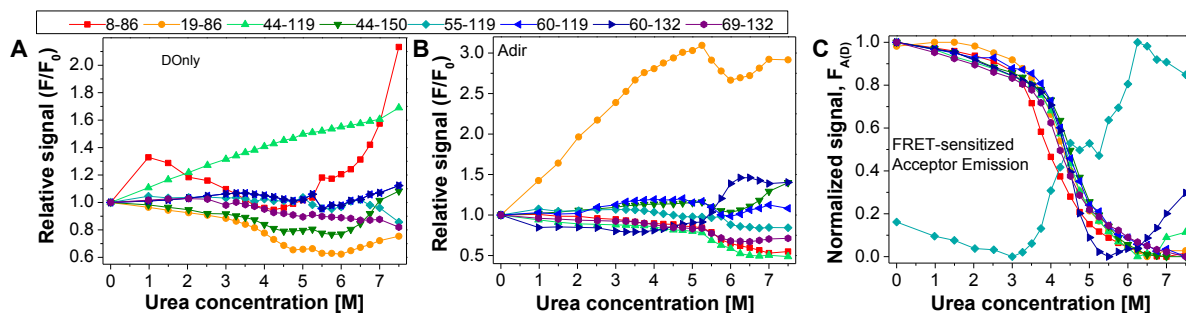


Figure 8.5.1 Influence of urea on the steady-state ensemble fluorescence. (A). Relative donor fluorescence of the DOnly sample corrected for dilution. For comparison purposes the dilution corrected signal was divided by the signal determined in the absence of urea (0 M). (B) Same as (A) for direct-acceptor excitation acceptor fluorescence corrected for dilution. (C) Acceptor fluorescence recorded after donor excitation, corrected for dilution and urea effects on the fluorophores (A, B). For comparison purposes the signals were normalized. The color code is the same for all figures.

The FRET signal recorded when exciting the donor fluorophore and observing the acceptor fluorescence shows the typical sigmoidal progress for a two-state unfolding system (Figure 8.5.1C). The calculated midpoints of transitions (MoT) are summarized in Table 8.5.1 and coincide – as expected - with the MoT's determined from time-resolved measurements (see Chapter 4). Interestingly, the signal of the N55pAcF R119C variant shows an increase in steady-state FRET signal, whereas the fluorescence averaged fluorescence lifetime shows a slight increase (3.2 ns to 3.4 ns). However, the changes in both observables for this variant are rather small.

Table 8.5.1 MoT's of the steady-state ensemble titration for the different variants.

Variant	MoT [M urea]
R8pAcF P86C	3.9
K19pAcF P86C	4.3
S44pAcF R119C	4.7
S44pAcF I150C	4.4
N55pAcF R119C	4.3
K60pAcF R119C	4.4
K60pAcF N132C	4.7
Q69pAcF N132C	4.2

PRESENTATIONS & PUBLICATIONS

Part of this work has been presented at several conferences.

Talks & Poster presentation

International Discussion Meeting on FRET in life sciences 2, Göttingen, 04/2016, - Oral Presentation: ‘Multiparameter Fluorescence spectroscopic toolkit resolves the heterogeneity of unfolded states’

Biophysics Department Colloquium, Clemson University, Clemson SC, USA, 03/2016, - Oral Presentation: ‘Multiparameter Fluorescence Detection sheds light on unfolding pathways and unfolded states of proteins’

60th Annual Meeting of the Biophysical Society, Los Angeles, USA, 02/2016, - Poster Presentation: ‘Resolving the heterogeneity of the ensemble of unfolded states by a combination of fluorescence spectroscopic methods’

25th Faltertage Regensburg, Regensburg, 10/2014, - Poster Presentation & Talk: ‘Conformational fluctuations of Phage T4 Lysozyme under denaturing conditions: Deciphering energy landscape and intermediate structures’

20th International Workshop on “Single Molecule Spectroscopy and Ultra Sensitive Analysis in the Life Sciences”, Berlin, 09/2014, - Oral presentation: ‘Conformational fluctuations of Phage T4 Lysozyme under denaturing conditions: Deciphering energy landscape and intermediate structures’

58th Annual Meeting of the Biophysical Society, San Francisco, USA, 02/2014, - Poster Presentation: ‘Deciphering Folding pathways of Phage T4 Lysozyme’

Biostruct Symposium, Düsseldorf, 11/2013, - Poster Presentation: ‘Deciphering Folding pathways of Phage T4 Lysozyme’

Biostruct Symposium, Düsseldorf, 11/2013, - Oral presentation: ‘Deciphering the structure and energy landscape of T4L using fluorescence methods’

24th Faltertage Regensburg, Regensburg, 10/2013 - Poster Presentation: ‘Deciphering Folding pathways of Phage T4 Lysozyme’

Publications (in preparation)

Sanabria, H., Rodnin, D., Hemmen, K., Peulen T., Felekyan, S., Fleissner, M., Dimura, M., Koberling, F., Kühnemuth, R., Hubbell, W., Gohlke, H., Seidel, C. A. M.; Dynamics and function of transient states in single enzyme molecules.

Hemmen, K., Rodnin, D., Rohrbeck, D., Rezaei-Adariani, S., Felekyan, S., Kühnemuth, R., Sanabria, S., Seidel, C. A. M.; Multiparameter Fluorescence spectroscopy reveals sub-millisecond intermediate transitions in protein unfolding.

Hemmen, K., Rodnin, D., Markovic, I., Felekyan, S., Kühnemuth, R., Sanabria, S., Seidel, C. A. M.; Fluorescence spectroscopic toolkit resolves the heterogeneity of denatured states.

Rodnin, D., Hemmen, K., Sanabria, S., Seidel, C. A. M.; Denaturation of T4 lysozyme and the effects of the non-ionic surfactant Tween20.

Tsytlonok, M., Sanabria, H., Wang, Y., Felekyan, S., Hemmen, K., Yun, M., White, S., Waddell, B., Park, C., Vaithiyalingam, S., Iconaru, L., White, S.-W., Tompa, P., Seidel, C. A. M., Kriwacki, R.; Dynamic anticipation by Cdk2/Cyclin A-bound p27 mediates signal integration in cell cycle regulation

Further publications

Hemmen, K., Reinl, T., Buttler, K., Behler, F., Dieken, H., Jänsch, L., Wilting, J., Weich, H. A.; High-resolution mass spectrometric analysis of the secretome from mouse lung endothelial progenitor cells; Angiogenesis; Jan 15 2011

Badar, M., Hemmen, K., Nimtz, M., Stieve, M., Stiesch, M., Lenarz, T., Hauser, H., Möllmann, U., Vogt, S., Schnabelrauch, M., Müller, P.P.; Evaluation of Madura-hydroxylactone as a Slow Release Antibacterial Implant Coating; The Open Biomedical Engineering Journal, 2010, 4, 263-270

DANKSAGUNG

An dieser Stelle möchte ich mich herzlich bei den Menschen bedanken, ohne die es nicht zu dieser Arbeit gekommen wäre:

Prof. Dr. Claus Seidel für die Betreuung, die Bereitstellung des Themas und den zahlreichen Diskussionen.

Prof. Dr. Dieter Willbold für die Übernahme des Zweitgutachtens.

Prof. Dr. Hugo Sanabria für die gute Betreuung, die auch mit seinem Wegzug aus Deutschland nicht geendet hat.

Annemarie Greife, Jakub Kubiak, Hugo Sanabria und vor allen Dingen Thomas Peulen für das Korrekturlesen dieser Arbeit.

Suren Felekyan und Ralf Kühnemuth für die Hilfe beim Durchführen und Auswerten der Korrelationsmessungen.

Meinen Eltern für's immer-da-sein.

Zuletzt möchte ich mich für die finanzielle Unterstützung durch die NRW Forschungsschule ***BioStruct*** und der International Heinrich-Heine Graduate School of Protein Science ***iGRASP_{seed}*** während der ersten drei Jahre meiner Promotion bedanken.

EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich, dass ich die Promotion mit dem Thema

“Structure and conformational fluctuations of Phage T4 Lysozyme under native and denaturing conditions”

am Institut 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 03.06.2016

Katherina Hemmen