



HEINRICH HEINE  
UNIVERSITÄT DÜSSELDORF

Mapping protein structure and dynamics by Förster  
Resonance Energy Transfer *in vitro* and in live cells

Inaugural-Dissertation

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

vorgelegt von

Thomas-Otavio Peulen

aus Kiel

Düsseldorf, Dezember 2015

Institute of Molecular Physical Chemistry,  
Heinrich-Heine University Düsseldorf

Printed with the permission of  
Department of Mathematics and Natural Sciences  
Heinrich-Heine-University Düsseldorf

1. Referee: Prof. Dr. C.A.M. Seidel  
2. Referee: Prof. Dr. H. Gohlke

Date of the oral examination: 10.03.2016

## Mapping protein Structure and Dynamics by Förster Resonance Energy Transfer *in vitro* and in live cells

Structural, biochemical and cellular aspects have to be considered to obtain a full picture on the native function of proteins. In this context fluorescence spectroscopy with multi-parameter fluorescence detection (MFD) combined Förster Resonance Energy transfer (FRET) with is an ideal method, as it informs on protein structure and dynamics in living cells. MFD detects spectral and polarization resolved fluorescent photons with picosecond resolution, while FRET measures distances with Ångström accuracy. Therefore, a combination of MFD, FRET and computational methods is ideal to study biochemical-, structural and dynamics. Cellular aspects are considered by a combination of MFD with microscopic imaging (multi-parameter image spectroscopy, MFIS) allowing to solve the multifactorial problem of understanding biomolecular function in their native environment.

FRET-measurements are rarely used to determine quantitative structural models. Therefore, a set fluorescence methods was established and combined with computational simulations to cumulate multiple FRET-measurements into a quantitative structural models and to determine equilibrium constants in in living cells. These methods were tested on a *static* biomolecular structure. As biomolecules are often dynamic and hence heterogeneous, an analytical method for fluorescence intensity decays was developed to resolve structural heterogeneities of conformational ensembles by FRET. Next, a theory describing multi-dimensional FRET-efficiency histograms of was established to reliably detect conformational dynamics of biomolecules. Finally, methods were developed to determine equilibrium constants in living cells by MFIS.

Using the developed methods as toolkit T4 lysozyme was studied as a model system and a transiently populated conformational state was resolved *in vitro*. Next, the dimerization of guanylate binding proteins (GBPs) was studied by *in vitro* and live-cell. *In vitro* a previously postulated binding mode of the human GBP1 (hGBP1) was confirmed and new binding mode was found. Combining the developed analysis methods and the *in vitro* information GBP oligomerization was quantify in live-cell by equilibrium constants. Finally, the hGBP1 monomer was studied in detail *in vitro* by ensemble and single-molecule measurements. Here, two conformational states were resolved and described by structural models suggesting a potential dimerization pathway. Multiple FRET fluorescence correlation spectroscopy measurements quantify the transitions among the two monomeric states and mapping conformational dynamics to structure. These presented studies demonstrate how fluorescence and FRET-measurements improve our understanding of molecular machines.



## Acknowledgments

The research presented in this dissertation was conducted from January 2010 to October 2015 at the chair of Molecular Physical Chemistry at the Heinrich Heine University Düsseldorf under the supervision of Prof. Dr. Claus A.M. Seidel.

First, I'm most grateful to Prof. Dr. Claus A.M. Seidel who gave me the opportunity to work on numerous different exciting projects using state of the art equipment to explore highly promising applications of fluorescence. I also like to thank Ralf Kühnemuth and Suren Felekyan for their constant endeavor to keep the laboratory equipment and software functional. Without their efforts this work would certainly not be possible. Next, I thank Alessandro Valeri who introduced me into single-molecule measurements and the workgroup. Additionally, I thank the collaborators of the hGBP1 project: Carola Hengstenberg, Yathrib Ajaj, Johann Klare, Andreas Stadler and Christian Herrmann. Moreover, I'd like to thank Hugo Sanabria for being a patient co-author and excellent leader on the T4L-project. I also thank Katharina Hemmen, Dmytro Rodnin for our joint work on T4L. I additionally thank Dmytro Rodnin for providing the experimental data on IF3. Qijun Ma, Annemarie Greife and Elisabeth Kravets I thank for our joint work on live-cell imaging. Mykola Dimura I thank for being an excellent partner for scientific discussions and our joint work on hGBP1, T4L and TGR5. Oleg Opanasyuk I thank for our fruitful discussions on time-resolved fluorescence measurements and his critical comments on our joint manuscript. I also thank Mark Bowen, McCann, Jakub Kubiak for providing samples and time-resolved fluorescence measurements; Paul Rothwell, Silvia Berger for providing single-molecule measurements used in the manuscript "A combined graphical and analytical method for accurate analysis of time-resolved FRET measurements". Furthermore, thank Katherina Hemmen, Aishwaria Prakash, Qijun Ma, Suren Felekyan and Ralf Kühnemuth for proof-reading of this dissertation. I thank Olga Doroshenko, and Deborah Sandrin, Denis Dörr and Peter Zentis for being nice office mates and excellent colleagues.

Finally, I thank the International Helmholtz Research School of Biophysics and Soft Matter (IHRS BioSoft) for funding

### Overview

The following manuscripts are either published or currently in revision can be found in the supplement of this dissertation:

- A toolkit and benchmark study for FRET-restrained high-precision structural modeling
- Triphosphate induced dimerization of human guanylate binding protein 1 involves association of the C-Terminal helices: A joint double electron-electron resonance and FRET Study
- Guanylate binding proteins (GBPs) directly attack *T. gondii* via supramolecular complexes

The following chapters of this dissertation focus on fluorescence measurements as analytical technique to reveal dynamic and structural features of biomolecules:

- A combined graphical and analytical method for accurate analysis of time-resolved FRET measurements: FRET between tethered fluorophores in presence of local quenching
- Experimental reference samples for accurate time-resolved FRET analysis: Handling positional effects of fluorescent dyes in bis-functional proteins
- A generalized description of multidimensional single-molecule FRET histograms: Theoretical treatment of FRET parameters of multistate dynamics systems
- Quantitative FRET-analysis on measurements of living cells

The chapter “Quantitative FRET-analysis on measurements of living cells” is a summary of the developed analytical methods for live cell imaging. The other chapters represent the current stage of manuscripts of planned publications.

The developed methods were applied to a set of proteins: the HIV reverse transcriptase (HIV-RT), the human guanylate binding protein 1 (hGBP1), T4 lysozyme (T4L), the G protein-coupled bile acid receptor TGR5 and murine guanylate binding proteins (mGBPs). The measurement on mGBPs and TGR5 are live-cell measurements while all other measurements were performed *in vitro*. The strength of the methods in determining dynamic protein structures is demonstrated using hGBP1 and T4L as prime examples. The relevance of the methods is exemplified in the following chapters:

- Refining fluorescence tools to resolve hidden conformational states of T4L by FRET
- Mapping kinetics and structure of a large GTPase resolves a hidden state essential for oligomerization

The chapter on T4L summarizes the main contribution to the T4L-project. The “large GTPase”-chapter (hGBP1) summarizes findings on hGBP1 by SAXS, EPR and FRET. The hGBP1-chapter is a manuscript in early draft stage.

---

**This thesis is based on the following papers and manuscripts:****Published or in revision**

1. **T.-O. Peulen**, C.A.M. Seidel, “Struktur und Dynamik von Biomolekülen mit high precision-FRET“ (2011) *BIOspektrum* 17(7), 4590-4600
2. S. Kalinin, **T.-O. Peulen**, S. Sindbert, P.J. Rothwell, S. Berger, T. Restle, R.S. Goddy, H. Gohlke, C.A.M. Seidel, “A toolkit and benchmark study for FRET-restrained high-precision structural modeling”, (2012) *Nature methods* 9 (12), 1218-1225 (**Supplement A**)
3. **T.-O. Peulen**, T. Vöpel, C.S. Hengstenberg, Y. Ajaj, C.A.M. Seidel, C. Herrmann, J.P. Klare “Triphosphate induced dimerization of human guanylate binding protein 1 involves association of the C-Terminal helices: A joint double electron-electron resonance and FRET Study”, (2014) *Biochemistry* 53 (28), 4590-4600 (**Supplement B**)
4. E. Kravets, D. Degrandi, Q. Ma, **T.-O. Peulen**, S. Felekyan, R. Kühnemuth, S. Weidtkamp-Peters, C.A.M. Seidel, K. Pfeffer “Guanylate binding proteins (GBPs) directly attack *T. gondii* via supramolecular complexes”, in revision (eLife) (**Supplement C**)

**Manuscripts in preparation**

5. **T.-O. Peulen**, O. Opanasyuk, C.A.M. Seidel “A combined graphical and analytical method for accurate analysis of time-resolved FRET measurements: FRET between tethered fluorophores in presence of local quenching”, Manuscript written (**Chapter A**)
6. **T.-O. Peulen**, C.S. Hengstenberg, J. Kubiak, M. Richert, J.J. McCann, M. Bowen, C. Herrmann, C.A.M. Seidel: “Experimental reference samples for accurate time-resolved FRET analysis: Handling positional effects of fluorescent dyes in bis-functional proteins”, Manuscript written (**Chapter B**)
7. **T.-O. Peulen**, H. Sanabria, S. Kalinin, S. Felekyan, C.A.M. Seidel “A generalized description of multidimensional single-molecule FRET histograms: Theoretical treatment of FRET parameters of multistate dynamics systems”, Manuscript written (**Chapter C**)
8. Q. Ma, M. Somssich, S. Kalinin, **T.-O. Peulen**, R. Kühnemuth, Y. Stahl, R. Simon, S. Weidtkamp-Peters, C.A.M. Seidel, “Revealing structural features and affinities of protein complexes in living cells by MFIS-FRET analysis“, Manuscript written (**Chapter D**)
9. A. Greife, S. Felekyan, Q. Ma, C.G.W. Gertzen, M. Dimura, **T.-O. Peulen**, L. Spomer, C. Wöhler, D. Häussinger, H. Gohlke, V. Keitel, C.A.M. Seidel “Structural assemblies of the di- and oligomeric G-protein coupled receptor TGR5 in live cells: an MFIS-FRET and integrative modeling study”, Manuscript written (**Chapter D**)
10. H. Sanabria, D. Rodnin, K. Hemmen, **T.-O. Peulen**, S. Felekyan, M.R. Fleissner, M. Dimura, F. Koberling, R. Kühnemuth, W. Hubbell, H. Gohlke, C.A.M. Seidel “Dynamics and function of transient states in single enzyme molecules”, Manuscript written (**Chapter E**)
11. **T.-O. Peulen**, M. Dimura, K. Hemmen, D. Rodnin, H. Gohlke, C.A.M. Seidel “FRET restrained structural modelling resolves an excited state in the catalytic cycle of T4L”, Manuscript planed (**Chapter E**)
12. H. Sanabria, J. Kubiak, **T.-O. Peulen**, S. Felekyan, D. Rodnin, R. Kühnemuth, C.A.M. Seidel, “Determination of biomolecular structures and their dynamics by fluorescence spectroscopic methods”, Manuscript written (Review).
13. **T.-O. Peulen**, C.S. Hengstenberg, R. Biehl, M. Dimura, A. Valeri, S. Ince, T. Vöpel, B. Faragó, H. Gohlke, C. Herrmann, J.P. Klare, A. Stadler, C.A.M. Seidel, “Mapping kinetics and structure of a large GTPase resolves a hidden state essential for oligomerization”, Manuscript in preparation (**Chapter F**)

## Contributions

### Chapter A:

#### **A combined graphical and analytical method for accurate analysis of time-resolved FRET measurements: FRET between tethered fluorophores in presence of local quenching**

T.-O. Peulen, O. Opanasyuk, C.A.M. Seidel

- designed the research: TOP, CAMS
- developed software: TOP, OO
- analyzed and interpreted the data: TOP, OO, CAMS
- drafted and revised the manuscript: TOP, CAMS

AG Seidel in total: 100%*		
TOP	OO	CAMS
60 (60)	10 (10)	30 (30)

\* In brackets the relative contribution within the Seidel group is shown.

Detailed personal contribution of TOP:

- a. Development of theory
- b. Writing of the manuscript
- c. All figures except for the statistical error-estimation

### Chapter B:

#### **Experimental reference samples for accurate time-resolved FRET analysis: Handling positional effects of fluorescent dyes in bis-functional proteins**

T.-O. Peulen, C.S. Hengstenberg, J. Kubiak, M. Richert, J.J. McCann, M. Bowen, C. Herrmann, C.A.M. Seidel

- designed the research: TOP, CSH, CAMS
- performed the measurements: TOP, MR, JK
- prepared the samples: CSH, JK, JJM
- developed analysis software: TOP
- analyzed and interpreted the data: TOP, JK
- drafted and revised the manuscript: TOP, CSH, JK, MB, CH, CAMS

AG Seidel in total: 60%*			
TOP	JK	MR	CAMS
50 (30)	15 (10)	5 (3)	30 (18)

\* In brackets the relative contribution within the Seidel group is shown.

Detailed personal contribution of TOP:

- a. Development of and implementation of analytical methods
- b. Theory
- c. Data-analysis
- d. Writing of manuscript
- e. Figures of manuscript

**Chapter C:****A generalized description of multidimensional single-molecule FRET histograms: Theoretical treatment of FRET parameters of multistate dynamics systems**

H. Sanabria , T.-O. Peulen, S. Kalinin, S. Felekyan, C.A.M. Seidel

- designed the research: HS, TOP, SK, CAMS
- drafted and revised the manuscript: HS, TOP, SK, SF, CAMS

AG Seidel in total: 100%*				
HS	TOP	SK	SF	CAMS
25 (25)	25 (25)	15 (15)	10 (10)	25 (25)

\* In brackets the relative contribution within the Seidel group is shown.

Detailed personal contribution of TOP:

- Generalization of FRET-lines by the FRET-induced donor decay
- Clarification of the relation between FRET-efficiency, fluorescence averaged lifetimes as the moments of lifetime distribution
- Writing of approx. half of the manuscript
- Approx. half of the figures

**Chapter D:****Revealing structural features and affinities of protein complexes in living cells by MFIS-FRET analysis**

Q. Ma, M. Somssich, S. Kalinin, T.-O. Peulen, R. Kühnemuth, S. Felekyan, Y. Stahl, R. Simon, S. Weidtkamp-Peters, C. A.M. Seidel

- designed the research: RS, SWP and CAMS
- calibrated the experimental setup and performed the experiments: QM
- prepared the samples and performed the experiments: MS
- developed software: SK and SF, TOP
- analyzed and interpreted the data: QM, MS, TOP, RK, SF, YS, RS, SWP and CAMS
- QM, MS, RS, SWP and CAMS drafted and revised the manuscript.

AG Seidel in total: 65%*					
QM	SK	TOP	RK	SF	CAMS
46 (30)	8 (5)	8 (5)	8 (5)	8 (5)	22 (15)

\* In brackets the relative contribution within the Seidel group is shown.

Detailed personal contribution of TOP:

- Development of simulations and simulation analysis software for conformational sampling of eGFP-mCherry
- Programming of model functions for fluorescence decay analysis
- Application of FRET induced donor decay to live-cell measurements

### Chapter D:

#### Structural assemblies of the di- and oligomeric G-protein coupled receptor TGR5 in live cells: an MFIS-FRET and integrative modeling study

A. Greife, S. Felekyan, Q. Ma, C. G.W. Gertzen, L. Spomer, M. Dimura, T.-O. Peulen, C. Wöhler, D. Häussinger, H. Gohlke, V. Keitel, C. A.M. Seidel

- designed the research : AG, DH, HG, VK and CAMS
- performed the experiments: AG, QM, LS and CW
- performed the computer simulation: CGWG, MD and TOP
- analyzed and interpreted the data: AG, SF, QM, CGWG, LS, MD, TOP, HG, VK and CAMS
- drafted and revised the manuscript: AG, SF, QM, CGWG, LS, MD, TOP, HG, VK and CAMS

AG Seidel in total: 45%*					
AG	SF	QM	MD	TOP	CAMS
36 (16)	18 (8)	11 (5)	9 (4)	9 (4)	18 (8)

\* In brackets the relative contribution within the Seidel group is shown.

Detailed personal contribution of TOP:

- a. Method development to determine  $K_D$  by multi-dimensional analysis of concentrations and transfer-efficiencies
- b. Together with Mykola Dimura development of “implicit dye-linker” model for GFPs by weighting of accessible volume simulations
- c. Direct fitting of experimental fluorescence decays by simulated structural models.

### Chapter E:

#### Dynamics and function of transient states in single enzyme molecules

H Sanabria, D Rodnin, K Hemmen, T.-O. Peulen, S. Felekyan, M.R. Fleissner, M. Dimura, F. Koberling, R. Kühnemuth, W. Hubbell, H. Gohlke, C.A.M. Seidel

- purified and labeled the protein: HS, KH
- measured and analyzed the experiments: HS, TOP, KH, DR
- performed structural screening: TOP, MD, HG
- performed and analyzed EPR-experiment: MRF
- performed and study design and EPR analysis: WH
- developed fluorescence analysis tools: SF, TOP
- developed fluorescence instrumentation and gave technical support: FK, RK
- discussed the results and commented on manuscript: all authors
- wrote the paper: HS, WH, CAMS
- supervised the project: CAMS

AG Seidel in total: 89%*							
HS	DR	KH	TOP	SF	MD	RK	CAMS
22 (20)	22 (20)	22 (20)	9 (8)	5 (4)	4 (3)	4 (3)	12 (11)

\* In brackets the relative contribution within the Seidel group is shown.

Detailed personal contribution of TOP:

- a. Development of analysis software and fluorescence models
- b. Software for statistical error error estimation

**Chapter E:****FRET restrained structural modelling resolves an excited state in the catalytic cycle of T4L**

M. Dimura, T.-O. Peulen, K. Hemmen, D. Rodnin, H. Sanabria, H. Gohlke, C.A.M. Seidel

- measured and analyzed experiment: TOP, KH, DR
- performed molecular simulations: HG, MD
- implemented simulation workflow: TOP, MD
- wrote the manuscript: MD, TOP, KH
- supervised project: HS, HG, CAMS
- discussed the results and commented on manuscript: all authors

AG Seidel in total: 100%*						
MD	TOP	KH	DR	HG	CAMS	HS
20 (20)	20 (20)	20 (20)	10 (10)	10 (10)	10 (10)	10 (10)

\* In brackets the relative contribution within the Seidel group is shown.

Detailed personal contribution of TOP:

- Influence of state-specific donor fluorescence
- Analysis of the fluorescence data

**Chapter F:****Mapping the motions and structure of a state necessary for oligomerization of a large GTPase**

T.-O. Peulen, C.S. Hengstenberg, R. Biehl, M. Dimura, A. Valeri, S. Ince, T. Vöpel, B. Faragó, H. Gohlke, C. Herrmann, J.P. Klare, A. Stadler, C.A.M. Seidel

- measured and analyzed experiment: TOP, RB, AV, JPK, AS
- performed molecular simulations: HG, MD
- implemented simulation workflow: TOP, MD
- wrote the manuscript: TOP, CSH, RB, MD, HG, JPK, AS, CAMS
- supervised project: CH, AS, CAMS
- discussed the results and commented on manuscript: all authors

AG Seidel in total: 25%*			
TOP	MD	AV	CAMS
50 (13)	15 (4)	5 (1)	30 (8)

\* In brackets the relative contribution within the Seidel group is shown.

Detailed personal contribution of TOP:

- Fluorescence analysis
- Coarse grained rigid body docking
- Screening of the conformational ensembles
- Unifying SAXS, EPR and FRET results

## Preamble

---

### Supplement A:

#### A toolkit and benchmark study for FRET-restrained high-precision structural modeling

S. Kalinin, T.-O. Peulen, S. Sindbert, P.J. Rothwell, S. Berger, T. Restle, R.S. Goddy, H. Gohlke, C.A.M. Seidel

- designed the research: SK, CAMS, RSG
- performed the simulations: HG
- performed the measurements: PJR, SB
- prepared the samples: PJR, TR
- developed the analysis software: SK
- analyzed and interpreted the data: SS, TOP
- drafted and revised the manuscript: SK, SS, TOP, CAMS

AG Seidel in total: 80%*					
SK	TOP	SS	PJR	SB	CAMS
30 (24)	15 (12)	15 (12)	5 (4)	5 (4)	30 (24)

\* In brackets the relative contribution within the Seidel group is shown.

Detailed personal contribution of TOP:

- a. Analysis of approx. half of the 44 single-molecule fluorescence data sets
- b. Rigid body docking and screening of molecular dynamic simulations
- c. Writing parts of the manuscript
- d. Figures of the manuscript

### Supplement B:

#### Triphosphate induced dimerization of human guanylate binding protein 1 involves association of the C-terminal helices: A joint double electron-electron resonance and FRET study

T. Vöpel, C.S. Hengstenberg, T.-O. Peulen, Y. Ajaj, C.A.M. Seidel, C. Herrmann, J.P. Klare

- designed the research: CH, CAM
- performed the measurements: YA, TOP, JPK
- prepared the samples: TV, CSH
- developed analysis software: TOP
- analyzed and interpreted the data: TOP, JPK, TV, CSH
- drafted and revised the manuscript: TV, CSH, TOP, CAMS, CH, JPK

AG Seidel in total: 33%*		
TOP	YA	CAMS
60 (20)	10 (3)	30 (10)

\* In brackets the relative contribution within the Seidel group is shown.

Detailed personal contribution of TOP:

- a. Development of error-estimation and data-analysis software for fluorescence intensity decays
- b. Data-analysis of the presented fluorescence measurements
- c. Validation of structural models by FRET
- d. Writing parts of the manuscript

**Supplement C, Chapter D:****Guanylate binding proteins (GBPs) directly attack *T. gondii* via supramolecular complexes**

E. Kravets, D. Degrandi, Q. Ma, T.-O. Peulen, S. Felekyan, R. Kühnemuth, S. Weidtkamp-Peters, C.A.M. Seidel, K. Pfeffer

- designed the research: CAMS and KP
- performed the research: EK, DD, QM and SWP
- analyzed and interpreted the data: EK, DD, QM, TOP, SF, RK, SWP, CAMS and KP
- drafted and revised the manuscript: EK, DD, QM, SWP, CAMS, and KP

AG Seidel in total: 45%*				
QM	TOP	SF	RK	CAMS
42 (19)	22 (10)	11 (5)	2 (1)	22 (10)

\* In brackets the relative contribution within the Seidel group is shown.

Detailed personal contribution of TOP:

- Simulation of mGBP2 dimer
- Determination of  $K_D$  values by analysis of multiple FRET-pairs
- Fitting model of homo-FRET analysis

**Review: Determination of biomolecular structures and their dynamics by fluorescence spectroscopic methods**

H. Sanabria, J. Kubiak, T.-O. Peulen, S. Felekyan, D. Rodnin, R. Kühnemuth, C.A.M. Seidel

- wrote the manuscript: all authors

AG Seidel in total: 100%*						
HS	JK	TOP	SF	DR	RK	CAMS
18 (18)	20 (20)	14 (14)	10 (10)	10 (10)	10 (10)	18 (18)

\* In brackets the relative contribution within the Seidel group is shown.



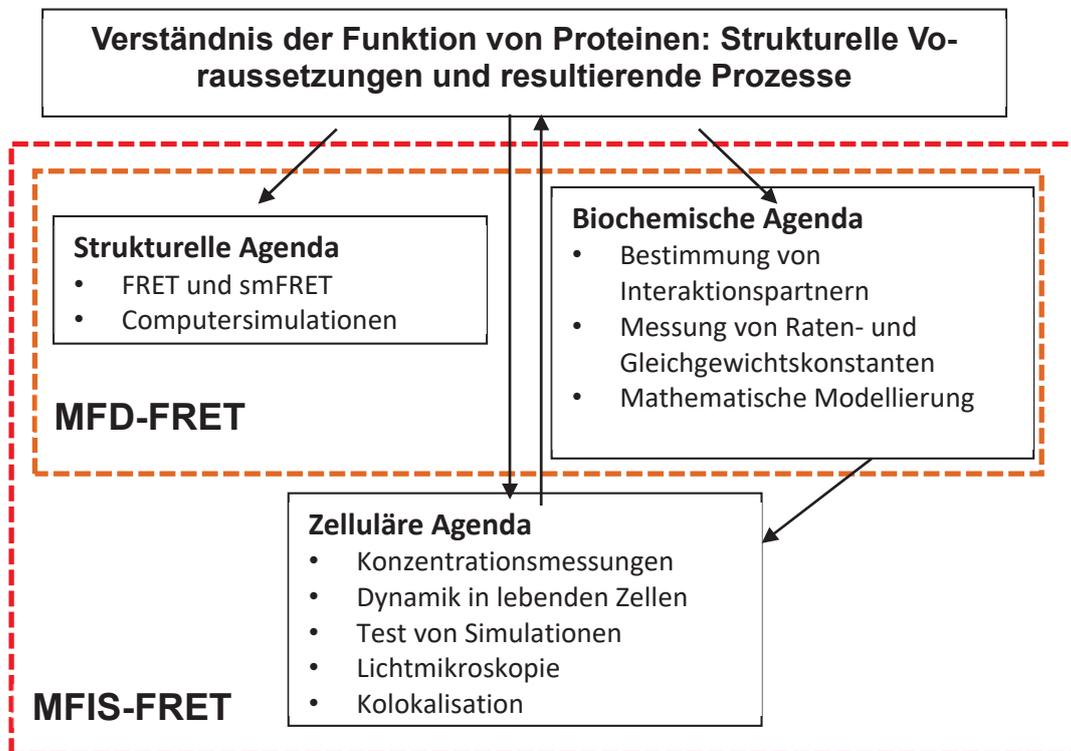
### Ziele

Um die Funktion von Biomolekülen zu verstehen, ist Wissen über ihre strukturellen und biochemischen Eigenschaften erforderlich. Zusätzlich muss, um ihre biologische Funktionen zu erfassen, ein Bezug zum ihren zellulären Kontext hergestellt werden. Dies wird im Folgenden anhand der in dieser Arbeit untersuchten Proteinklasse der Guanylatebindeproteinen (GBPs) dargelegt. GBPs sind in der Immunantwort auf intrazelluläre Parasiten beteiligt. In Säugetierzellen wird ihre Expression durch Interferon-gamma getriggert. GBPs gehören zur Familie der „dynamine-like proteins“ (engl.: Dynamine ähnliche Proteine), und hydrolysieren Guanosin-triphosphat (GTP). In Gegenwart von GTP bilden GBPs Oligomere. Die Bildung eines GBP-Oligomeres hängt sowohl von der GBP-Konzentration als auch von der GTP-Konzentration ab. Bei der GTP-Hydrolyse ändern GBPs ihre Konformation. Die GBP-Oligomere heften sich an Membranen an und attackieren, wie in dieser Arbeit gezeigt wurde, Membranen intrazellulärer Parasiten wie *Toxoplasma gondii*. Dieses Beispiel verdeutlicht, wieso strukturelle (Konformationsänderung bei Hydrolyse), biochemische (Konzentrationsabhängigkeit der Oligomerisierung) als auch zelluläre Aspekte (Angriff der intrazellulären Parasiten) gemeinsam betrachtet werden müssen, um die biologische Funktion der Proteinklasse (zelluläre Immunantwort) zu verstehen.

Bei Biomolekülen stellt die Verknüpfung funktioneller, dynamischer und struktureller Aspekte eine besondere Herausforderung dar, da die Kinetiken welche für ihre Funktion relevant sind mehrere Größenordnungen umfassen: einige Pikosekunden im Falle von Seitenkettenrotationen, bis zu Sekunden bei großen Konformationsänderungen. Daher werden analytische Methoden mit einer hohen Zeitauflösung benötigt. Um hochaufgelöste molekulare Strukturmodelle von Biomolekülen zu entwickeln müssen zudem Abstände in Ångström-Genauigkeit vermessen werden. Zur Untersuchung von Biomolekülen in lebenden Zellen sind hochselektive und gleichzeitig sensitive mikroskopische bildgebende Methoden erforderlich. In diesem Zusammenhang ist Fluoreszenzspektroskopie, insbesondere in Kombination mit mikroskopischen bildgebenden Verfahren, eine äußerst vielversprechende Technik, da Fluoreszenz hochselektiv und sensitiv detektiert werden kann. Multiparameter Fluoreszenz Detektion (MFD) verspricht eine hohe Zeitauflösung, da die detektierten Photonen pikosekundengenau registriert werden, während durch Förster-Resonanzenergietransfer (FRET) Abstände mit Ångström-Genauigkeit zu vermessen werden. Daher ist die Kombination aus MFD und FRET (MFD-FRET) ergänzt durch Computersimulationen eine vielversprechende Methode, um biochemische, strukturelle

## Zusammenfassung

und dynamische Aspekte von Proteinen zu untersuchen. Wie in dieser Arbeit gezeigt wird können durch die Kombination aus MFD-FRET mit Computersimulationen, biomolekulare Strukturen („Strukturelle Agenda“) als auch die Wechselwirkung mit Interaktionspartnern („Biochemische Agenda“) *in vitro* untersucht werden (siehe Abbildung 1).



**Abbildung 1.** Diagramm eines synthetisch reduktionistischen Ansatzes zur Charakterisierung von Proteinfunktionen und biologischer Prozesse. Die Kombination aus MFD-FRET Messungen und Computersimulationen erlaubt uns strukturelle, biochemische Aspekte zu beleuchten. MFIS-FRET adressiert zusätzlich zelluläre Aspekte durch bildgebende Spektroskope in lebenden Zellen. (angepasstes Schema von (Pollard 2013, "No Question about Exciting Questions in Cell Biology.")

Um zelluläre Aspekte zu beleuchten, wird MFD mit bildgebenden mikroskopischen Verfahren zu „Multi-Parameter Image Spectroscopy“ (MFIS) kombiniert. Durch die Kombination in einer Messung können biochemische, zelluläre und strukturelle Aspekte miteinander verknüpft werden (siehe Abbildung 1). Es können Gleichgewichtskonstanten bestimmt, Bindungspartner identifiziert und Kinetiken in lebenden Zellen quantifiziert und dynamische Strukturmodelle generiert werden. Damit werden Einzelbeobachtungen vereint und in den biologischen Kontext gestellt. Momentan liefern *in vitro* MFD-Messungen eine signifikant höhere zeitliche als auch strukturelle Auflösung im Vergleich zu MFIS-Messungen. Daher wurden in dieser Arbeit die Messungen in lebenden Zellen mit zusätzlichen *in vitro* Messungen unter definierten Messbedingungen kombiniert.

## Zusammenfassung

Das Ziel Biomoleküle bei der Arbeit zu beobachten und durch molekulare Strukturmodell zu beschreiben wurde durch einen Satz neu entwickelter Analysemethoden und Techniken verfolgt. Da FRET selten für Strukturmodellierungen genutzt wird, war die Entwicklung neuer Analysemethoden war erforderlich. Die entwickelten Methoden wurden durch *in vitro* Messungen und Messungen in lebenden Zellen etabliert. Zuerst wurde eine Methode entwickelt, Messergebnisse aus mehreren FRET-Messungen in einem Strukturmodell konsolidieren. Diese Methode wurde auf an einem *statischen* Protein Nukleinsäurekomplex angewendet (siehe Anhang A). Später wurde diese Methode erweitert, um dynamische Biomoleküle zu beschreiben. Dazu wurde eine Methode zur Analyse von Fluoreszenzintensitätsabklingkurven entwickelt, welche in der Lage ist Heterogenitäten zu auflösen (siehe Kapitel A). Diese Methode ist die Basis für die Gewinnung von Abstandsinformationen zur strukturellen Modellierung von dynamischen Systemen und erfordert Referenzen von hoher Qualität. In Kapitel B ist beschrieben, wie man solche Referenzen experimentell erhält. Als nächstes wurde eine Theorie für die Analyse mehrdimensionale Einzelmolekül MFD-Histogramme entwickelt, welche es erlaubt es Dynamik mittels Einzelmolekülmessungen zuverlässig zu detektieren (siehe Kapitel C). Auf Basis der in Kapitel A entwickelten Theorie Gleichgewichtskonstanten in lebenden Zellen durch MFIS-Messungen zu bestimmen (Kapitel D). Dadurch konnten strukturelle, biochemische als auch zelluläre Aspekte verknüpft werden.

Die Methoden zur Strukturaufklärung mittels FRET wurden anhand eines Protein/DNA-Komplexes der HIV Reverse Transkriptase validiert (Anhang A). Die entwickelten spektroskopischen Methoden lösten in dem Modellsystem T4 Lysozyme (T4L) *in vitro* einen transient populierten, kurzlebigen Zustandes auf. Die komplexere Proteinklasse der Guanylatebindeproteinen (GBPs) wurde *in vitro* und in lebenden Zellen untersucht. *In vitro* wurde die Dimerisierung des menschlichen Guanylatebindeprotein 1 (hGBP1) durch zeitaufgelöste Fluoreszenzmessungen und EPR-Messungen untersucht (Anhang B) und ein bis dato unbekanntes Bindungsmodus entdeckt während ein postulierter Bindungsmodus bestätigen wurde. Die Dimer-Strukturmodelle in Kombination mit den in Kapitel D präsentiert Methoden erlaubten es Homo- und Hetero-Oligomerisierung von GBPs in lebenden Zellen zu untersuchen (Anhang C). Zuletzt wurde die Struktur von hGBP1 durch Ensemble- und Einzelmolekülmessungen aufgelöst und wurden zwei Konformationszustände entdeckt (Kapitel F). Diese Zustände wurden mit Hilfe von molekularen Simulationen modelliert während die Kinetik des Austausches zwischen den Zuständen durch Fluoreszenzkorrelationspektroskopie quantifiziert, um ein dynamisches Bild der

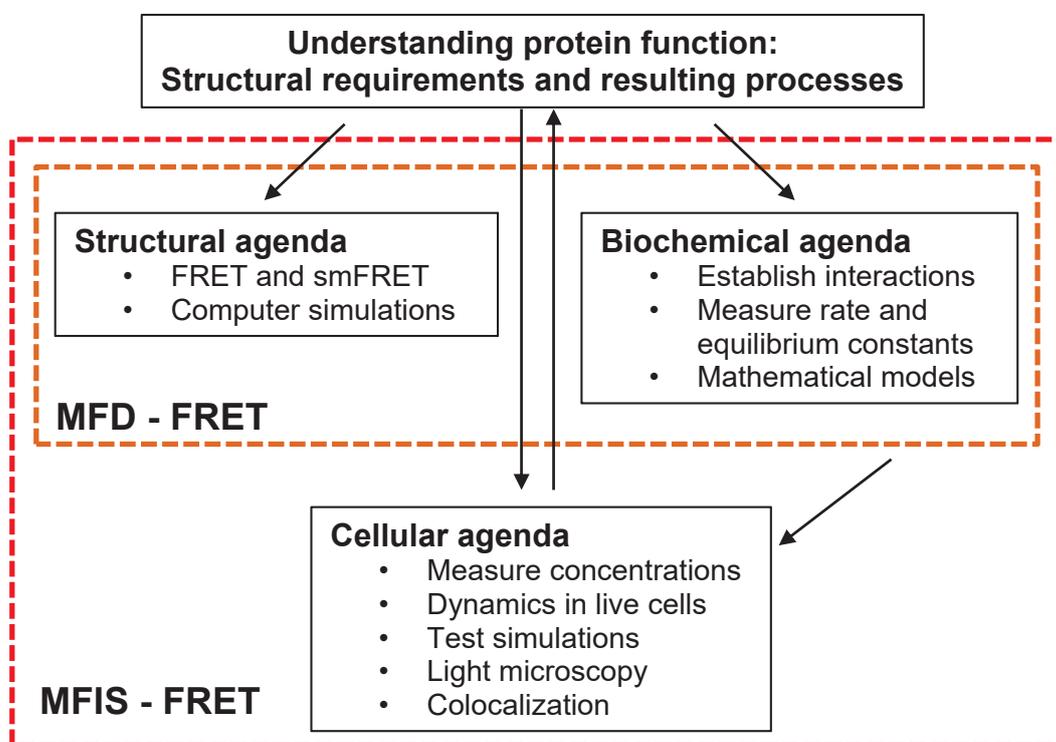
beiden Konformeren zu zeichnen. Die gefundenen Konformere des hGBP1 Monomers liefern Hinweise auf einen möglichen Reaktionspfad der GBP-Oligomerisierung. Die Ergebnisse der hGBP1 Studien wurden gestützt durch Kleinwinkel-Röntgenstreuung, Neutronen-Spin Echo Spektroskopie (Andreas Stadler, FZ-Jülich) und durch Elektronenspinresonanz-Messungen (Johann Klare, Universität Osnabrück). Zusammen zeigen die entwickelten Methoden und Ergebnisse, wie Fluoreszenz und FRET-Messungen unser Verständnis von Struktur und Dynamik von Biomolekülen erweitert.

Um Rückschlüsse über ein System durch Fluoreszenzmessungen zu ziehen sind dedizierte Modelle zur Beschreibung der experimentellen Observablen erforderlich, welche über eine rein formale Beschreibung des Experimentes hinausgeht. So muss beispielsweise in einem FRET-Experiment die räumliche Verteilung der Fluorophore berücksichtigt werden, um die beobachteten Fluoreszenzintensitäten korrekt zu interpretieren. Weiter erschwert Dynamik die Interpretation der Observablen signifikant, da die Zeitskalen von Pikosekunden bis zu Stunden berücksichtigt werden müssen. Um ein System über mehrere zeitliche Dekaden zu beschreiben und Rückschlüsse auf das beobachtete System treffen zu können müssen Modelle verschiedener Komplexität miteinander kombiniert werden. In dieser Arbeit wurden u.a. die Vorwärtsmodellierungen von chemischen Gleichgewichten, grobkörnige molekulare Monte-Carlo Simulationen (Kapitel D), sowie Simulationen der Brownschen Diffusion von Farbstoffen (Kapitel A) kombiniert.

## Summary

### Aims

To understand protein function, knowledge on their structural features, biochemical properties and their role in the cellular context is essential. For instance, in mammalian cells the expression of a protein class studied in this work, the guanylate binding proteins (GBPs), is triggered by interferon-gamma. This protein class belongs to the family of dynamin like proteins. It is involved in the immune response to intra-cellular pathogens. In presence of guanosine triphosphate (GTP) GBPs form oligomers. Thus, the formation of these complexes depends on the GBP- and GTP-concentration. Moreover, GBPs undergo reversible conformational changes upon GTP-hydrolysis, which allow them to attach to membranes. By this mechanism murine GBPs attacks cellular parasite such as *Toxoplasma gondii*.



**Figure 1.** Diagram of the reductionist-synthetic strategy to characterize the role of proteins in biological systems. MFD-FRET measurements in combination with computer simulations allow us to address structural and biochemical aspects of protein function. MFIS-FRET addresses additionally cellular aspects by live-cell imaging. (adapted from (Pollard 2013, "No Question about Exciting Questions in Cell Biology."))

This brief overview on GBPs clarifies why structural, biochemical and cellular aspects have to be studied together to obtain a full picture on protein function (see Figure 1). By omitting a single aspect, the function of a protein might remain elusive.

The function of proteins is directly related to their structure and dynamics. Analysis of protein kinetics is highly challenging, as their relevant kinetics covers several order of magnitudes: a

## Summary

---

few picoseconds, in case of side-chain rotations, up to several seconds, in case of large conformational changes. Therefore, analytical methods with high temporal resolution are required. At the same time, distances have to be measured with Angström accuracy, to be able to develop molecular models. Additionally, microscopic techniques are required to keep track of the cellular context in living cells. Finally, to allow for experiments in living cells as rigorous as in a test tube, highly selective and sensitive analytical techniques are required.

In this context, fluorescence spectroscopy, particular in combination with microscopic imaging, is a highly promising and not yet fully assessed analytical technique: Fluorescence detection assures high selectivity and sensitivity down to the single-molecule level. Multi-parameter fluorescence detection (MFD) promises high time-resolution, as fluorescence photons are detected with picosecond resolution and Förster resonance energy transfer (FRET) can be used to measure distances within biomolecules and biomolecular assemblies with Angström accuracy. Therefore, the combination of MFD and FRET (MFD-FRET), when united by computational techniques, which interpret the spectroscopic data in terms of structural models, can be used to study biochemical, structural and dynamical aspects of proteins (see Figure 1). To address cellular aspects, the combination of MFD and microscopic imaging (multi-parameter image spectroscopy, MFIS) offers unique possibilities function, as structural-, biochemical- and cellular-aspects are consolidated in a single measurement (see Figure 1). This allows to address the multifactorial problems to understand protein function by a single measurement.

To achieve the aim of understanding protein function under physiological conditions in living cells, several key aspects have to be addressed. Equilibrium constants have to be quantified, binding partners have to be identified, conformational dynamics has to be related to protein function, structural models have to be constructed by quantitative measurements, and finally these observations have to be put into context, meaning the biological relevance of the findings. Currently, *in vitro* MFD-measurements achieve a significant higher resolution as compared to MFIS-measurements in living cells. Hence, measurements in living cells are best corroborated by additional *in vitro* ensemble and single-molecule measurements under defined conditions.

## Summary

To solve this task a whole set of new analysis techniques and methods was established for *in vitro* and live cell measurements. This was necessary as FRET is rarely used for quantitative structural modelling. Hence, first a method, which culminates multiple FRET-measurements

into a quantitative structural model, was developed and tested on a *static* protein nucleic acid complex (Supplement A). Later, methods to study *dynamic* protein structures were established. A method to analyze fluorescence intensity decays with high resolution was developed which is able to resolve structural heterogeneities of conformational ensembles by FRET (Chapter A). This technique requires reference samples of high quality, which can be obtained as described in Chapter B. Next, the meaning of the fluorescence observables were clarified. This allows to reliably detect conformational dynamics by MFD single-molecule measurements and to describe multi-dimensional FRET-efficiency histograms (Chapter C). Finally, methods equilibrium constants in living cells are quantified by MFIS of the G-protein coupled receptor TGR5 and GBPs (chapter D).

Having this toolkit at hand, *in vitro* measurements on the model system T4 lysozyme (T4L) were analyzed to determine structural models of a transiently populated, short lived (hidden state) conformational state. FRET-guided molecular simulations and cross-validation of the applied fluorescence decay model achieved this. Next, guanylate binding proteins (GBPs) were studied by *in vitro* and live cell measurements. First, the dimerization of the human guanylate binding protein 1 (hGBP1) was investigated by ensemble FRET and EPR-measurements (Supplement B). Here a new binding mode was found and a previously postulated binding mode was confirmed. This structural model together with the methods presented in Chapter D allowed to study the homo- and hetero-oligomerization of GBPs in living cells and determine equilibrium constants (Supplement C). Finally, the free hGBP1 was examined by ensemble and single-molecule measurements and its solution structure given by two dynamic exchanging conformational states was solved. The kinetics of transitions between these states were quantified by lifetime filtered FCS and mapped to structural models. This conformational states might explain the previously found dimerization modes. The hGBP1 study was corroborated by small angle X-ray scattering (SAXS), neutron-spin echo spectroscopy (NSE) (Andreas Stadler, FZ-Jülich) and electron paramagnetic resonance measurements (EPR) (Johann Klare, Universität Osnabrück). Overall these studies demonstrate how fluorescence and FRET-measurements may improve our understanding on the movement and mechanism in molecular machines.

### Abstract of Chapters

#### **A – A combined graphical and analytical method for accurate analysis of time-resolved FRET measurements: FRET between tethered fluorophores in presence of local quenching**

Fluorescence intensity decay curves can be interpreted in terms of inter-fluorophore distance distributions by a component analysis. The resolution of this method is bound to the noise level of the recorded fluorescence intensity decay and not the detection count-rate. Hence, conformers can be distinguished even in presence of fast exchange kinetics. Distances are obtained by measuring the time-resolved fluorescence intensity of the donor in presence and absence of FRET. To facilitate the interpretation of the experimental data we introduce an intensity independent quantifier for FRET (the FRET-induced donor decay). The FRET-induced donor decay allows to interpret fluorescence intensity decays visually in terms of FRET-rate constants and distances. We generalize this method to systems of multiple donor- and multiple FRET-states. Next, we study systematic errors that may arise due to fluorescence quenching of flexible donor by its local environment by simulating dynamic quenching of the donor and quenching by FRET. Our simulations demonstrate that the expected relative error of the mean distance within a protein conformer is approximately 2 % if correlations between FRET and local quenching are not considered explicitly. Finally, we study statistical and systematic errors. We demonstrate that systematic errors due to incorrect donor-references typically outweigh statistical errors in particular for long separation distances.

#### **B – Experimental reference samples for accurate time-resolved FRET analysis: Handling positional effects of fluorescent dyes in bis-functional proteins**

In bis-functional proteins with identically labeling chemistry of the donor and the acceptor fluorophore, reference samples for FRET-measurements are not easily obtained. Without an appropriate reference sample, systematic errors of the recovered donor-acceptor distances are anticipated. We address this problem by different analytical and preparative methods. We label a set of FRET-samples either first with a donor- or by an acceptor-fluorophore. In both cases the distribution of the fluorophores among the labeling-sites is unknown. However, if the sample was labeled first by the acceptor, it is inevitable to determine the fluorophore distribution among the labeling sites by two additional reference samples with precisely known donor positions. In a FRET-sample, which was labeled first by a donor, this inconvenience can be circumvented as an experimental representative donor reference can be obtained. For both cases we present a set of analytical methods which is capable of determining the distribution of the fluorophores among the labeling sites. Knowing this

distribution, the two experimental donor-reference decays are combined to a representative donor reference decay. Additionally, we demonstrate by simulations of weakly interacting fluorophores that the FRET-rate constant distribution is basically independent of the distribution of the donor- and the acceptor-fluorophores among the labeling sites. This is particularly the case if both donor and acceptor are tethered to the biomolecule with linkers of comparable length.

### **C – A generalized description of multidimensional single-molecule FRET histograms: Theoretical treatment of FRET parameters of multistate dynamics systems**

Multidimensional single molecule Forster Resonance Energy Transfer (smFRET) experiments are rich on information regarding biomolecular kinetics from subnanosecond timescale to timescales of the burst integration in milliseconds. Here, we present a rigorous analytical treatment of the time dependent FRET efficiency distributions to find parametric relationships between the FRET indicators measured by time resolved and intensity based methods connecting the fast dynamics in the sub microseconds domain to the slower dynamic events happening within the burst duration. Moreover, we show that these relationships represent the “FRET lines” used in analysis of smFRET experiments data. Mathematically, these FRET lines can be described as the relationship between the first two moments on the lifetime distribution. These moments are used to define the “static” and the “dynamic” relationships useful in identifying exchange processes that occur at timescales between nanoseconds and milliseconds. Moreover, these FRET lines include most experimental parameters such as detection efficiencies, quantum yield and the linker dynamics of the fluorophores. Finally, we provide an example where FRET lines are indispensable for solving complex dynamic schemes.

### **D – Quantitative FRET-analysis of measurements in live cells**

In MFIS mainly fluorescent proteins (FPs) instead of small organic fluorophores are used. The use of FPs has severe implications for data-analysis as well as the general applicability of MFIS-measurements. With FPs the dynamic  $\kappa^2$ -approximation does not apply. Therefore, for every donor-acceptor distance a distribution of orientation factors has to be considered. Additionally, the linkers used to couple FPs to the host molecules are at least 5 times longer than the ones used in single-molecule studies. Thus, a correct averaging over the sterically allowed space is crucial. We consider these effects by dedicated Monte-Carlo (MC) simulations, which sample the sterically allowed conformational space of the FPs. The MC simulations explicitly consider orientation effects and the entropic effects of the dye-linker. It

is shown how the results of these simulations can be converted into time-resolved FRET-observables for homo- and hetero-FRET.

Additionally, two distinct methodologies to determine equilibrium constants by live-cell measurements are presented. The first methodology relies on a formal description and quantification of multiple FRET-measurements, which are later interpreted in terms of equilibrium constants, once a structural model is suggested and concentrations were determined. In the second approach conformational ensembles are generated by molecular simulations. Later fluorescence decays are interpreted in terms of species fraction of the respective simulations. These species fractions are used to determine equilibrium constants. The first approach was applied to live-cell measurements on the G-protein coupled receptor TGR5. The later approach was applied to the murine guanylate binding proteins. The advantages and disadvantages of both approaches are briefly discussed.

### **E – Refining fluorescence tools to resolve hidden conformational states of T4L by FRET**

In the frequently studied model system T4 lysozyme (T4L), we recently found a new transiently populated conformational state (hidden state) within its catalytic cycle. Some key findings, which lead to the discovery of this “hidden” state, are summarized. A formal description of the fluorescence decay model is based on discrete FRET-rate constants and was used to corroborate the finding of the “hidden” state. The shortcomings of the formal description to determine structural models by FRET are discussed. In particular, the effect of spatial dye distributions and the effect of dynamic quenching by the local environment are addressed.

A hybrid approach is presented which welds FRET-guided molecular dynamics simulations, simulations of dynamic donor quenching and adapted fluorescence decay models together. After generating structures by FRET-guided simulations, the determined structural models of the conformers are tested for their ability to quench the donor dye. This is accomplished by simulating the donor diffusion and quenching by aromatic amino acids. This allows to cross-validate the initially used fluorescence decay model against the obtained structures. In T4L, the simulated dynamic quenching of the donor showed only minor differences among the found conformers. This validated our determined distances and enabled us to resolve structural models of a short-lived conformational state.

## **F – Mapping kinetics and structure of a large GTPase resolves a hidden state essential for oligomerization**

Large GTPases or guanylate binding proteins (GBPs) are involved in the immune response against intracellular pathogens. We determined the solution structure of a representative of this protein class (the human guanylate binding protein 1, hGBP1) by small angle X-ray scattering (SAXS), electron paramagnetic resonance measurements (EPR) and Förster resonance energy transfer (FRET). Large GTPases are composed out of three different domains: a nucleotide binding LG-domain, a middle domain, and a helical domain, here addressed as helices  $\alpha 12/\alpha 13$ . We find that hGBP1 adopts two distinct conformations. In one conformer the C-terminal helices  $\alpha 12/\alpha 13$ , which are important for oligomerization, are bound to the “left” side of the LG-domain. In the other, they are bound to the “right” side of the LG-domain. This conformational transition is associated with a complex rearrangement within the middle-domain.

To study the conformational dynamics of hGBP1 we combine NSE and lifetime filtered species cross-correlation (fFCS). By this unique combination of methods, we fully characterize the relaxation time distribution from nanoseconds to milliseconds. The dynamical amplitude determined by NSE in the nanosecond regime is small. This unambiguously demonstrates that the conformational kinetics is fully captured by fFCS and happens in the microsecond to millisecond regime. The determined amplitude distribution of the relaxation time spectrum suggests that the conformational changes happen on a rugged energy landscape. By combining multiple FRET-measurements we outline a distinct kinetic pattern of the protein and map the flexibility to protein function. We find that the C-terminal helices  $\alpha 12/\alpha 13$  are highly flexible relatively to the middle-domain of hGBP1. This finding is notably as we previously showed that helix  $\alpha 13$  associates in the dimeric-form of the protein (see Supplement B). However, if the protein remains in its major state at room temperature, an association of the helices  $\alpha 13$  is sterically impossible. The presence of a second constantly exposed interface may be of general relevance to explain differences in the oligomerization equilibria among different GBP family members.

## **Supplemental papers**

### **A – A toolkit and benchmark study for FRET-restrained high-precision structural modeling**

We first established a toolkit that allows to interpret FRET-measurements on biomolecules and their complexes in terms of structural models. FRET provides inter-fluorophore distance information in the range of 20 to 80 Å. However, as the fluorophores are coupled by long

## Summary

---

linkers ( $\sim 20$  Å), no direct structural information is obtained. As the precise position of the fluorophores is unknown, FRET is often considered a low-resolution technique with errors in the order of the linker-length. Nevertheless, we demonstrate that accurate structural models can be obtained if the spatial distribution of the fluorophores is explicitly considered. This dramatically improves the precision of FRET-derived structures. We determine the precision and the confidence levels of the structural models by rigorous error estimation and show the accuracy of this method by docking a DNA primer-template to HIV-1 reverse transcriptase. We obtained FRET-derived models that agree with the known X-ray structure with a root mean squared deviation of 0.5 Å. Additionally, we introduced a method to ‘screen’ a large conformational ensemble for instance generated by molecular dynamics simulations. By this hybrid approach, we determine a formerly unknown configuration of the flexible single-strand template overhang.

### **B – Triphosphate Induced Dimerization of Human Guanylate Binding Protein 1 Involves Association of the C-Terminal Helices: A Joint Double Electron-Electron Resonance and FRET Study**

The human guanylate binding protein 1 (hGBP1) is a member of the subclass of the dynamin superfamily. During GTP hydrolysis, the protein undergoes structural changes leading to self-assembly. Previous studies have suggested dimerization of the protein by means of its large GTPase (LG) domain and significant conformational changes in helical regions near the LG domain and at its C-terminus. We applied intra-molecular time-resolved FRET measurements together in combination with pulsed electron paramagnetic resonance and time-resolved fluorescence spectroscopy for structural investigations on hGBP1 dimerization and conformational changes of its C-terminal helix  $\alpha 13$ . Consistent distance measurements by double electron–electron resonance (DEER, also named pulse double electron resonance = PELDOR) spectroscopy and Förster resonance energy transfer (FRET) measurements using model-free analysis approaches revealed a close interaction of the two  $\alpha 13$  helices in the hGBP1 dimer formed upon binding of the non-hydrolysable nucleoside triphosphate derivate GppNHp. In molecular dynamics (MD) simulations, these two helices form a stable dimer in solution. Our data show that dimer formation of hGBP1 involves multiple spatially distant regions of the protein, namely, the N-terminal LG domain and the C-terminal helices  $\alpha 13$ . The contacts formed between the two  $\alpha 13$  helices and the resulting juxtaposition are expected to be a key step for the physiological membrane localization of hGBP1 through the farnesyl groups attached to the end of  $\alpha 13$ .

### **C – Guanylate binding proteins (GBPs) directly attack *T. gondii* via supramolecular complexes**

Guanylate binding proteins (GBPs) are key players in the immune response against intracellular pathogens. To gain molecular insights on how GBPs control the replication of the cellular parasite *T. gondii* in mice, we performed mutational analyses and in live-cell MFIS FRET-measurements. MFIS combines microscopy and spectroscopy in a single measurement. This allows for the correlation of structural, biochemical and cellular aspects. We were able to determine spatially resolved intracellular concentrations of GBP's via the fluorescence intensity. Additionally, we determined structural aspects on the cellular- as well as the molecular level by monitoring the mGBP co-localization and their interaction monitored by FRET. By correlating this information, we were able to quantify the equilibrium constants between different GBPs and to monitor the dynamics in live cell.

We show that the control of *T. gondii* replication by mGBP2 requires GTP hydrolysis and isoprenylation, thus, enabling reversible oligomerization in vesicle-like structures. mGBPs reside in at least two discrete subcellular reservoirs and attack the parasitophorous vacuole membrane (PVM) as orchestrated, supramolecular complexes forming large, densely packed multimers comprising up to several thousand monomers. This dramatic mGBP enrichment results in the loss of PVM integrity, followed by a direct assault of mGBP2 on the plasma membrane of the parasite. We studied biochemical aspects of this attack by quantifying the homo- and hetero-oligomerization of mGBP2, indispensable in the host response, with mGBP1, mGBP3 and mGBP6. We find dimer formation with mGBP2, mGBP1 and mGBP3 but not with mGBP6. Furthermore, the tendency to form oligomers decreases in the following order: mGBP2, mGBP1 mGBP3. These findings provide important insights as they demonstrate that mGBPs are not only players in the control of *T. gondii* but directly attacking the parasitophorous vacuole membrane as a fine-tuned ensemble of molecules.

### **Conclusion**

To base conclusions on a system of interest on fluorescence observables, highly adapted models are required that go beyond a formal description. For instance, the spatial distributions of the fluorophore have to be considered to interpret the result of a fluorescence experiment. In static systems mainly the spatial distributions of the dyes are of relevance. These distributions can be for instance determined by accessible volume calculations, which mimic the conformational ensemble of fluorophores attached to a biomolecule. However, once dynamical systems are studied, the analysis of the fluorescence observables significantly complicates: The fluorescence lifetime of fluorophores is in the nanosecond regime. Dynamic

## Summary

---

quenching coupled to dye diffusion also happens in the nanosecond regime, while large scale diffusion of dye-linkers happens within hundreds of nanoseconds. The time-scales of conformational dynamics within proteins span from microseconds up to seconds, while the chemical processes occurring in living cells occur within seconds to hours. Hence, the time-scales, which have to be considered to understand all aspects of fluorescence, range from picoseconds to hours.

The origin of the observed fluorescence observables can only be understood by combining specifically adapted models. Forward modelling of chemical equilibria may be used to describe processes within cells. Accessible volume calculations describe reliably the spatial distribution of fluorophores attached to biomolecules. Brownian dynamics simulations provide insights on dynamic quenching of dyes. Coarse-grained Monte-Carlo simulations of fluorescence proteins provide the possibility to test models against fluorescence observables. Finally, molecular dynamics simulations provide the most detailed insights at high computational costs.

All these types of models have been developed and combined to relate measurable fluorescence observables to protein structure, kinetics and equilibrium constants. These models were evolved into analytical methods, which were applied to study protein function *in vitro* and in living cells.

## Chapter A: A combined graphical and analytical method for accurate analysis of time-resolved FRET measurements: FRET between tethered fluorophores in presence of local quenching

<b>1</b>	<b>Introduction.....</b>	<b>1</b>
<b>2</b>	<b>Models and Results.....</b>	<b>4</b>
	2.1 Time-resolved fluorescence.....	4
	2.1.1 Definitions.....	4
	2.1.2 Distance-dependence.....	6
	2.2.3 FRET induced donor decay.....	6
	2.2.4 Treating a reference with several partially quenched donor states.....	12
	2.2 Mobile dyes and FRET.....	15
	2.2.1 Concept.....	15
	2.2.2 Simulation of dynamic donor quenching.....	18
	2.2.3 Impact of dye diffusion on FRET.....	20
	2.2.4 Impact of dye diffusion and quenching on FRET.....	22
	2.3 Error-estimation.....	23
	2.3.1 Accuracy using a standard reference.....	23
	2.3.2 Accuracy with sample specific reference.....	25
	2.3.3 Statistical precision.....	26
<b>3</b>	<b>Conclusion.....</b>	<b>29</b>
<b>4</b>	<b>Supplement.....</b>	<b>30</b>
<b>5</b>	<b>References.....</b>	<b>31</b>

## Chapter B: Experimental reference samples for accurate time-resolved FRET analysis: handling positional effects of fluorescent dyes in bis-functional proteins

<b>1.</b>	<b>Introduction.....</b>	<b>37</b>
<b>2</b>	<b>Materials and Methods.....</b>	<b>40</b>
	2.1 Distances by FRET.....	40
	2.2 Systematic errors.....	41
	2.3 Accessible-volume simulations.....	42
	2.4 Recombinant protein expression and purification.....	43
	2.5 Labeling.....	44
	2.6 Enzymatic proteolysis.....	45
	2.7 MFD-measurements.....	45
	2.8 Anisotropy PDA.....	46
	2.9 FCS.....	47
	2.10 Fluorescence decay analysis.....	48
<b>3</b>	<b>Results and Discussion.....</b>	<b>50</b>
	3.1 Position dependencies of mobile dyes.....	50
	3.2 Labeling.....	52
	3.3 Determination of dye locations.....	55

## Table of contents

---

3.3.1 Enzymatic proteolysis .....	55
3.3.2 FCS, Anisotropy .....	56
3.3.3 Fluorescence decay analysis .....	58
3.3.4 Method comparison .....	59
4 Conclusion .....	60
5 Supplement .....	61
6 References .....	62

## Chapter C: A generalized description of multidimensional single-molecule FRET histograms

1 Introduction.....	65
1.1 Theoretical background: Förster theory .....	66
1.2 FRET efficiency of single- and multiple- species .....	67
1.2.1 Single species .....	67
1.2.2 Mixture of multiple species .....	69
1.2.3 Experimental observables .....	70
2 FRET-lines.....	73
2.1 FRET-lines for static host-molecules .....	74
2.1.1 Absence of dye-linker dynamics .....	75
2.1.2 Presence of dye-linker dynamics.....	75
2.2 FRET-lines for dynamic host-molecules.....	79
2.2.1 Absence of dye-linker dynamics .....	79
2.2.2 Presence of dye-linker dynamics.....	80
2.2.3 Dynamic FRET-lines and relation to the species fractions .....	81
2.4 Multistate systems.....	82
2.3 Alternative implementation of FRET-lines .....	84
2.3.1 Polynomial Approximation .....	84
2.3.2 Quadratic approximation .....	85
2.3.3 Alternative representation of FRET lines.....	86
2.3.4 FRET Lines and timescale of dynamics .....	88
3 Real case and complex FRET-lines .....	89
4 Conclusions .....	92
5 Materials and Methods.....	93
5.1 Multiparameter Fluorescence Detection.....	93
5.2 Single molecule simulator (Brownian Dynamics) .....	94
5.3 WLC Polymer model.....	94
5.4 IF3 production and labeling .....	95
6 Appendix .....	96
A: Polynomial approximation coefficients .....	96
B: Quadratic approximation (Analytical Solution) .....	96
B.1 Species averaged lifetime .....	96
B.2 Fluorescence averaged lifetime .....	97
7 References .....	98

## Chapter D: Quantitative FRET-analysis of measurements in live cells

<b>1</b>	<b>Introduction</b> .....	<b>101</b>
<b>2</b>	<b>GFPs and Structure</b> .....	<b>104</b>
	2.1 Dye-dimensions and orientation effects .....	104
	2.2 Weighting of distribution .....	106
	2.3 Monte-Carlo simulations.....	108
	2.4 Effect of orientation factor.....	110
	2.5 Quantifiers in MFIS .....	113
<b>3</b>	<b>Determination of association constants</b> .....	<b>115</b>
	3.1 Approaches to determine equilibrium constants .....	115
	3.2 Forward modelling of oligomer species .....	117
	3.4 Individual model free analysis .....	118
	3.3 Individual pattern analysis .....	122
<b>4</b>	<b>Conclusion</b> .....	<b>124</b>
<b>5</b>	<b>References</b> .....	<b>125</b>

## Chapter E: Refining fluorescence tools to resolve hidden conformational states of T4L by FRET

<b>1</b>	<b>Introduction</b> .....	<b>127</b>
<b>2</b>	<b>Identification of a hidden state</b> .....	<b>131</b>
	2.1 Formal fluorescence decay FRET-analysis .....	131
	2.2 Detection of a hidden state .....	133
	2.3 Biochemical trapping.....	134
	2.4 Limitations of formal FRET-analysis .....	135
<b>3</b>	<b>Refined fluorescence tools for structure validation</b> .....	<b>136</b>
	3.1 Global analysis of fluorescence decays.....	136
	3.2 Considering FRET and fluorescence quenching in structural modelling .....	138
	3.3 Structural model validation .....	140
<b>4</b>	<b>Conclusion</b> .....	<b>143</b>
<b>5</b>	<b>References</b> .....	<b>144</b>

## Chapter F: Mapping kinetics and structure of a large GTPase resolves a hidden state essential for oligomerization

<b>1</b>	<b>Introduction .....</b>	<b>147</b>
<b>2</b>	<b>Material and Methods.....</b>	<b>150</b>
2.1	Protein expression and labeling.....	150
2.2	FRET .....	151
2.2.1	Experimental methods .....	151
2.2.2	Analysis methods.....	152
2.3	Simulations.....	158
2.3.1	Rigid-body docking.....	158
2.3.2	Normal mode based geometric simulations .....	159
2.3.3	Principle component analysis.....	159
2.3.4	Accessible volume simulations and FPS.....	160
<b>3</b>	<b>Results .....</b>	<b>161</b>
3.1	Experimental equilibrium distributions .....	161
3.2	Identification and quantification of molecular kinetics .....	164
3.3	Essential motions determined by molecular dynamics simulations.....	169
3.4	Structural modelling by experimental constraints .....	170
<b>4</b>	<b>Discussion.....</b>	<b>174</b>
<b>5</b>	<b>Supplement.....</b>	<b>178</b>
5.1	Principle component analysis .....	178
5.2	Rigid body docking .....	178
5.3	MFD-Histograms .....	181
5.4	Sub-Ensemble TCPC.....	187
5.5	Ensemble TCPC.....	190
5.6	Filtered FCS .....	191
5.7	Error estimation.....	194
5.8	EPR-measurements .....	195
5.9	Parametrization of MTSSL for Accessible volume calculations .....	196
<b>6</b>	<b>References.....</b>	<b>196</b>

### Supplements

- A**    **A toolkit and benchmark study for FRET-restrained high-precision structural modeling**
- B**    **Triphosphate induced dimerization of human guanylate binding protein 1 involves association of the C-terminal helices: A joint double electron-electron resonance and FRET study**
- C**    **Guanylate binding proteins (GBPs) directly attack *T. gondii* via supramolecular complexes**

## **Chapter A - A combined graphical and analytical method for accurate analysis of time-resolved FRET measurements: FRET between tethered fluorophores in presence of local quenching**

---

**Thomas-Otavio Peulen\***, **Oleg Opanasyuk**, **Claus A.M. Seidel<sup>§</sup>**

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

### **1 Introduction**

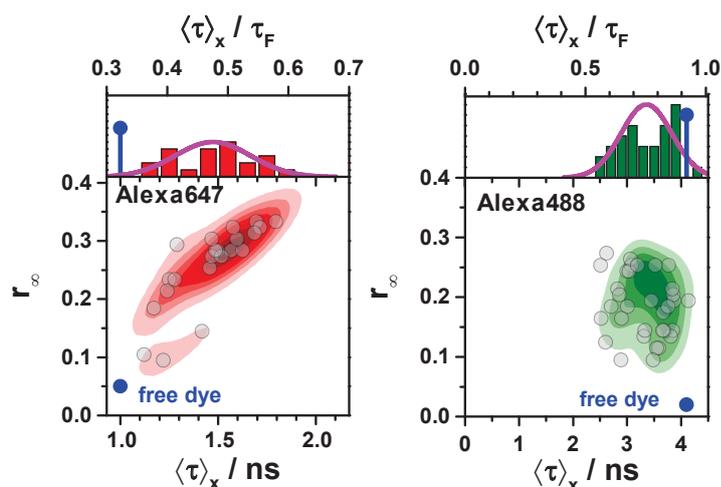
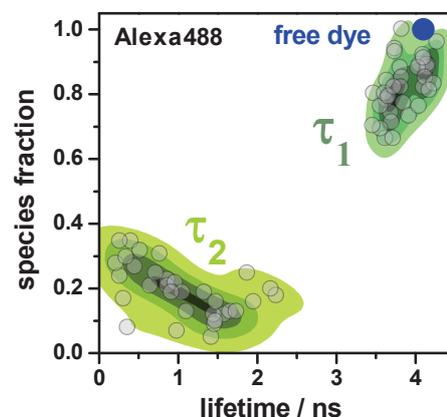
Fluorescence measurements by time-resolved techniques, such as multi-parameter fluorescence detection (MFD), which preserve the spectral emission window and the polarization of all detected photons with picosecond time-resolution (Widengren et al., 2006), provide a wealth of information, in particular when combined with pulse interleaved excitation (PIE) (Kudryavtsev et al., 2012). Given the experimental recordings a whole set of parameters, such as the fluorescence intensities, the fluorescence anisotropy and the fluorescence lifetime, can be determined by efficient estimators even for single molecules (Schaffer et al., 1999, Maus et al., 2001). Among those parameters the fluorescence lifetime is the most informative, as the excited state may serve as clock to study the dynamics of diffusion governed reactions and short lived excited states (Rice, 1985). The fluorescence lifetime is altered by short- and long-range interactions. Short-ranged collisional quenching reports on the local environment of the dye while long-range dipolar coupling, i.e. Förster resonance energy transfer (FRET), reports distances donor and an acceptor fluorophores (Doose et al., 2009, Stryer and Haugland, 1967). As the distance range with the highest sensitivity in FRET (20-80 Å) matches the length scale of biomolecules, FRET-measurements are an ideally tool to study biomolecules under physiological conditions (Vöpel et al., 2014), in living cells (Sustarsic and Kapanidis, 2015, Somssich et al., 2015, Stahl et al., 2013, Weidtkamp-Peters et al., 2009). As optical method FRET can be combined with force spectroscopy (Gaiduk et al., 2007) and given the current progress in super-resolution microscopy (Hell and Wichmann, 1994, Rust et al., 2006, Betzig et al., 2006, Dickson et al., 1997) FRET has tremendous potential to resolve molecular structures in live-cell to capture biomolecules in action. Typically, FRET is quantified by the steady-state transfer efficiency  $E$ , which measures the yield of the FRET-process. The steady-state transfer efficiency is an extremely robust quantifier of the FRET-process, easily and straight forward determined, and thus applied frequently to grasp structural features of biomolecule by ensemble- (Clegg et al., 1994, Clegg et al., 1992, Mekler et al., 2002) and single-molecule measurements (Rothwell et al., 2013, Kalinin et al., 2012, Wozniak et al., 2008, Andrecka et al., 2009). Being a relative quantity, the FRET-efficiency is highly valuable, because it is independent of absolute intensities and instrumental parameters. It is determined using steady-state fluorescence intensities  $F$  which are cor-

rected for the spectral sensitivity of the instrument and sample specific background fluorescence. To determine  $E$ , either the relative changes of the donor fluorescence intensity in absence  $F_{D|D}^{D0}$  and presence of an acceptor  $F_{D|D}^{DA}$  (eq. 1a), or the FRET-sensitized acceptor emission  $F_{A|D}^{DA}$  (eq. 1b), using the total corrected intensity of the donor and acceptor  $F_{A|D}^{DA} + \gamma' \cdot F_{D|D}^{DA}$  as a reference, are monitored:

$$E = \frac{F_{D|D}^{D0} - F_{D|D}^{DA}}{F_{D|D}^{D0}} = \frac{F_{A|D}^{DA}}{F_{A|D}^{DA} + \gamma' \cdot F_{D|D}^{DA}} \quad \text{with} \quad \gamma' = \frac{\Phi_F^{A0}}{\Phi_F^{D0}} \quad (1)$$

Herein,  $\Phi_F^{D0}$  and  $\Phi_F^{A0}$  are the fluorescence quantum yields of the donor and acceptor, respectively. The sub-scripts indicate the detection and excitation channels and the super-scripts specify the sample. For instance, the sub-script  $D|D$  symbolizes donor detection ( $D|D$ ) given donor excitation ( $D|D$ ) while  $A|D$  corresponds to acceptor detection ( $A|D$ ) given donor excitation ( $A|D$ ). The superscript  $DA$  represents a FRET-sample, while  $D0$  and  $A0$  correspond to samples containing solely a donor and acceptor fluorophore, respectively. Thus, with known quantum yields and calibrated experimental setups absolute transfer efficiencies  $E$  can be determined (Sisamakris et al., 2010).

The fluorescence properties of the fluorophores commonly used in FRET studies are sample depend. To distinguish FRET from other deactivation pathways processes, sample specific references compulsory. If the transfer efficiency is determined by the change of the donor fluorescence intensity, a separate sample without acceptor  $F_{D|D}^{D0}$  is needed as a reference (eq. 1a). Alternatively, the calibration factor  $\gamma'$  has to be determined (eq. 1b). Differences of the fluorescence properties among samples are caused by the local environment of the dyes, i.e. the excited state of xanthenes is quenched by the sidechains of aromatic amino acid by photo-induced electron transfer (PET) (Neubauer et al., 2007, Seidel et al., 1996, Eggeling et al., 1998, Doose et al., 2009) while the fluorescence of cyanines is influenced by solvent effects and steric constraints (Sauerwein et al., 1992, Buschmann et al., 2003, Widengren and Schwille, 2000, Levitus and Ranjit, 2011, Chibisov et al., 1995, Widengren et al., 2001). For Alexa647N (a cyanine dye) this dependence results in a positive correlation between the residual anisotropy and the species averaged lifetime (see Figure 1A). As changes of the microenvironment of cyanines alter their brightness, they cause broadening of single-molecule FRET-efficiency histograms (Kalinin et al., 2008). This can be utilized to sense interactions of proteins and nucleic acids using a single fluorophore as reporter (Hwang and Myong, 2014).

**A Average lifetimes and residual anisotropies**

**B Fluorescence lifetime distributions**


**Figure 1.** Fluorescence lifetimes  $\tau$  and residual anisotropies  $r_\infty$  of the fluorophores Alexa647 and Alexa488 attached via maleimide or hydroxylamine chemistry to different amino acids of various proteins (human guanylate binding protein 1, T4 lysozyme, Postsynaptic density protein 95, lipase foldase of *Pseudomonas aeruginosa* and the cyclin-dependent kinase inhibitor 1B). **(A)** For each sample the species weighted averaged lifetimes  $\langle \tau \rangle_x$  and  $r_\infty$  are shown as dots overlaid by a Gaussian kernel density estimation (Sheather and Jones, 1991). Using radiative lifetimes of  $\tau_F = 3.1$  ns and  $\tau_F = 4.5$  ns for Alexa647 (Klehs et al., 2014) and Alexa488, respectively, the relative brightness  $\langle \tau \rangle_x / \tau_F$  of Alexa647 and Alexa488 species were calculated. The average of all Alexa647 and Alexa488 samples are  $\langle \tau \rangle_x / \tau_F = 0.43 \pm 0.07$  and  $\langle \tau \rangle_x / \tau_F = 0.76 \pm 0.11$ , respectively. The average residual anisotropy of Alexa647 and Alexa488 of all samples is  $\langle r_\infty \rangle = 0.25 \pm 0.07$  and  $\langle r_\infty \rangle = 0.18 \pm 0.05$ , respectively **(C)** The fluorescence intensity decays of the Alexa488 samples were formally resolved into two components  $\tau_1$  and  $\tau_2$  with the respective fractions  $x_1$  and  $x_2 = 1 - x_1$ . For each sample the lifetimes and fractions are shown as open circles overlaid with a Gaussian kernel density estimation (green). The average lifetimes of the populations are  $\tau_1 = 3.9 \pm 0.2$  ns and  $\tau_2 = 1.0 \pm 0.5$  ns with species fractions of  $x_1 = 0.8 \pm 0.1$  and  $x_2 = 0.2 \pm 0.1$ , respectively.

As the transfer-efficiency only reports on the average yield of the FRET-process, heterogeneous mixtures are not resolved. This limitation is overcome by measuring the transfer-efficiency of single molecules. In single-molecule measurements the count rate limits the experimental resolution, as interconverting conformational states are not resolved if the time-scale of dynamics is shorter than the integration time (Gopich and Szabo, 2003). Hence, if the mean transfer efficiencies are used conformational heterogeneities may be concealed. Time-resolved fluorescence measurements overcome this limitation by recording averaged fluorescence intensity decay curves, as opposed to average transfer efficiencies. It is well-known that the fluorescence decays can be utilized to resolve conformational distributions and multiple models has been developed to resolve distance distributions and fast protein dynamics (van der Meer et al., 1993, Haas et al., 1975, Möglich et al., 2006). Reliable commercial instrumentation ensure highly reproducible experiments while fluorescence lifetime imaging allows to answer biological questions in living cells (Sun et al., 2011, Liptenok et al., 2014, Somssich et al., 2015). Unfortunately, the capabilities of the technique are unappreciated and data is misinterpreted. For instance, lifetime components obtained by formal analysis are interpreted in terms of distances.

In this paper we introduce a graphical representation for time-resolved FRET. By this and the presented equations we hope to facilitate the analysis of time-resolved fluorescence decays. The equations are founded current insights on the dynamics of fluorophore tethered to biomolecules and extensive simulations. Furthermore, we define clear limits set by statistical and systematic errors to provide confidence on statements based on the analysis results of their time-resolved fluorescence decays.

## **2 Models and Results**

### **2.1 Time-resolved fluorescence**

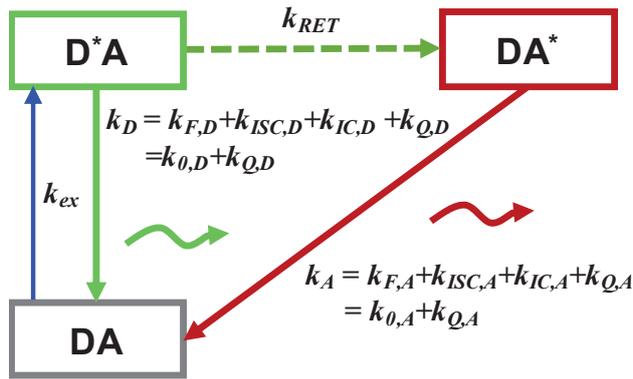
#### **2.1.1 Definitions**

Time-resolved techniques such as time-correlated single photon counting (TCSPC) consider the number of detected photons and their time of detection  $t$  with respect to an excitation pulse. By analyzing time-resolved fluorescence intensities  $f(t)$ , sample heterogeneities are resolved. This allows for instance to draw conclusions on protein dynamics (Möglich et al., 2006). The time-resolved fluorescence intensity  $f(t)$  is proportional to the number of molecules in an excited fluorescent state  $N^*(t)$  and to the radiative rate constant of fluorescence  $k_F$ :

$$f(t) = k_F \cdot N^*(t) \tag{2}$$

To analyze signal intensities, the fluorescence quantum yields and detection efficiencies of the experimental setup have to be considered. However, as usually only the decay shapes are analyzed, proportionality factors between  $N^*(t)$  and  $f(t)$  such as  $k_F$  are often omitted. In the global analysis of the donor  $f_{AD}^{DA}(t)$  and the FRET-sensitized acceptor emission  $f_{DA}^{DA}(t)$  such factors are essential to obtain properly scaled relative intensities (Beutler et al., 2008).

At low excitation the coupling between D and A can be described by a rate scheme as depicted in Figure 2 and D and A are excited at most by a single photon. Both, D and A are either in their ground state (D, A) or excited state ( $D^*$ ,  $A^*$ ). Thus, the DA-pair can be in one out of three distinct states: DA,  $D^*A$ ,  $DA^*$ .



**Figure 2.** Definition of states and rate constants of a system composed of a single donor (D) and acceptor (A) excited by a single photon. The star (\*) indicates an excited fluorophore: D\*A (excited donor, ground state acceptor), DA\* (excited acceptor, ground state donor), DA (ground state donor, ground state acceptor).  $k_{ex}$  is the rate constant of excitation,  $k_D$  and  $k_A$  are the rate constants of deactivation of the excited donor and acceptor state,  $k_{RET}$  is the rate constant of energy transfer from D to A.  $k_D$  and  $k_A$  are the sum of the respective radiative rate constant of fluorescence  $k_F$ , internal conversion  $k_{IC}$ , intersystem crossing  $k_{ISC}$  and the quenching rate constant  $k_Q$ .  $k_Q$  depends on the local environments of the dyes.  $k_F$ ,  $k_{ISC}$  and  $k_{IC}$  are dye specific and joined in the constants  $k_0$ .

The time-dependent population of these states is determined by the exchange rate-constants among the states. Here,  $k_{RET}$  is rate constant of energy transfer from D to A and the rate constants  $k_D$  and  $k_A$  are the total deactivation rate constant of D and A, respectively. These constants are given by the sum of all rate constants deactivating the fluorescent state: internal conversion  $k_{IC}$ , inter-system crossing  $k_{ISC}$ , radiative rate constant  $k_F$  of the respective fluorophore. The rate constant assigned to other environmental deactivation processes is summarized by  $k_Q$ . The probability of finding a molecule in a given state is obtained by solving the rate matrix associated to Figure 2. Usually, the rate of excitation  $k_{ex}$  is significantly smaller  $k_D$  and  $k_A$ . Thus, the ground state DA does not have to be considered and the change of fluorescent states is given by the following system of differential equations:

$$N_0 \cdot \frac{d}{dt} \begin{bmatrix} n_{D^*A} \\ n_{DA^*} \end{bmatrix} = N_0 \cdot \begin{bmatrix} -(k_D + k_{RET}) & 0 \\ k_{RET} & -k_A \end{bmatrix} \begin{bmatrix} n_{D^*A} \\ n_{DA^*} \end{bmatrix} \quad (3)$$

The time-dependent state populations of the excited states are obtained by solving this equation system. A solution for  $N_0$  initially excited donor molecules is given by:

$$N_{D^*A}(t) = N_0 \cdot e^{-(k_{RET}+k_D)t}$$

$$N_{DA^*}(t) = N_0 \cdot \frac{k_{RET}}{k_A - k_D - k_{RET}} \left( e^{-(k_{RET}+k_D)t} - e^{-k_A t} \right) \quad (4)$$

By combining equation (4) and equation (2) time-resolved fluorescence intensities of D and A are obtained. To go beyond formal description by rate-constants, the rate constants have to be related to system properties, i.e. rotation, intramolecular distances and motions (van der Meer et al., 1993). As experiments are usually insufficient to resolve complex, often reduced models

are applied. For instance, experimental decays are described by the means of DA-distance distribution (Gryczynski et al., 1988).

### **2.1.2 Distance-dependence**

The DA-coupling rate constant (or FRET-rate constant)  $k_{RET}$  depends to the DA-distance  $R_{DA}$  and the mutual DA-orientation. Quantitatively this is described by the Förster-relationship (Förster, 1948):

$$k_{RET} = k_{F,D} \cdot \kappa^2 \cdot \left( \frac{R_{0J}}{R_{DA}} \right)^6 = k_0 \cdot \left( \frac{R_0(\kappa^2)}{R_{DA}} \right)^6 \quad (5)$$

The mutual orientation is reflected by the orientation factor  $\kappa^2$  while  $R_{0J}$  is the spectral Förster-radius, a characteristic constant of the DA-pair. It is essentially defined by the refractive index of the medium  $n$  and the spectral overlap  $J$  integral of the donor fluorescence and the acceptor absorption spectrum:

$$R_{0J} = \left[ \frac{9(\ln 10)}{128\pi^5 \cdot N_A} \cdot \frac{J}{n^4} \right]^{\frac{1}{6}} \quad (6)$$

Usually the fluorophores undergo orientational motions. This results in a time-dependent orientation factor  $\kappa^2(t)$ . For small organic dyes typically FRET-rate constants are usually significantly smaller than their characteristic rate of mutual reorientation. Therefore, the time-dependent orientation factor  $\kappa^2(t)$  can be approximated by a mean orientation factor ( $\kappa^2(t) \approx 2/3$ ) (Kalinin et al., 2015). Contrary to small organic dyes with a typical characteristic depolarization times of a few hundred picoseconds the depolarization time of fluorescent proteins, given by their global rotation is considerably longer (Striker et al., 1999, Volkmer et al., 2000) resulting in additional broadening of the FRET-rate constant distributions (van der Meer et al., 2014).

### **2.2.3 FRET induced donor decay**

#### *Definition*

To obtain quantities independent of the fluorescence intensity references measurements are mandatory. However, contrary to the steady-state experiments with the FRET-efficiency as established measure of FRET, no wide-spread time-resolved analogue exists. Thus, experimental results are presented as fluorescence decay curves. This impedes the interpretation, as the effect of FRET is not sufficiently stressed. Therefore, we suggest the use of the ratio of the donor fluorescence decay in presence  $f_{DD}^{(DA)}$  and in absence  $f_{DD}^{(D0)}$  of the acceptor as experimental quantifier:

$$\varepsilon_D(t) = \frac{f_{D|D}^{(DA)}(t)}{f_{D|D}^{(D0)}(t)} \quad \text{or} \quad f_{D|D}^{(DA)}(t) = \varepsilon_D(t) \cdot f_{D|D}^{(D0)}(t) \quad (7)$$

We refer to this ratio as FRET-induced donor decay as it describes the time-dependent donor quenching by FRET. This factorization for single-exponential donor dyes was firstly introduced by Theodor Förster in his second paper (Förster, 1949) to separate the effect of FRET from emission of fluorescence. As shown below, the FRET-induced donor decay reveals information on the distribution of the underlying FRET rate constants. We illustrate the meaning of the FRET-induced donor decay by a few examples on a single exponential donor dye in more detail below. Later, we generalize this concept to the multi-exponential donor reference decays. Previously, van der Meer introduced a similar concept, the time-resolved FRET efficiency (TRE), by replacing the steady-state fluorescence intensities in equation (1) by the time-resolved fluorescence intensities (van der Meer et al., 2014). This approach also results in a time-dependent quantifier for FRET. However, the TRE has several disadvantages: (1) equations describing TRE are more complex, (2) the integral of the TRE diverges and (3) the TRE cannot be treated analogous to the time-resolved anisotropy.

Our definition of the FRET-induced donor decay  $\varepsilon_D(t)$  is completely analogous to the time-resolved anisotropy  $r(t)$  in the sense that two independent observables are used to derive a time-dependent quantity: The time-resolved anisotropy  $r(t)$  is given by the difference between the parallel and perpendicular intensities normalized to the total intensity while the donor FRET-decay  $\varepsilon_D(t)$  is given by the donor fluorescence intensity in presence of FRET normalized to its intensity in absence of FRET. The anisotropy decay  $r(t)$  describes the time-scale and degree of depolarization while the FRET-decay  $\varepsilon_D(t)$  describes the time-scale of FRET and fraction of molecules undergoing FRET. Further analogies are summarized in Table S1. Given the donor FRET decay, the corresponding steady-state observable  $E$  is obtained by the intensity weighted integration of  $\varepsilon_D(t)$ :

$$E = 1 - \frac{\int \varepsilon_D(t) \cdot f_{D|D}^{(D,0)}(t) dt}{\int f_{D|D}^{(D,0)}(t) dt} = 1 - \frac{\int f_{D|D}^{(D,A)}(t) dt}{\int f_{D|D}^{(D,0)}(t) dt} = 1 - \frac{F_{D|D}^{(D,A)}}{F_{D|D}^{(D,0)}} \quad (8)$$

This equation is analogous to the time-resolved anisotropy and the Perrin equation. It relates time-resolved information to the steady-state transfer efficiency. Similar relations can be formalized to obtain other observables as the fluorescence averaged lifetime (see chapter D). Analogous to the time-resolved anisotropy decay (Beechem et al., 2002) the FRET-induced donor decay is best fitted by a global analysis of experimental data.

*Single FRET species*

In case of a single-exponential donor reference and a single FRET-rate constant  $k_{RET}$  the fluorescence decay of a donor in absence  $f_{D/D}^{(D0)}(t)$  and in presence of an acceptor  $f_{D/D}^{(DA)}(t)$  are given by:

$$\begin{aligned} f_{D/D}^{(D0)}(t) &= k_{F,D} \cdot N_0 \cdot e^{-k_D t} \\ f_{D/D}^{(DA)}(t) &= k_{F,D} \cdot N_0 \cdot e^{-(k_{RET}+k_D)t} \end{aligned} \quad (9)$$

Thus, the FRET-rate constant  $k_{RET}$  can only be determined if both, the fluorescence decay in presence  $f_{D/D}^{(DA)}(t)$  and absence  $f_{D/D}^{(D0)}(t)$  of the acceptor, are known. Given both decays the FRET induced donor decay  $\varepsilon_D(t)$  provides the FRET-rate constant:

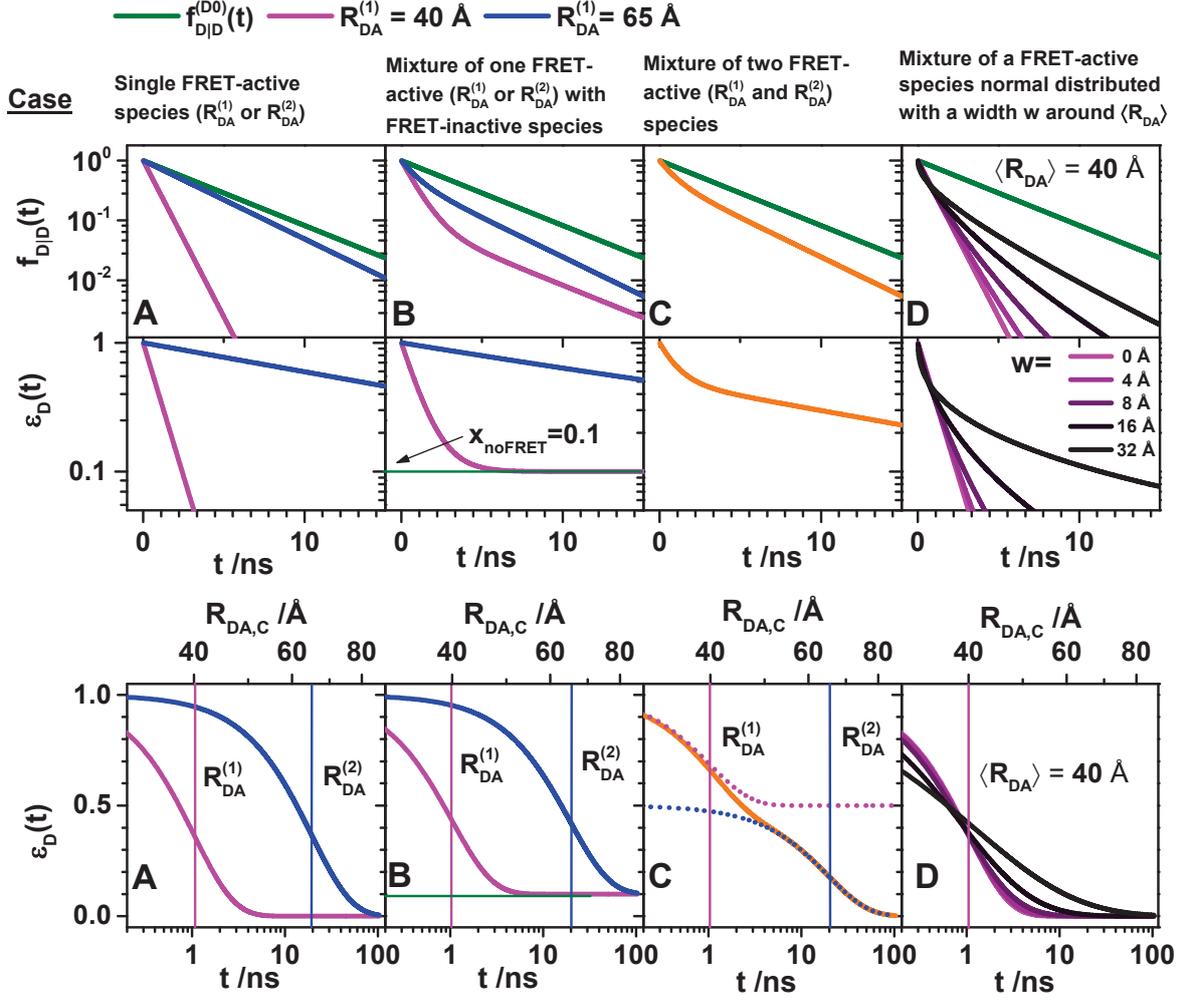
$$\varepsilon_D(t) = \frac{f_{D/D}^{(DA)}(t)}{f_{D/D}^{(D0)}(t)} = \frac{e^{-(k_D+k_{RET})t}}{e^{-k_D t}} = e^{-k_{RET}t} \quad (10)$$

Thus,  $\varepsilon_D(t)$  facilitates the interpretation of time-resolved FRET measurements, as the FRET-rate constant  $k_{RET}$  is directly obtained. Furthermore, the FRET-induced donor decay  $\varepsilon_D(t)$  rationalizes the global analysis of two independent decays. This is exemplified for a high FRET ( $R_{DA}=40 \text{ \AA}$ ) and a low FRET ( $R_{DA}=65 \text{ \AA}$ ) case in Figure 3A. In distances of the presented examples these correspond to FRET-rate constants of  $k_{RET}^{-1}=0.95 \text{ ns}$  and  $k_{RET}^{-1}=0.05 \text{ ns}$  ( $R_0=50 \text{ \AA}$ ,  $k_0^{-1}=4.0 \text{ ns}$ ). In a semi-logarithmic plot of the FRET-induced donor decay  $k_{RET}$  is obtained as the slope of the decay-curve (Figure 3, middle panel). Alternatively, the inverse of the FRET-rate constant  $k_{RET}^{-1}$  is obtained at the time  $t$  at which the FRET-induced donor decay  $\varepsilon_D(t)$  decayed to a value of  $1/e$ . This is best visualized in a semi-logarithmic plot of the time-axis (Figure 3, lower panel).

The interpretation of  $\varepsilon_D(t)$  in terms a DA-distances can be facilitated by expressing the time  $t$  in the dimension of a DA-distance by rewriting the Förster-relationship (eq. (5)):

$$R_{D,A,C}(t) = R_{0,J} \cdot \sqrt[6]{k_{F,D} \cdot \langle \kappa^2 \rangle(t) \cdot t} \approx R_0 \cdot \sqrt[6]{k_0 \cdot t} \quad (11)$$

Here,  $\langle \kappa^2 \rangle(t)$  is the average orientation factor of the DA-pair at that given time  $t$ . By this transformation the distance-axis  $R_{D,A,C}(t)$  contains the information on the mean orientation  $\langle \kappa^2 \rangle(t)$  and the Förster-radius  $R_0$  and the distance  $R_{DA}$  can be obtained visually at the distance  $R_{D,A,C}$  at which  $\varepsilon_D(R_{D,A,C})$  decayed to a value of  $1/e$  (Figure 3A lower-panel). This minimizes ambiguities, as the Förster-radius  $R_0$  as well as the mean orientation factor are implicitly considered (lower panels of Figure 3).



**Figure 3.** Fluorescence intensity decays of a donor  $f_{D|D}(t)$  (top row) in absence (green) and presence of FRET (blue, magenta and orange). The corresponding FRET-induced donor decays  $\varepsilon_D(t)$  are shown in the lower two rows. The fluorescence decays were calculated by eq. (11) (Single FRET-active species), eq. (14) (Mixture of FRET-active and FRET-inactive species) and eq. (15) (Mixture of FRET-species and distribution of FRET-species) ( $R_0 = 50 \text{ \AA}$  and  $k_D^{-1} = 4.0 \text{ ns}$ ). Information on FRET is obtained by comparing the fluorescence decay of the donor in presence of an acceptor (blue or magenta) to its reference given by the fluorescence decay in absence of FRET (green).  $\varepsilon_D(t)$  contains the reference implicitly. In the middle row  $\varepsilon_D(t)$  is shown in linear scale. In the lower row  $\varepsilon_D(t)$  is shown with a logarithmic time-axis and the time  $t$  between excitation and detection of fluorescence was converted into a critical donor-acceptor distance axis  $R_{DA,C}$  by eq. (13). This allows determining the characteristic times of FRET  $k_{RET}^{-1}$  and distances graphically at the point where  $\varepsilon_D(t)$  decayed to the value  $1/e$  (shown as vertical lines). These times  $t$  correspond to DA-distances of the FRET-process. (A) Single distances of  $R_{DA} = 40 \text{ \AA}$  (magenta) and  $R_{DA} = 65 \text{ \AA}$  (blue) (B) Mixture of a FRET-active  $R_{DA} = 40 \text{ \AA}$  (magenta) and  $R_{DA} = 65 \text{ \AA}$  (blue) and a FRET-inactive species (fraction,  $x_{noFRET} = 0.1$ ) (C) Mixture of two FRET-active species  $R_{DA}^{(1)} = 40 \text{ \AA}$  (50%) and  $R_{DA}^{(2)} = 65 \text{ \AA}$  (50%) (orange). The position and the height of the “steps” in the lowest plot relate to the FRET-rate constant and the species fractions of the individual species. For comparison the components (dotted blue and magenta lines) of the individual species are overlaid. (D) Normal distributed distance with a mean of  $\langle R_{DA} \rangle = 40 \text{ \AA}$  and a distribution width varying from 0 to  $32 \text{ \AA}$  (black to magenta).

### FRET Inactive species

The FRET-induced donor decay  $\varepsilon_D(t)$  facilitates the analysis of species mixtures. In species mixtures the total intensity is given by the sum of the individual fluorescence intensities decays weighted by the respective species fractions. Thus,  $\varepsilon_D(t)$  is particularly useful to distinguish

FRET-active from FRET-inactive species. If a FRET-sample consists of a mixture of FRET-active and FRET-inactive molecules with species fractions of  $1-x_{DOnly}$  and  $x_{DOnly}$  the fluorescence decay is given by:

$$f_{D/D}^{(DA)}(t) = k_F^D \cdot N_0 \cdot \left[ (1-x_{DOnly}) \cdot e^{-(k_D+k_{RET})t} + x_{DOnly} \cdot e^{-k_D \cdot t} \right] \quad (12)$$

This is illustrated in the top of Figure 3B for a low-FRET and a high-FRET case with a fraction of 10% FRET-inactive molecules.

The FRET-rate constants and the fraction of FRET-inactive molecules  $x_{DOnly}$  are difficult to estimate visually given only the fluorescence intensity decays. The fraction of FRET-inactive species  $x_{DOnly}$  is given by a constant offset of the FRET-induced donor decay:

$$\varepsilon_D(t) = (1-x_{DOnly}) \cdot e^{-k_{RET}t} + x_{DOnly} \quad (13)$$

Therefore,  $x_{DOnly}$  is visually easily estimated. In case of the presented high-FRET case the amplitude of  $\varepsilon_D(t)$  essentially decays to zero within a time of 15 ns (Figure 3B) and the species fraction of the FRET-inactive molecules is easily determined. In the low-FRET case the amplitude of  $\varepsilon_D(t)$  only decays to 47 % of the initial amplitude 15 ns after excitation (Figure 3B). Hence, the fraction of inactive molecules is less reliable as it has to be determined by extrapolation.

#### *Mixture of FRET species*

As fluorescence intensities are additive, the concept of the FRET-induced donor decay is easily generalized to species mixtures with species specific rate constants  $k_{RET}^{(i)}$  and species fractions  $x_{RET}^{(i)}$ . For a mixture of  $N$  species with a common single-exponential donor reference the FRET-induced donor decay is given by:

$$\varepsilon_D(t) = \sum_{i=1}^N x_{RET}^{(i)} \cdot \varepsilon^{(i)}(t) \quad (14)$$

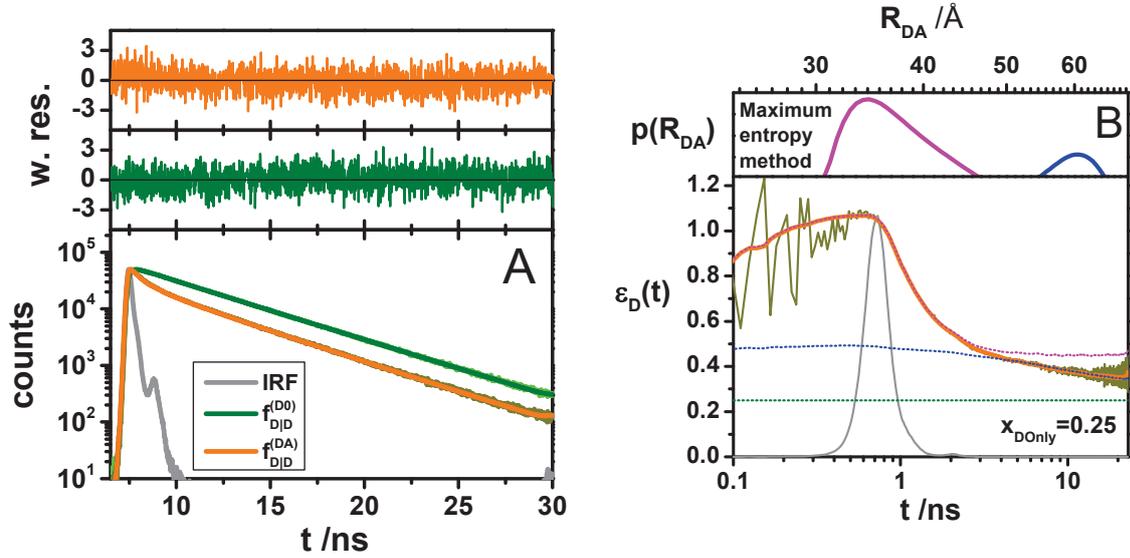
In Figure 3C this is shown for a mixture of high- and a low-molecules. This discrete case can be simply generalized to a continuous distribution of FRET-rate constants  $p(k_{RET})$  where the FRET-induced donor decay is given by an integral:

$$\varepsilon_D(t) = \int_0^{\infty} p(k_{RET}) \cdot \exp[-k_{RET} \cdot t] dk_{RET} \quad (15)$$

If the time-axis is transformed into a distance axis the FRET-decay is given by:

$$\varepsilon_D(R_{DA,C}) = \int_0^\infty p(R_{DA}) \cdot \exp\left[-\left(\frac{R_{DA,C}}{R_{DA}}\right)^6\right] dR_{DA} \quad (16)$$

Such case is illustrated in Figure 3D for a set of normal distributed distances centered at  $\langle R_{DA} \rangle = 40 \text{ \AA}$  with varying width from 0 to 32  $\text{\AA}$ . Most information is obtained visually by of plot of  $\varepsilon_D(t)$  a logarithmic time-axis (Figure 3, bottom panel). Here, the characteristic times and distances are obtained from the positions of the ‘‘steps’’ while species fractions by their height. Distance distributions or multiple states can be identified by deviations from the ideal exponential behavior (Figure 3D).



**Figure 4.** Experimental fluorescence decay (A) and FRET induced donor decay (B) from ensemble  $f$  measurements on human guanylate binding protein 1 (hGBP1) dimers induced by GTP $\gamma$ S (500  $\mu$ M) labeled by Alexa 488 ( $\tau_0=4.1$  ns) and Alexa 647N ( $R_0=52$   $\text{\AA}$ ) at the position Q577C. (A) Fluorescence intensity decays in absence ( $\tau_1=4.2$ ,  $x_1=0.94$ ,  $\tau_2=1.7$  ns,  $x_2=0.06$ ) (green) and presence (orange) of the FRET, instrument response function in (gray) derived (B) The distance distribution obtained by inversion of the time-resolved fluorescence intensity decays by MEM is displayed on top of the FRET-decay (blue high FRET, green low FRET, (yellow experimental, orange fit)). The time-axis measures the time between excitation and detection of donor photons. The distance axis is given by the Förster-relationship  $R_{DA}=R_0(t/\tau_0)^{1/6}$ . The distances corresponding to the two steps are visualized the components of the distance distributions (dashed green 38  $\text{\AA}$ , dashed magenta 58  $\text{\AA}$ ).

As demonstrated in Figure 4 on a dimers of the human guanylate binding protein 1 (hGBP1) induced by GTP $\gamma$ S this concept of  $\varepsilon_D(t)$  can be directly applied to experimental data, if the instrument response function (IRF) is sufficiently narrow and the donor in absence of FRET is close to single-exponential. In Figure 4A measured histograms of the donor fluorescence decays with and without acceptor are shown. The fluorescence decay of the FRET-sample is clearly distinguishable from the donor sample. However, a direct comparison only reveals a single species visually and no conclusion of the associated distances or fractions can be made. In  $\varepsilon_D(t)$  visually two distinct species are be identified (Figure 4B). A high-FRET species with a distance around 35  $\text{\AA}$  and a low-FRET species with a distance around 60  $\text{\AA}$ . For comparison we display

in Figure 4 to the right a model free computer analysis of the fluorescence intensity decay of the donor in presence of FRET by the maximum entropy method (MEM) (Livesey and Skilling, 1985, Brochon, 1994) that explicitly considers the instrument response function (IRF) and fluorescent background (Vöpel et al., 2014). The agreement between both analysis methods is quite good, but DA distance of the high-FRET species recovered by the MEM analysis is shifted by  $\sim 5 \text{ \AA}$  to smaller values. This highlights the necessity of accounting effects of IRF in the analysis or using short detector rise-time to resolve short DA-distances.

### 2.2.4 Treating a reference with several partially quenched donor states

In case of multi-exponential donor and acceptor reference all lifetime components and FRET-rate constant have to be treated individually. For a donor-state ( $i$ ) and acceptor-state ( $j$ ) with rate constants  $k_D^{(i)}$  and  $k_A^{(j)}$  the time-dependent population of the excited state for the DA-pair ( $i,j$ ) is given by:

$$n_{D^*A}^{(i,j)} = e^{-(k_{RET}^{(i,j)} + k_D^{(i)})t} = \varepsilon^{(i,j)} d^{(i)} \quad (17)$$

$$n_{DA^*}^{(i,j)} = w_{ij} \left( e^{-(k_{RET}^{(i,j)} + k_D^{(i)})t} - e^{-k_A^{(j)}t} \right) = w^{(i,j)} \left( \varepsilon^{(i,j)} d^{(i)} - a^{(j)} \right) \quad \text{with} \quad w^{(i,j)} = \left( 1 + \frac{k_D^{(i)} - k_A^{(j)}}{k_{RET}^{(i,j)}} \right)^{-1}$$

If the probability  $x_D^{(i)}$  of being in a donor states ( $i$ ) is independent from the probability  $x_A^{(j)}$  of the acceptor states ( $j$ ) the probability of the donor-acceptor pair ( $i,j$ ) is given by the product of the donor and acceptor probability  $x_D^{(i)} \cdot x_A^{(j)}$  and the fluorescence decay of A and the FRET sensitized acceptor fluorescence are obtained by summation over all pairs:

$$f_{D|D}^{(DA)}(t) = k_{F,D} \cdot N_0 \cdot n_{D^*A} = k_{F,D} \cdot N_0 \cdot \sum_{ij} x_A^{(j)} x_D^{(i)} \varepsilon^{(i,j)} d^{(i)} \quad (18)$$

$$f_{A|D}^{(DA)}(t) = k_{F,A} \cdot N_0 \cdot n_{DA^*} = k_{F,A} \cdot N_0 \cdot \sum_{ij} x_A^{(j)} x_D^{(i)} w^{(i,j)} \left( \varepsilon^{(i,j)} d^{(i)} - a^{(j)} \right)$$

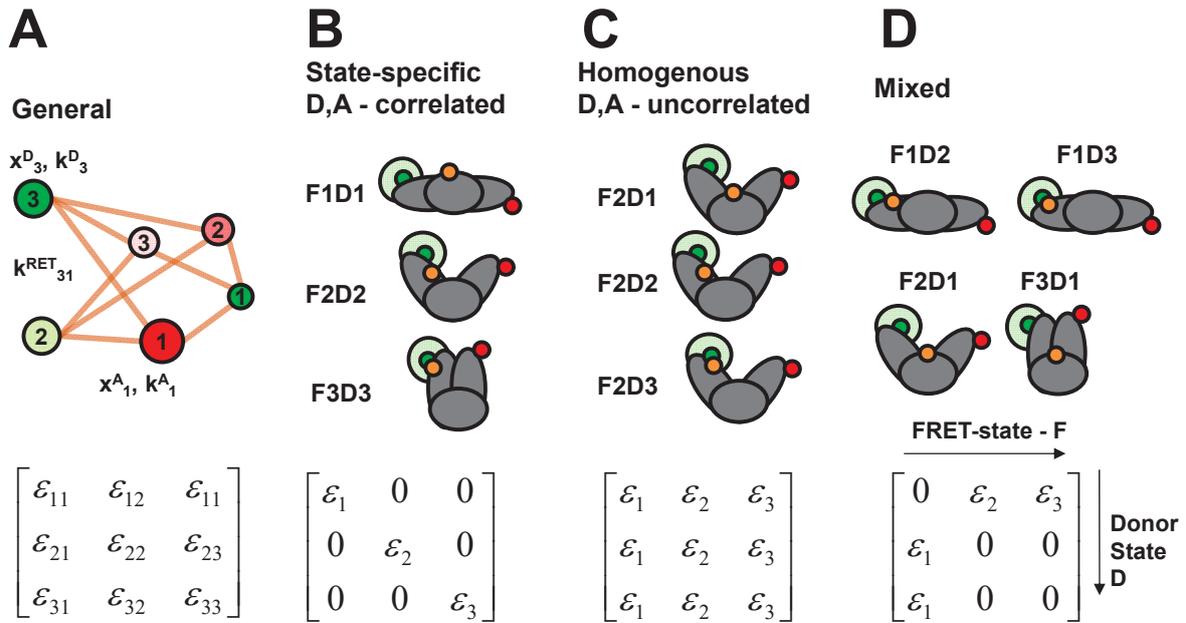
The aim of analyzing the fluorescence decays  $f_{A|D}^{(DA)}(t)$  and  $f_{D|D}^{(DA)}(t)$  is to obtain the fractions  $x_A^{(j)} x_D^{(i)}$  and the decays  $\varepsilon^{(i,j)}$ . Once  $x_A^{(j)} x_D^{(i)} = x_{DA}^{(i,j)}$  and  $\varepsilon^{(i,j)}$  are known a FRET-decay which is independent of the donor or acceptor fluorescence is obtained by summation over the matrix elements:

$$\varepsilon(t) = \sum_{ij} x_{DA}^{(i,j)} \varepsilon^{(i,j)} \quad (19)$$

For a single exponential donor reference the FRET-induced donor decay of  $\varepsilon_D(t)$  equals the FRET-decay  $\varepsilon(t)$ .

A key problem of this approach is that the number of free parameters increases the uncertainty of the obtained fitting results. For instance, in case of three donor- and three acceptor-states as

illustrated in Figure 5A overall 17 parameters (9 FRET-rate constants contained in the matrix elements  $\varepsilon^{(i,j)}$  and 8 fractions given by the pair probability  $x_{DA}^{(i,j)}$ ) have to be determined. The number of free parameters can be reduced by jointly analyzing the donor- and the acceptor references to impose restrictions on  $x_{DA}^{(i,j)}$ . The number of free parameters can be further reduced by making assumptions regarding the values of the matrix elements  $\varepsilon^{(i,j)}$  take.

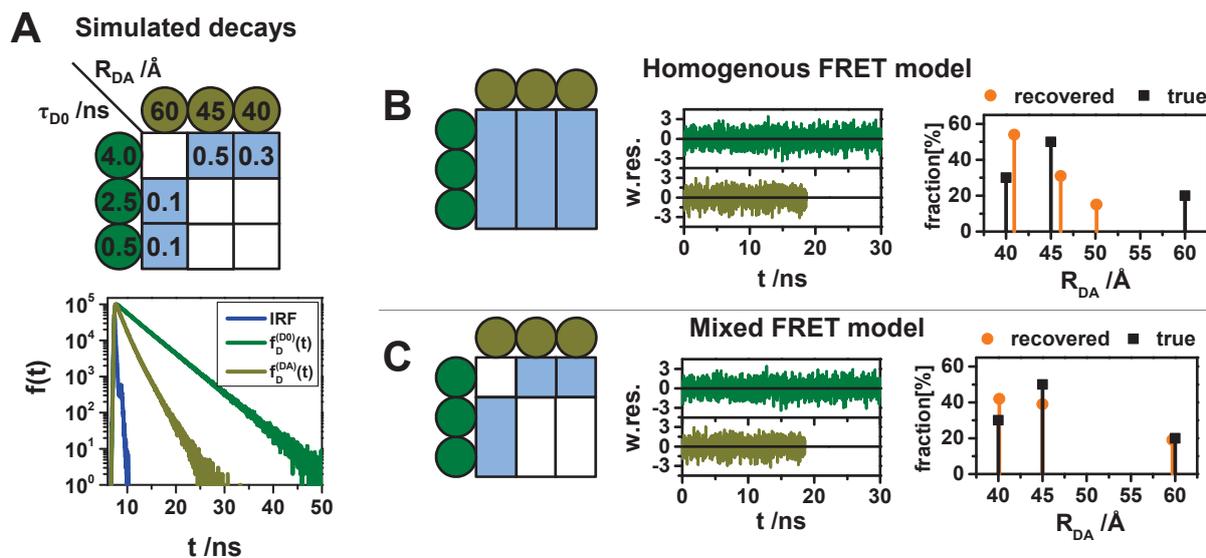


**Figure 5.** FRET-matrix  $\varepsilon_{ij}$  and graphical representations corresponding conformational states –The general case is illustrated by a network of discrete numbered donor states (green) and acceptor states (red). Each donor and acceptor state has a certain species fraction and decay constant  $x^D/k^D_i$  and  $x^A/k^A_j$ , respectively. To each donor-acceptor pair a FRET-rate constant  $k^{RET}_{ij}$  is associated. In specific cases some matrix elements vanish or are dependent. This is illustrated schematically by a protein (gray) with three FRET-states (F1, F2 and F3) labeled by a donor (green) and acceptor (red) where a quencher (orange) leads to three distinct donor states (D1, D2 and D3). Depending on the correlation between the FRET-states and the donor-states the matrix elements take different values.

For instance, it might be assumed a change in FRET correlate with changes of  $k_D^{(i)}$ . Such a case is illustrated in the upper panel of Figure 5B. Here, the donor fluorophore is not quenched the protein state F1 while it is slightly quenched in F2 and strongly quenched F3. Under these conditions the FRET-matrix  $[\varepsilon^{(i,j)}]$  is a diagonal matrix as shown in the lower panel of Figure 5B. Alternatively, it may be assumed that quenching of the donor by the host molecule is uncorrelated to quenching by FRET. We call this case uncorrelated or homogenous, as all donor-states are quenched by the same FRET-rate constant distribution. This is illustrated for the protein state F2 of a three state protein in Figure 5C. In all three FRET-states F1, F2 and F3 the donor lifetime distribution in absence of FRET are identical. Under these conditions the interpretation of the FRET-induced donor decay by a FRET-rate distribution is exact. This is a simplification

is often used but rarely stressed and may lead, as outlined below, to significant errors in particular for minor states. Practically, also mixtures of the presented limiting cases may arise. This is illustrated in Figure 5D. Here, the donor is unquenched in the protein in the FRET-states F2 and F3 while it is not quenched in the state F1.

If only  $f_{D|D}^{(DA)}(t)$  and  $f_{D|D}^{(D0)}(t)$  are monitored it is impossible to distinguish specific, homogenous or the mixed cases (compare Figure 5), as both FRET and quenching of the donor by the local environment shorten the donor fluorescence lifetime. This is exemplified in Figure 6, where fluorescence intensity decays  $f_{D|D}^{(DA)}(t)$  and  $f_{D|D}^{(D0)}(t)$  of a mixed case were simulated and analyzed by the homogenous model and by the correct model. Overall,  $12 \cdot 10^6$  photons were simulated for FRET-sample  $f_{D|D}^{(DA)}(t)$  and  $30 \cdot 10^6$  photons for the donor sample  $f_{D|D}^{(D0)}(t)$ , to mimic typical FRET-experiments in terms of photon statistics. As highlighted in Figure 6B/C both models differ significantly in the recovered distances and fractions. Nevertheless, they are indistinguishable as judged by the quality of the fits. Clearly, the incorrect homogenous model affects the minor FRET-state  $\varepsilon^{(3)}$  (60 Å) stronger as compared to the major state. For the minor state and the incorrect homogenous model, the recovered distance differs by roughly 10 Å from the correct distance. The two major states  $\varepsilon^{(1)}$  and  $\varepsilon^{(2)}$  are less affected by the choice of the model and differ only by 1 Å from the correct value (Figure 6B). Compared to the distances the amplitudes are stronger influenced by the choice of the model and differ by at most by 22% from the correct amplitudes. If the correct fitting model is used the simulated distances and amplitudes are recovered (Figure 6C). This exemplifies that the interpretation of the experimental results is ambiguous, if no knowledge about the connectivity of the donor and FRET-states is available. However, as we will demonstrate below by simulations, the homogenous approximation works surprisingly well for flexible coupled dyes.



**Figure 6.** (A) Simulated time-resolved fluorescence decay histograms with 100.000 photons in peak (bin-width 14.1 ps) of a system with three donor states 4 ns (80%), 2.5 ns (14%) and 0.5 ns (6%) and three FRET-states 40 Å (30%), 45 Å (50%), and 60 Å (20%) ( $R_0=50$  Å,  $k_0^{-1}=4$  ns). The 40 Å, and 45 Å are associated to the donor lifetime of 4 ns, 60 Å state is associated to the donor lifetimes 2.5 ns and 0.5 ns. This connectivity of the states is illustrated by the table the fractions are shown as numbers (B) Analysis result assuming all donor states are homogenous quenched by FRET (C) Analysis result given the correct model.

## 2.2 Mobile dyes and FRET

### 2.2.1 Concept

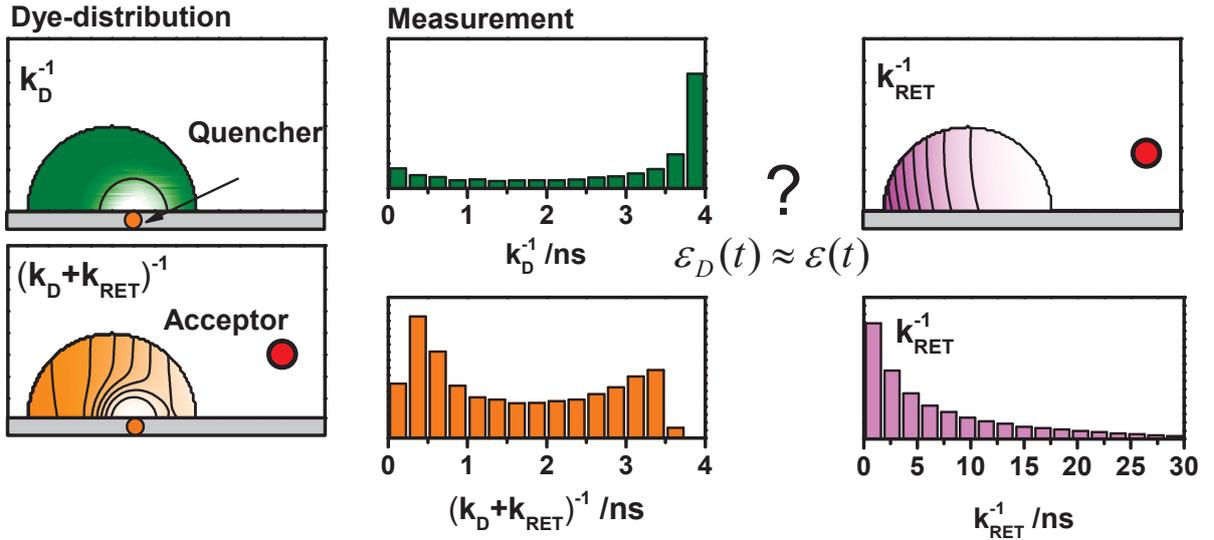
Instead of interpreting the fluorescence decays by FRET-rate constants, our ultimate goal are quantitative structural models of the biomolecules. To achieve this goal, the spatial and dynamic properties of the fluorophores have to be considered. Usually, the fluorophores are tethered to biomolecules via flexible linkers with a length around 20 Å. This results in DA-distributions, which can be quantified by the accessible volume (AV) simulations (Sindbert et al., 2011). In AV simulations the spatial equilibrium distribution of the fluorophores is approximated by their sterically allowed conformational space. Given this conformational space, fluorescence observables are calculated and compared to experiments. This approach is used in the FRET Positioning and Screening (FPS) toolkit to derive quantitative structural models (Kalinin et al., 2012).

The effect of dye quenching has not been considered so far in FPS. It is known that quenchers within the AV alter the fluorescence lifetime, i.e. xanthene dyes are quenched by aromatic amino acids via photo-induced electron transfer (PET) (Doose et al., 2009). The rate constant of electron transfer in PET can be described by an exponential distance dependence (Marcus and Sutin, 1985). The characteristic length of this process is in the order of a few Ångström (Moser et al., 1992, Gray and Winkler, 1996). This modulates, even in absence of diffusion, the fluorescence lifetime of the dye within its AV as illustrated on the top of Figure 7. Alternatively,

the distance dependence in PET is approximated by a step-function. If the fluorophore quencher distance is smaller than  $R_{rad}$  then the fluorophore is quenched with a rate constant  $k_Q$ :

$$k_Q(R_e) = \begin{cases} k_Q & : R_e \leq R_{rad} \\ 0 & : R_e > R_{rad} \end{cases} \quad (20)$$

Such approximation is called radiation boundary model. The characteristic length  $R_{rad}$  (the radiation boundary) is significantly smaller than the linker-length. Therefore, the lifetime distribution is unevenly distributed within the AV. Additionally, dye motilities has to be considered, as the dye motion results in a time-dependent quenching rate constant  $k_Q(t)$ . For freely diffusing fluorophores and quenchers analytical solutions of the radiation boundary model exists and found application in various experimental studies (Flannery, 1982, Periasamy et al., 1988, Joshi et al., 1987).



**Figure 7.** Effect of a quencher on the fluorescence lifetime distribution of a donor in presence and absence of FRET within its accessible volume (shown as a half-circle). The lines in the half-circles are iso-lines of the lifetimes (left) or the inverse of FRET-rate constants (right). Experimentally, the donor fluorescence lifetime distribution (middle) in absence (top) and presence (bottom) of FRET is accessible and spatial correlations between FRET and quenching cannot be considered within knowing the position of the quencher. This raises the question if the experimental accessible FRET-induced donor decay  $\varepsilon_D(t)$  (eq. 7) may be interpreted as FRET-decay  $\varepsilon(t)$  (eq. 19).

Experimentally spatial correlations between the donor-lifetimes and FRET are inaccessible, if only the donor fluorescence decays  $f_{D|D}^{(DA)}(t)$  and  $f_{D|D}^{(D0)}(t)$  are monitored (Figure 7). This raises the question which accuracy may be achieved for a single flexible coupled DA-pair if the FRET-induced donor decay  $\varepsilon_D(t)$  is interpreted by a homogenous model. Lacking experimental data and analytical solutions of the time-dependent population of the excited state for complex en-

vironments, we simulate the donor in absence and presence of the acceptor. By these simulations we estimate the accuracy of the homogenous approximation. In the simulations the dye diffusion, quenching by the local environment, and FRET are considered. As previously shown, the effects of dye mobilities can be assessed by calibrated molecular dynamic (MD)-simulations (Kalinin et al., 2015). However, to assure sufficient sampling, unpractically long MD simulations are needed. Therefore, we developed a coarse grained model (qtAV) that combines the merits of fast accessible volume (AV) calculations, atomistic knowledge on the local protein structure, amino-acid specific dynamic quenching rate constants and prior knowledge on the dye motilities, to study the effects of local quenching the fluorescence decays.

Ideally, the time-dependent position of the donor  $r_D(t)$  and of the acceptor  $r_A(t)$  are considered to derive the population  $N^*(r_D, r_A, t)$  of the excited donor. Changes of  $N^*(r_D, r_A, t)$  due to dye diffusion, FRET, and the rate constant  $k_D(r_D)$  are given by:

$$\frac{\partial}{\partial t} N^*(r_D, r_A, t) = \left[ \frac{\partial}{\partial r_A} \left( D_A \frac{\partial}{\partial r_A} \right) + \frac{\partial}{\partial r_D} \left( D_D \frac{\partial}{\partial r_D} \right) - k_0 \left( \frac{R_0}{|r_D - r_A|} \right)^6 - k_D(r_D) \right] \quad (21)$$

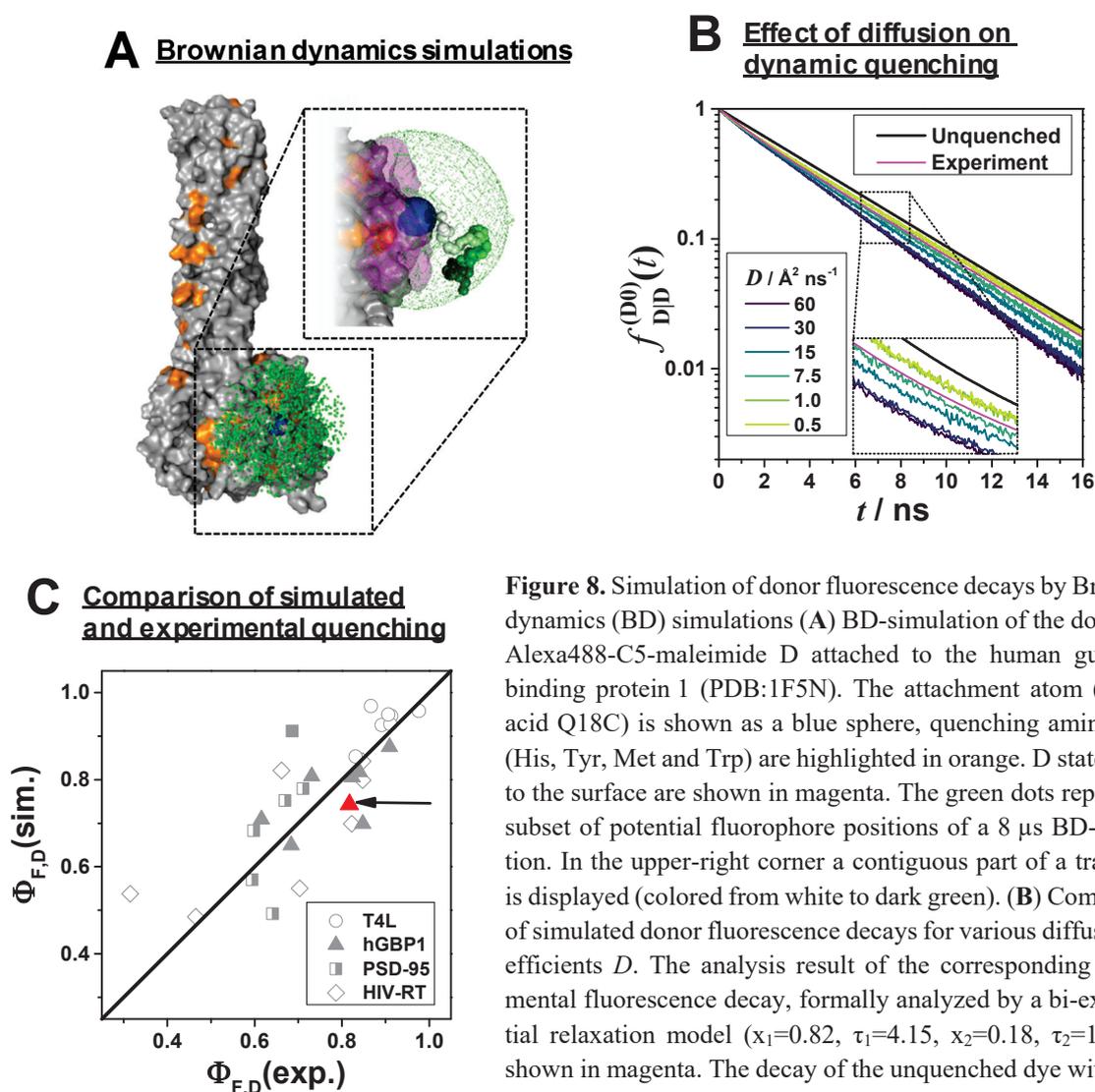
This is a modified form of the Haas-Steinberg equation which accounts the donor and the acceptor diffusion separately. This allows to consider correlations between the FRET-rate constant and  $k_D(r_D)$ . Following eq. 20 the position dependent rate constant of the donor  $k_D(r_D)$  is obtained by summation of the distance dependent quenching rate constant. Experimentally the ensemble average over the donor and acceptor positions  $\langle N^*(t) \rangle_{r_D, r_A}$  can be determined.

We solve this partial differential equation, by simulating the of D- and A-diffusion by Brownian dynamics (BD) simulations. Similar simulations have been recently used for insights into reaction on electrodes (Huang and White, 2013). After simulating the dye diffusion, we calculate in a subsequent step the average time-dependent population of the excited state. These simulations are visualized in Figure 8A using a structure of the human guanylate binding protein 1 (hGBP1) labeled at the amino acid Q18C by Alexa488. First, we determine the AV of the dye (inset Figure 8A, green net) and we identify all potential quenching amino acids of the structure (Figure 8A, orange surfaces). Next, we simulate the diffusion of the dye within its AV by BD-simulations where the dye is slowed down based in the vicinity of the molecular surface (inset Figure 8A, magenta volume). We assume, that the rotational is fast. Thus, we approximate the dye by a single sphere. Following the BS-simulations the coordinates of the fluorophore are converted into a trajectory of rate constants  $k_D(t)$  using the distances to the quenching amino

acids. Given the  $k_D(t)$ -trajectory the fluorescence decays are calculated by numeric integration, either by simulating the photon emission or by direct calculation of fluorescence decay curves.

### 2.2.2 Simulation of dynamic donor quenching

The coarse grained quenching and traps accessible volume (qtAV) dye model presented in Figure 8 requires assumptions with respect to the diffusion coefficients of the dyes and the quenching rate constants. Amino acid specific quenching rate constants were obtained by scaling an initial estimated by experimental dynamic quenching rate constants measured in bulk solution (Chen et al., 2010). As diffusion coefficient for Alexa488 and Alexa647 values of  $10 \text{ \AA}^2/\text{ns}$  and  $5.0 \text{ \AA}^2/\text{ns}$ , respectively, have been used.



**Figure 8.** Simulation of donor fluorescence decays by Brownian dynamics (BD) simulations (A) BD-simulation of the donor dye Alexa488-C5-maleimide D attached to the human guanylate binding protein 1 (PDB:1F5N). The attachment atom (amino acid Q18C) is shown as a blue sphere, quenching amino acids (His, Tyr, Met and Trp) are highlighted in orange. D states close to the surface are shown in magenta. The green dots represent a subset of potential fluorophore positions of a  $8 \mu\text{s}$  BD-simulation. In the upper-right corner a contiguous part of a trajectory is displayed (colored from white to dark green). (B) Comparison of simulated donor fluorescence decays for various diffusion coefficients  $D$ . The analysis result of the corresponding experimental fluorescence decay, formally analyzed by a bi-exponential relaxation model ( $x_1=0.82$ ,  $\tau_1=4.15$ ,  $x_2=0.18$ ,  $\tau_2=1.35$ ), is shown in magenta. The decay of the unquenched dye with a fluorescence lifetime of 4.1 ns is shown in black. (C) Simulated fluorescence quantum yields of fluorescent species  $\Phi_{F,D}(\text{sim.})$  for a diffusion coefficient of  $D = 15 \text{ \AA}^2/\text{ns}$  vs. experimentally determined quantum yields  $\Phi_{F,D}(\text{exp.})$  for a set of variants of the proteins T4L, hGBP1, PSD-95 and HIV-RT. The black line shows a 1:1 relationship.  $\Phi_{F,D}(\text{exp.})$  was determined by ensemble TCSPC (hGBP1, T4L, PSD-95) or single-molecule measurements (HIV-RT). The data-point highlighted by the red arrow corresponds to the experiment shown in (B).

fluorescence quantum yields of fluorescent species  $\Phi_{F,D}(\text{sim.})$  for a diffusion coefficient of  $D = 15 \text{ \AA}^2/\text{ns}$  vs. experimentally determined quantum yields  $\Phi_{F,D}(\text{exp.})$  for a set of variants of the proteins T4L, hGBP1, PSD-95 and HIV-RT. The black line shows a 1:1 relationship.  $\Phi_{F,D}(\text{exp.})$  was determined by ensemble TCSPC (hGBP1, T4L, PSD-95) or single-molecule measurements (HIV-RT). The data-point highlighted by the red arrow corresponds to the experiment shown in (B).

For the proteins T4 lysozyme (T4L), human guanylate binding protein 1 (hGBP1), post synaptic density protein 95 (PSD-95) and the reverse transcriptase of HIV (HIV-RT) reference measurements with Alexa488 tethered to different positions were performed and protein structures are known: T4L (PDB: 148L, 172L), hGBP1 (PDB: 1F5N), HIV-RT (PDB: 1RTD). In case of PSD-95 the structural model determined by CNS and multiple FRET-measurements was used as a reference (McCann et al., 2011). For the given protein structures fluorescence decays of Alexa488 were simulated and compared to the experimental decays using the fluorescence quantum yield of the dynamically quenched dye as a sum parameter. All simulations were performed with the same simulation parameters as reported in Table S2-S4. To obtain simulation comparable to the experiments, we simulate photon trajectories which were binned in fluorescence decay histograms. These fluorescence decay histograms were analyzed analogous to the experimental data by fitting a multi-exponential relaxation model to yield the species averaged lifetime. The species averaged lifetime were used to calculate the fluorescence quantum yields (see Figure 8B/C). The reported experimental quantum yields were either determined by ensemble TCSPC (PSD-95, T4L), sub-ensemble TCSPC (hGBP1) or the molecular brightness as estimated by lifetime filtered FCS (HIV-RT) using rhodamine 110 as a reference.

Even though the qtAV dye model is relatively simple, it correctly predicts complex fluorescence decays due to dynamic quenching based on biomolecular structures (see Figure 8C). This is surprising, as certain aspects are not considered: (1) strong interactions between the biomolecule and the dyes; (2) orientational effects of PET; (3) protein dynamics. Some of the mentioned shortcomings seem of minor importance. (1) For instance, do not have to be considered, as strong ground state complexes result in static quenching. (2) Describing the dye by a sphere proved to be a good approximation in sterically undemanding environments (Sindbert et al., 2011). (3) Even though, orientational effects in PET are of importance, side-chains rotations are fast ( $\sim 100$  ps), compared to the dye diffusion. Hence, the dye experiences a mean quenching efficiency of the side chain. (3) The presence of multiple conformational states of the biomolecule is a major problem when comparing experiment and simulations. The proteins PSD-95, T4L and HIV-RT are known for their conformational dynamics, while little is known on hGBP1. PSD-95 is highly dynamic, as two presumably rigid PDZ-domains are connected by a flexible linker (McCann et al., 2012). HIV-RT is known for its conformational dynamics associated to its “finger” (Rothwell et al., 2013), while T4L is known for its a global hinge bending motion (Yirdaw and Mchaourab, 2012). Even though, in all samples multiple conformational

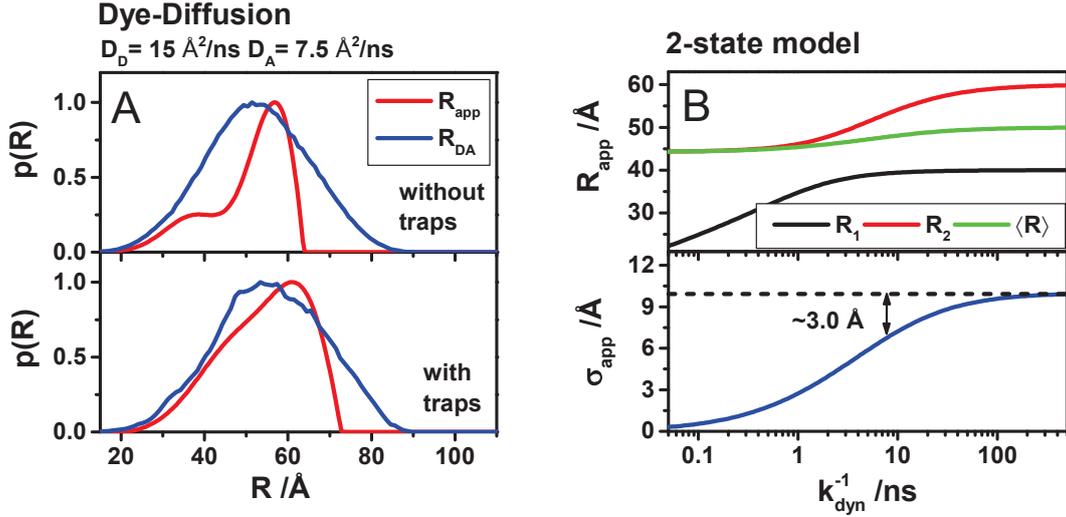
states are presumably present, the correlation between the experiments and the simulations (Figure 8C) indicates, that the dye environment is rather independent on the conformational state of the biomolecule.

The diffusion coefficients of the dyes clearly influence the time-resolved fluorescence intensity decay. This is demonstrated in Figure 8B where simulated fluorescence decays for varying diffusion coefficients, ranging from 0.5 to 60 Å<sup>2</sup>/ns, are shown. The simulations agree best with the experiment, if in the simulations the dye diffuses with a diffusion coefficient around 10 Å<sup>2</sup>/ns. Given that the diffusion coefficient of the free dye is the range of 40 to 45 Å<sup>2</sup>/ns (Gendron et al., 2008, Müller et al., 2008) this is reasonable, as the dye is hindered by the tether.

### **2.2.3 Impact of dye diffusion on FRET**

Before testing the precision of the homogenous AV approximation, we investigate the effect of dye diffusion on the apparent distance distributions  $p(R_{app})$  recovered by equation (16). Actual DA-distributions  $p(R_{DA})$  are only recovered from fluorescence decays, if the applied analysis model explicitly accounts for the dye mobilities, or if the time-scale of dynamics is significantly slower than the time-scale of fluorescence. Diffusional effects are a well-known phenomenon and results shortened apparent distances (Jacob et al., 2013). This can be exploited to study chain diffusion of unfolded proteins (Möglich et al., 2006). Calibrated MD-simulations of the dyes Alexa488 and Cy5 tethered to DNA indicate that the fluorescence intensity decays can be described by fast rotating but otherwise static dyes (Kalinin et al., 2015).

Our dye model allows for the long simulation times (64 μs). This minimizes the simulation noise of the obtained fluorescence decays. To facilitate the inversion of the simulated fluorescence decays to the apparent distance distributions  $p(R_{app})$  by eq. (16), we calculate them without simulating fluorescent photons. In Figure 9A apparent distance distributions  $p(R_{app})$  are shown for donor and acceptors dyes with diffusion coefficients of 15 Å<sup>2</sup>/ns and 7.5 Å<sup>2</sup>/ns, respectively. To highlight the effect of the dyes interacting (sticking) with the biomolecular surface, two simulations were performed. In the first, the dyes diffuse with a uniform diffusion coefficient within their AV. In the second, the dyes in vicinity of the molecular surface diffuse slower (stick). For a non-interacting dyes, a main peak and a shoulder are visible in  $p(R_{app})$ . These features are not present in  $p(R_{DA})$  (Figure 9A, top). For the interacting dyes, a broad distribution  $p(R_{app})$  is recovered resembling better the shape of  $p(R_{DA})$  (Figure 9A, bottom). To assure, that the shoulder observed in  $p(R_{app})$  is indeed caused by the dye diffusion and not an artifact of the inversion method, the simulated fluorescence decays were additionally fitted by normal distributions and the diffusion coefficient was changed. With increasing diffusion coefficients, the shoulder became more pronounced (not shown).



**Figure 9.** Effect of the dye diffusion on the apparent donor-acceptor distance distribution (A) Simulations on hGBP1 labeled on amino-acid F379C (donor) and D467C (acceptor) with and without surface trapping (B) Apparent distances of a two-state system in dependence of the exchange rate constant  $k_{exc}$  between the two states – the two states are equally populated with the limiting states of 40 Å and 50 Å ( $R_0=50$  Å). The resulting bi-exponential time-dependent FRET decay was converted to two apparent distances (red and black) and the average distance (green). The apparent distances as well as the apparent species fractions change with the exchange rate constant. The resulting mean and standard-deviation of the apparent distances is shown as blue line. The black dotted line serves as reference and gives the standard deviation in the absence of dynamic exchange between the two states.

The distorted shape of the apparent distance distribution  $p(R_{app})$  compared to the DA-distance distribution  $p(R_{DA})$  and the shift of its mean compared to the average of  $p(R_{DA})$  can be rationalized by a dynamic two-state system. For a two-state system with equal probability of the states an analytical solution of the time-resolved fluorescence decay is known (Westlund and Wennerstrom, 1993). If the two states exchange with a rate constant of  $k_{dyn}$  and have the FRET-rate constants of  $k_{RET}^{(a)}$  and  $k_{RET}^{(b)}$  the FRET-induced donor decay is given by:

$$\begin{aligned} \varepsilon(t) &= \frac{1}{2} \left( 1 - \frac{k_{dyn}}{\sqrt{\Delta^2 + k_{dyn}^2}} \right) e^{-t(\Sigma + k_{dyn} + \sqrt{\Delta^2 + k_{dyn}^2})} + \frac{1}{2} \left( 1 + \frac{k_{dyn}}{\sqrt{\Delta^2 + k_{dyn}^2}} \right) e^{-t(\Sigma + k_{dyn} - \sqrt{\Delta^2 + k_{dyn}^2})} \\ &= x_{app}^{(1)} \cdot e^{-t \cdot k_{app}^{(1)}} + x_{app}^{(2)} \cdot e^{-t \cdot k_{app}^{(2)}} \end{aligned} \quad (22)$$

Above  $\Sigma = 1/2 \cdot (k_{RET}^{(a)} + k_{RET}^{(b)})$  and  $\Delta = 1/2 \cdot (k_{RET}^{(a)} - k_{RET}^{(b)})$  solely depend on the FRET-rate constant while the pre-factors  $x_{app}^{(1)}$ ,  $x_{app}^{(2)}$  and rate constants  $k_{app}^{(a)}$ ,  $k_{app}^{(b)}$  additionally depend on the exchange with a rate constant of  $k_{dyn}$ . Thus, only apparent species fraction and FRET-rate constants are obtained if the exchange rate constant  $k_{dyn}$  is big and the rate constants  $k_{app}^{(a)}$ ,  $k_{app}^{(b)}$  are treated as FRET-rate constants. The effect of  $k_{dyn}$  on the resulting apparent distances  $R_{app}^{(a)}$ ,  $R_{app}^{(b)}$  is illustrated in see Figure 9B for an average distance and width comparable to simulations in Figure 9A. With increasing  $k_{dyn}$  both apparent distances of the two-state model decrease. However, as the

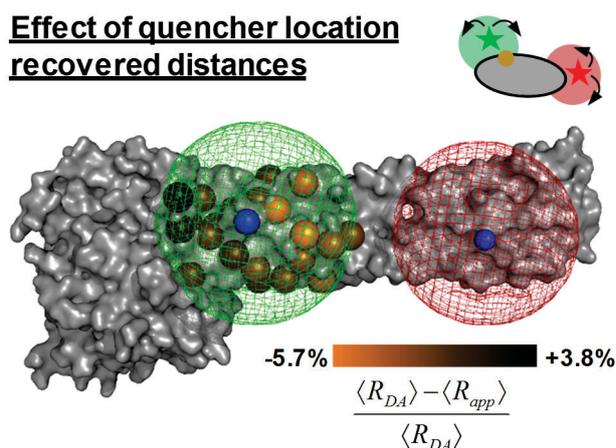
amplitude of the shorter distance decreases faster with increasing  $k_{dyn}$  compared to the amplitude of the longer distance, the apparent width (Figure 9B, lower panel) decreases stronger compared to the average distance (Figure 9B, upper panel). In the shown two-state model a decrease of 3 Ångström in width, as found for the qtAV simulations, corresponds to  $k_{dyn} \sim 0.1 \text{ ns}^{-1}$  (arrow in lower panel, Figure 9B).

As in the simulations the average distance of  $p(R_{app})$  and is only shifted by at most by 2 Å compared to  $p(R_{DA})$ , the average of  $p(R_{DA})$  can be approximated by the average of  $p(R_{app})$ . Relative differences of the width between  $p(R_{app})$  and  $p(R_{DA})$  are more pronounced – the width of  $p(R_{app})$  is  $\sim 3 \text{ Å}$  narrower compared to  $p(R_{DA})$ . This is in line with experiments which report width of  $p(R_{app})$  which are usually 1-3 Å narrower than the width determined by MD or AV (Sindbert et al., 2011, Hofig et al., 2014). Previously, this was attributed to the uniform weighting of the AV (Hofig et al., 2014). However, dye diffusion describes such narrowing equally well. To our knowledge no results on the shape of  $p(R_{app})$  of tethered dyes have been published so far. We demonstrate that dye diffusion influences the shape of  $p(R_{app})$  and cause “ghost peaks” which might be misinterpreted as a minor states. Finally, these results might be also applicable to cytosolic GFPs as comparable diffusion coefficients ( $8.7 \pm 0.2 \text{ Å}^2/\text{ns}$ ) been reported (Swaminathan et al., 1997).

### 2.2.4 Impact of dye diffusion and quenching on FRET

Next, we study the effect of quenching amino acids within the AV of the donor dye on the average of the apparent distance distribution. First all quenching amino-acids in proximity of the donor dye were removed. Next, a single amino acid in proximity of the AV was replaced by a tryptophan and the donor fluorescence decays in presence and absence of FRET was simulated by the qtAV dye model. Using these fluorescence decays the homogenous approximation was applied to recover an average apparent distance  $\langle R_{app} \rangle$ . Finally, the deviation between  $\langle R_{app} \rangle$  and of the average distance  $\langle R_{DA} \rangle$  was calculated. Overall 23 simulations with a quencher located at different positions were performed. For each simulation the calculated deviation was mapped to the side-chain of the quenching (see Figure 10).

### Effect of quencher location recovered distances



**Figure 10.** The effect of the quencher location on the mean apparent distance between the donor D and acceptor A ( $\langle R_{app} \rangle$ ) is illustrated using a crystal structure (PDB:1F5N) of the human guanylate binding protein 1. A set of 23 simulations (quencher located at acid number: 156, 158, 299, 313, 317, 321, 325, 326, 329, 336, 329, 336, 374, 378, 379, 382, 387, 390, 393, 524, 532, 538, 539, 542) was performed. The simulations consider dye diffusion and D quenching. The relative distance difference between the average distance ( $\langle R_{DA} \rangle$ ) (52 Å) and the average apparent distance ( $\langle R_{app} \rangle$ ) was mapped to the C $_{\beta}$ -atom of the respective quencher. The D and A accessible volume are shown as green and red net, respectively. The blue spheres mark the attachment points of D (F379C) and A (D467C).

As highlighted by Figure 10 a quenching amino acid within the donor AV results in systematic differences between  $\langle R_{DA} \rangle$  and  $\langle R_{app} \rangle$ . If the quencher is located in closer to the acceptor (correlated quenching)  $\langle R_{app} \rangle$  overestimates  $\langle R_{DA} \rangle$  (55 Å instead of 52 Å). Otherwise, the  $\langle R_{DA} \rangle$  is underestimated by  $\langle R_{app} \rangle$  (50 Å instead of 52 Å). In the latter case FRET and quenching are anti-correlated. This effect of a quencher on  $\langle R_{app} \rangle$  can be easily rationalized. If the dye is located in proximity of a quencher less fluorescent light is emitted. Thus, if a quencher is located in a high-FRET region less fluorescence of the high FRET species is emitted. Consequently, in average the distance between the donor and acceptor will be over-estimated. If a quencher is located in a low-FRET region the average distance will be under estimated, as the low FRET species emits less fluorescence as expected.

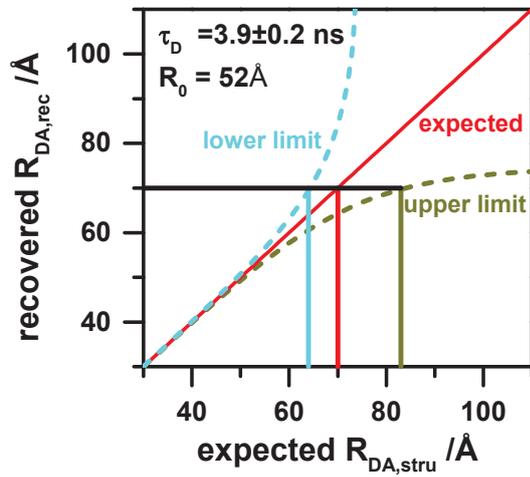
## 2.3 Error-estimation

### 2.3.1 Accuracy using a standard reference

Without knowing the fluorescence decay of the donor in absence of FRET, distances and distance distributions cannot be resolved by the shape of the fluorescence decay, even if the FRET-sensitized acceptor emission is monitored (the convolution of the donor fluorescence decay in presence of FRET and the acceptor fluorescence decay). Hence, always a donor reference sample in absence of FRET is needed. Such reference sample introduces systematic errors, if the fluorescence lifetime of the donor  $\tau_D$  in absence of FRET deviates by  $\Delta\tau_D$  from corresponding lifetime in the FRET-sample. Such deviation  $\Delta\tau_D$  propagates to the donor acceptor distance  $R_{DA}$ . If  $\tau_D$  is determined with a uncertainty  $\Delta\tau_D$  the recovered distance will be within the bounds  $R_{DA,\pm}$ :

$$R_{DA,\pm} = R_0 \left( \left( \frac{R_0}{R_{DA}} \right)^6 \pm \frac{\tau_D}{\Delta\tau_D} \right)^{-1/6} \quad \text{or} \quad \frac{R_{DA,\pm}}{R_{DA}} = \left( 1 \pm \left( \frac{R_{DA}}{R_0} \right)^6 \cdot \frac{\tau_D}{\Delta\tau_D} \right)^{-1/6} \quad (23)$$

To visualize the factor  $R_{DA,\pm} / R_{DA}$  by which the recovered distance deviates from  $R_{DA}$  the upper and lower error bounds  $R_{DA,\pm}$  are shown in dependence of  $R_{DA}$  in Figure 11 for a fluorescence lifetime  $k_D^{-1} = 3.9$  ns and an uncertainty of  $\Delta k_D^{-1} = 0.2$  ns. This fluorescence lifetime and uncertainty correspond to the average and the standard deviation of the references shown in Figure 1C, respectively. As highlighted in Figure 11 a uncertain reference mainly influences long distances. For a reference fluorescence lifetime that mismatches the correct fluorescence lifetime by 200 ps significant systematic errors (bigger than 10%) anticipated if  $R_{DA}$  is bigger than 65 Å.



**Figure 11.** Expected distance  $R_{DA,STRU}$  from structure for an experimentally recovered distance single-distance  $R_{DA,REC}$  of the FRET-pair Alexa488/Alexa647N (Förster-radius 52 Å) in case no sample specific reference was taken and the average donor lifetime  $\tau_{D(0)} = 3.9 \pm 0.2$  ns representative for the set of proteins as shown in Figure 1 is taken as reference. The cyan line defines the lower, the yellow line the upper limit for a given measured distance. The red line is the most likely distance.

In intensity based measurements rely on a calibrated detection efficiency ratio  $\gamma$  and the acceptor quantum yield  $\Phi_F^{A0}$  to determine a DA-distance  $R_{DA} = R_{0r} \left( \Phi_F^{A0} \cdot \gamma \cdot F_{D|D}^{(DA)} / F_{A|D}^{(DA)} \right)^{1/6}$ . Only considering the calibration parameters  $\gamma$  and  $\Phi_F^{A0}$  the relative error of  $\Delta R_{DA} / R_{DA}$  is constant:

$$\frac{\Delta R_{DA}}{R_{DA}} = \frac{1}{6} \left( \left( \frac{\Delta\gamma}{\gamma} \right)^2 + \left( \frac{\Delta\Phi_F^{A0}}{\Phi_F^{A0}} \right)^2 \right)^{1/2} \quad (24)$$

As in intensity based measurements the relative error  $\Delta R_{DA} / R_{DA}$  scales only weakly with the distance  $R_{DA}$  they are less demanding with respect to the sample quality. This comes at the cost of instrument calibrations and averaging over the integration time.

### 2.3.2 Accuracy with sample specific reference

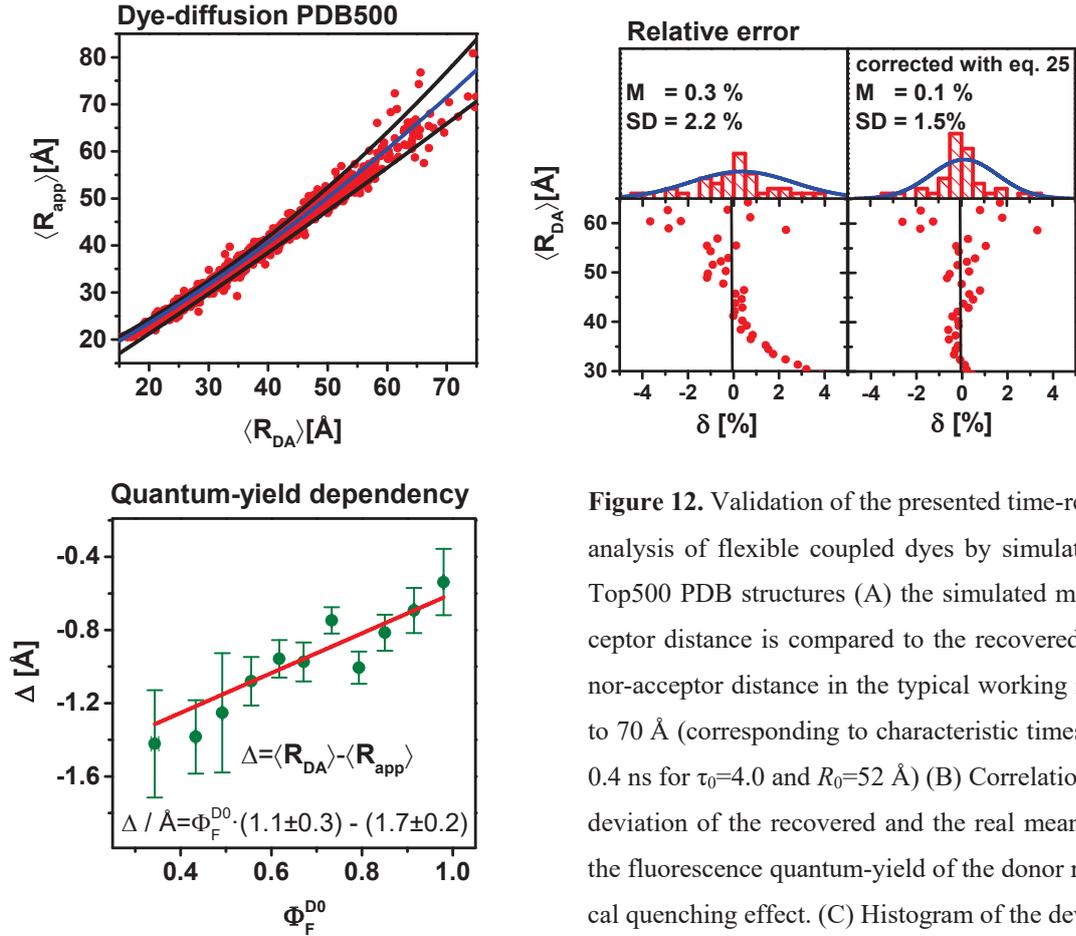
Quenchers within the accessible volume of a donor not explicitly accounted for in the analysis of the fluorescence decays increase the uncertainty of the recovered DA-distances. As demonstrated above (section 2.2.4) a single quencher may result in systematic deviations of the recovered mean DA distance  $\langle R_{app} \rangle$  compared to the mean donor acceptor distance  $\langle R_{DA} \rangle$ . Practically, it is not feasible to explicitly account the quenchers, as their spatial distribution is usually unknown. Therefore, we test for flexible coupled dyes the accuracy of the homogenous approximation by comparing the average recovered apparent distance  $\langle R_{app} \rangle$  to the mean donor acceptor distance  $\langle R_{DA} \rangle$  using the qtAV dye model in simulated experiments.

To quantify the resulting typical deviations between  $\langle R_{app} \rangle$  and  $\langle R_{DA} \rangle$  we used the Top500 (Lovell et al., 2003) as representative set of protein structures. For every structure within the Top500 six surface accessible labeling positions were chosen at random. Next, trajectories of the donor and acceptor dyes were simulated using the parameters as reported in Figure 8. Afterwards, we calculate fluorescence intensity decays of the donor in absence and presence of the acceptor and the corresponding the FRET-induced donor decay  $\varepsilon_D(t)$ . Finally, apparent distance distributions  $p(R_{app})$  were calculated using  $\varepsilon_D(t)$  and equation (16) to compare  $\langle R_{app} \rangle$ , the mean of  $p(R_{app})$ , to the mean donor-acceptor distance  $\langle R_{DA} \rangle$ .

Both means  $\langle R_{app} \rangle$  and  $\langle R_{DA} \rangle$  are compare in Figure 12. To quantify relate both, we have to consider that the distances are not uniformly sampled and short distances ( $<40 \text{ \AA}$ ) are overrepresented. Therefore, we bin the range from  $30 \text{ \AA}$  to  $65 \text{ \AA}$ , and use the mean of each bin to approximate the relation between  $\langle R_{app} \rangle$  and  $\langle R_{DA} \rangle$  an empirical polynomial conversion function:

$$\left( \frac{\langle R_{DA} \rangle}{[\text{\AA}]} \right) = 1.17 \cdot \left( \frac{\langle R_{app} \rangle}{[\text{\AA}]} \right) - 0.0013 \cdot \left( \frac{\langle R_{app} \rangle}{[\text{\AA}]} \right)^2 - 5.11 \quad (25)$$

$$\left( \frac{\langle R_{app} \rangle}{[\text{\AA}]} \right) = 0.60 \cdot \left( \frac{\langle R_{DA} \rangle}{[\text{\AA}]} \right) + 0.0040 \cdot \left( \frac{\langle R_{DA} \rangle}{[\text{\AA}]} \right)^2 + 9.59$$



**Figure 12.** Validation of the presented time-resolved FRET analysis of flexible coupled dyes by simulation using the Top500 PDB structures (A) the simulated mean donor-acceptor distance is compared to the recovered apparent donor-acceptor distance in the typical working range of 35 Å to 70 Å (corresponding to characteristic times of 25 ns and 0.4 ns for  $\tau_0=4.0$  and  $R_0=52$  Å) (B) Correlation between the deviation of the recovered and the real mean distance and the fluorescence quantum-yield of the donor reduced by local quenching effect. (C) Histogram of the deviations of the recovered apparent mean distance  $\langle R_{app} \rangle$  from the true mean distance used to estimate the error of the method. A clear shift towards longer distances is visible (mean=1.0 Å, standard deviations=1.7 Å)

As visible in Figure 12A  $\langle R_{app} \rangle$  and  $\langle R_{DA} \rangle$  follow nearly a 1:1 relationship within the typical working range of time-resolved FRET (30 Å to 65 Å). To quantify deviations we use the difference  $\Delta$  and the relative deviation  $\delta$ :

$$\Delta = \langle R_{DA} \rangle - \langle R_{app} \rangle \quad (26)$$

$$\delta = \frac{\langle R_{DA} \rangle - \langle R_{app} \rangle}{\langle R_{DA} \rangle}$$

In Figure 12B the relative difference  $\delta$  is shown in dependence of  $\langle R_{DA} \rangle$ . Without further correction a relative difference of  $\delta \sim 4\%$  is expected. However, at for short distances  $\langle R_{DA} \rangle$  below 30 Å and long  $\langle R_{DA} \rangle$  above 55 Å significant  $\langle R_{app} \rangle$  differs systematically from  $\langle R_{DA} \rangle$ . By correcting these deviations using the polynomial given by equation (25) in an relative average error of  $\delta = 1.5\%$  is expected. To relate the deviation  $\Delta$  is plotted in dependence of the donor fluorescence quantum yield (see Figure 12C). This reveals a clear trend: For small donor quantum

yields the distances are systematically over-estimated. However, even in worst case scenarios this effect is smaller than 2 Å.

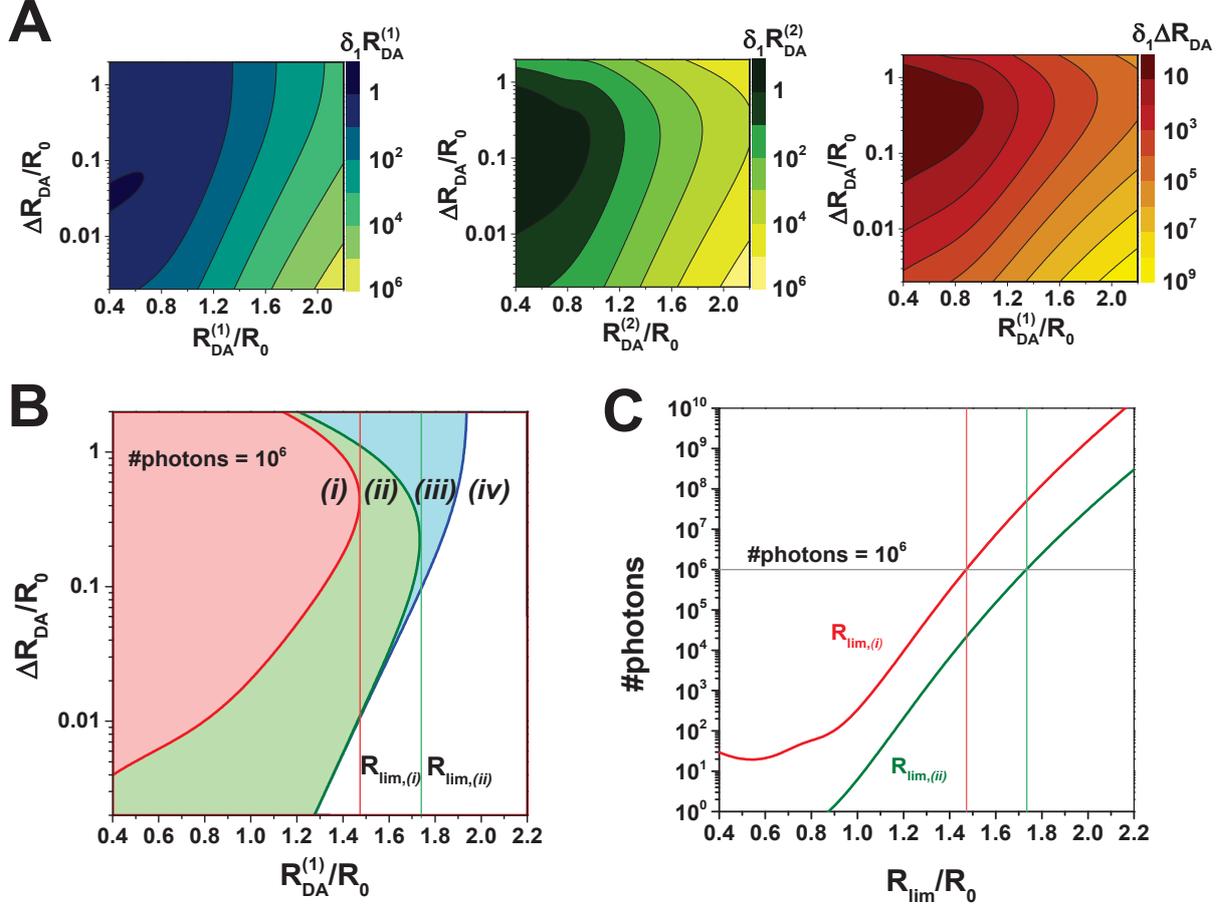
### 2.3.3 Statistical precision

Finally, we estimate the statistical errors for a given noise level of the experiment. Practically, this can be accomplished by exhaustive sampling of the model parameter space (Vöpel et al., 2014, Straume et al., 2002). However, to stress the fundamental limitations, we consider the simplest possible distance distribution and ask under which condition this distribution is still resolved. Our distribution is a two component system with equal species fractions and FRET distances  $R_{DA,1}$  and  $R_{DA,2}=R_{DA,1}+\Delta R_{DA}$ . For such a system the shape of the donor fluorescence decay of the donor in presence of the acceptor is given by:

$$f_{D|D}^{(DA)}(t, R_{DA,1}, R_{DA,2}) = e^{-k_D \left[ 1 + \left( \frac{R_0}{R_{DA,1}} \right)^6 \right] t} + e^{-k_D \left[ 1 + \left( \frac{R_0}{R_{DA,2}} \right)^6 \right] t} \quad (27)$$

In time-correlated single photon counting (TCSPC) the noise is determined by the number of detected photons. Therefore, using this model we address the questions: (1) how many photons have to be measured at least to determine  $R_{DA,1}$  and  $R_{DA,2}$  (or  $\Delta R_{DA}$ ) with a given confidence and (2) what is the distance  $R_{DA,lim}$  at which any incremented, second distance  $R_{DA,lim}+\Delta R_{DA}$  is unresolved.

To answer these questions, we estimate a lower limit of the statistical variances using the Cramér–Rao inequality and the Fisher information matrix (FIM). The results of this estimation for a fluorophore pair with a Förster radius 50Å and a donor lifetime of  $k_D^{-1}=4$  ns are presented in Figure 13. Relative errors ( $\delta_1(R) = \sqrt{\text{var}_1(R)} / R$ ) of  $R_{DA,1}$ ,  $R_{DA,2}$  and  $\Delta R_{DA}$  per counted photon are shown in Figure 13A. To determine errors for another number of counted photons  $N_C$  the presented values should be divided by square root of  $N_C$ . Here we assume that a parameter is “reliably” estimated if its relative error is smaller than 0.5. Figure 12B shows isolines for  $\delta = 0.5$  and detected  $N_C = 10^6$  photons for  $R_{DA,1}$ ,  $R_{DA,2}$  and  $\Delta R_{DA}$  in case if. Using this  $\delta = 0.5$  as criterion, we partition the parameter space into four regions (Figure 13B). In the region (i) all three parameters ( $R_{DA,1}$ ,  $R_{DA,2}$  and  $\Delta R_{DA}$ ) are resolved. In the region (ii) the distances  $R_{DA,1}$ ,  $R_{DA,2}$  are reliably determined while the error of their difference  $\delta(\Delta R_{DA})$  increases above value 0.5. In the region (iii) only the shortest distance  $R_{DA,1}$  can be determined. This means that in the region (iii) the distance distribution is only partially resolved, i.e. species with smaller FRET rate constants cannot be distinguished from non-FRET species. Finally, in the region (iv) none of the parameters are resolved.



**Figure 13.** Statistical error estimates of a two distance model with distances  $R_{DA,1}$  and  $R_{DA,2}=R_{DA,1}+\Delta R_{DA}$ , fluorescence lifetime 4ns and a time-window of  $12.5 \cdot \tau_{D(0)}$ . (A) Relative error  $\delta_1$  per counted photon of the distances  $R_{DA,1}$ ,  $R_{DA,2}$  and their difference  $\Delta R_{DA}$  of per counted. (B) Isolines  $\delta_1(R_{DA,1})=0.5$  (blue line),  $\delta_1(R_{DA,2})=0.5$  (green line) and  $\delta_1 \Delta R_{DA}=0.5$  (red line) for  $10^6$  counted photons. The isolines partition the parameter space in four regions (i) All three parameters are resolved, (ii) the distances  $R_{DA,1}$ ,  $R_{DA,2}$  can be reliably determined while the error of their difference  $\delta_1(\Delta R_{DA})$  increases above value 0.5, (iii) only the shortest distance  $R_{DA,1}$  can be determined. (iv) none of parameters is resolved. Vertical lines indicate the limiting distances  $R_{DA,lim}$  for parameters  $R_{DA,2}$  (green) and  $\Delta R_{DA}$  (red). (C) Limiting distances  $R_{DA,lim}$  for a given number of detected photons for parameters  $R_{DA,2}$  (green) and  $\Delta R_{DA}$  (red).

The isolines of the relative errors  $\delta_1(\Delta R_{DA})$  and  $\delta_1(R_{DA,2})$  reveal a special distance  $R_{DA,1}$ , above which neither the difference between two distances  $\Delta R_{DA}$  nor the second longer distance  $R_{DA,2}$  is resolved. The dependence of this limiting distances  $R_{DA,lim}$  on the number of detected photons  $N_C$  is shown in Figure 13C. Remarkably, this maximum resolvable distance depend only weakly (nearly logarithmically) on the number of detected photons. For instance, to estimate the parameters  $R_{DA,1} \approx 1.2R_0$  and  $\Delta R_{DA} \approx 0.5R_0$  roughly  $N_C = 10^4$  photons have to be detected. At a count-rate of 100 kHz such a limiting distance ( $1.2R_0$ ) is already achieved within 0.1 seconds. However, for a limiting distance of  $R_{DA,lim} = 1.5R_0$  already two orders of magnitude more photons

$N_c = 10^6$  have to be detected. This sets practical limitation, which in near future may be overcome for instance by CMOS-SPAD arrays which allow for significant higher count rates (Krstajic et al., 2015). Note that the presented values are lower bounds of the real errors. In practice the errors will be bigger, if additionally, the fractions of components have to be determined and further experimental nuisance parameters are considered.

### **3 Conclusion**

We presented an analysis method for time-resolved FRET measurements which rationalizes the use of appropriate references and the global analysis of the donor fluorescence decays. By introducing the FRET-induced donor decay we visualize donor acceptor distances distributions. This facilitates significantly the communication of experimental results and allows non-experts to interpret time-resolved experiments. We generalized this approach to multi-exponential donor fluorophores and consider correlations between the FRET quenching the donor by other pathways. Such correlations introduce ambiguity into the interpretation of the fluorescence decays, if the connectivity between donor- and FRET-states is unknown. On an example we showed that such correlations are of minor importance for majorly populated conformational states.

We addressed three different types of errors that may arise in data-analysis: systematic errors due to inappropriate reference samples, the influence of dye distributions and finally statistical errors. Flexible coupled dyes may be dynamically quenched by their local environment, which may affect the recovered distances. We studied this effect by unifying the accessible volume (AV) approach, which quantifies the conformational space of the dye, with dynamic simulations and structural information of protein. This allowed us to estimate the fluorescence properties for given structural models and predict dynamic quenching semi-quantitatively. Using this approach, we studied the effect of diffusion and dynamic quenching on FRET. In average the experimentally recovered distributions decreased in width by approximately 2-3 Å and are shifted by approximately 1 Å towards longer distances whereas the mean distances are recovered with an accuracy of  $\pm 1.7$  Å. This demonstrates for a given conformational state of a protein that the approximation that all fluorescence lifetimes of the donor are quenched by the same FRET rate constant distribution works very well. In typical FRET-measurements a main error source are inappropriate reference samples, in particular, if distances longer than the Förster radius are studied.

## 4 Supplement

### Tables

**Table S1.** Analogy between FRET and anisotropy

	<b>FRET</b>	<b>Anisotropy</b>
Definition	$\varepsilon(t) = \int p(k_{FRET}) \cdot e^{-t \cdot k_{FRET}} \cdot dk_{FRET}$	$r(t) = \int p(\rho_i) \cdot e^{-t/\rho_i} \cdot d\rho_i$
Signal	$f_{D D}^{(D,A)}(t) = N_0^{(D,A)} \frac{1}{N_0^{(D,0)}} f_{D D}^{(D,0)}(t) \cdot \varepsilon(t)$	$f_{\perp}(t) = f_M(t) \cdot (1/3 - 1/3 \cdot r(t))$ $f_{\parallel}(t) = f_M(t) \cdot (1/3 + 2/3 \cdot r(t))$
Reference	$f_{D D}^{(D,0)}(t)$	$f_M(t)$
	$\varepsilon(t) = \frac{N_0^{(D,0)} f_{D D}^{(D,A)}(t)}{N_0^{(D,A)} f_{D D}^{(D,0)}(t)}$	$r(t) = \frac{f_{\parallel}(t) - f_{\perp}(t)}{f_{\parallel}(t) + 2 \cdot G \cdot f_{\perp}(t)}$
Steady-state	$E_{SS} = 1 - \frac{\int \varepsilon(t) \cdot f_{D D}^{(D,0)}(t) dt}{\int f_{D D}^{(D,0)}(t) dt}$	$r_{SS} = \frac{\int r(t) \cdot f_M(t) dt}{\int f_M(t) dt}$
single exponential donor	$E_{SS} = 1 - \frac{\tau_{(D,0)} - \tau_{(D,A)}}{\tau_{(D,0)}}$	$r_{SS} = \frac{r_0}{1 + \tau/\rho}$ (Perrin equation)

**Table S2.** Simulation parameters

Name	Value [Å]
Linker length/ /	22.0
dye radius	3.0
linker width	2.0
Critical distance	dye radius + 5.5
Slow radius	8.0

**Table S4.** Dynamic quenching rate constants of single side-chains

Amino acid	$k_Q(R=0)$ [1/ns]
MET	1.4
TRP	2.8
TYR	2.8
HIS	1.2

**Table S3.** Diffusion coefficients

Name	Value / [Å <sup>2</sup> /ns]
Free donor	10.0
Free acceptor	5.0
Trapped donor	0.9
Trapped acceptor	0.4

### The estimation of statistical-errors

To estimate statistical errors due to the photon noise we use the Fisher information matrix (FIM) approach and the Cramér–Rao inequality. The Cramér–Rao inequality states that the variance-covariance matrix  $\Sigma$  is bigger than the inverse of the FIM  $I$  ( $\Sigma \geq I^{-1}$ ). For two model parameters  $\alpha$  and  $\beta$  the elements of the FIM are given by:

$$I_{\alpha,\beta} = E\left[\partial_{\alpha} \ln \ell(\alpha, \beta) \partial_{\beta} \ln \ell(\alpha, \beta)\right] \quad (28)$$

Here  $\ell(\alpha, \beta)$  is a likelihood function which quantifies the agreement between the model and the experiment. In a TCSPC experiment with  $n$ -detection bins and  $N$  detected photons with  $N_i$  photons per bin the likelihood is given by a multinomial probability mass function:

$$\ell(\alpha, \beta) = N! \prod_{i=1}^n \frac{f_i^{N_i}(\alpha, \beta)}{N_i!} \quad (29)$$

Here  $f_i$  is the probability of detecting a photon in a bin  $i$  assumed by the model-function. Thus,  $Nf_i$  is the expectation value in the bin  $i$ . The probability  $f_i$  is obtained from the continuous model function  $f(t, \alpha, \beta)$  describing the experiment by piecewise integration and normalization:

$$f_i(\alpha, \beta) = \frac{\int_{(i-1)\Delta t}^{i\Delta t} f(t, \alpha, \beta) dt}{\int_0^{n\Delta t} f(t, \alpha, \beta) dt} \quad (30)$$

As the likelihood function is multinomial the FIM takes the simple form:

$$I_{\alpha, \beta} = N \sum_{i=1}^n \frac{\partial_{\alpha} f_i(\alpha, \beta) \partial_{\beta} f_i(\alpha, \beta)}{f_i(\alpha, \beta)} \quad (31)$$

Under these conditions the FIM does not contain the experimental information. Hence, the variances and covariances can be predicted *a priori* given a model function.

## References

- ANDRECKA, J., TREUTLEIN, B., ARCUSA, M. A. I., MUSCHIELOK, A., LEWIS, R., CHEUNG, A. C. M., CRAMER, P. & MICHAELIS, J. 2009. Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex. *Nucleic Acids Research*, 37, 5803-5809.
- BEECHEM, J. M., GRATTON, E., AMELOOT, M., KNUTSON, J. R. & BRAND, L. 2002. The global analysis of fluorescence intensity and anisotropy decay data: Second-generation theory and programs. *Topics in Fluorescence Spectroscopy*. Springer.
- BETZIG, E., PATTERSON, G. H., SOUGRAT, R., LINDWASSER, O. W., OLENYCH, S., BONIFACINO, J. S., DAVIDSON, M. W., LIPPINCOTT-SCHWARTZ, J. & HESS, H. F. 2006. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313, 1642-1645.
- BEUTLER, M., MAKROGIANNELI, K., VERMEIJ, R. J., KEPPLER, M., NG, T., JOVIN, T. M. & HEINTZMANN, R. 2008. satFRET: estimation of Förster resonance energy transfer by acceptor saturation. *European Biophysics Journal*, 38, 69-82.
- BROCHON, J. C. 1994. Maximum-Entropy Method of data-analysis in time-resolved spectroscopy. *Methods in Enzymology*, 240, 262-311.
- BUSCHMANN, V., WESTON, K. D. & SAUER, M. 2003. Spectroscopic study and evaluation of red-absorbing fluorescent dyes. *Bioconjugate Chemistry*, 14, 195-204.
- CHEN, H. M., AHSAN, S. S., SANTIAGO-BERRIOS, M. B., ABRUNA, H. D. & WEBB, W. W. 2010. Mechanisms of quenching of Alexa fluorophores by natural amino acids. *Journal of the American Chemical Society*, 132, 7244-7245.

- CHIBISOV, A. K., ZAKHAROVA, G. V., GÖRNER, H., SOGULYAEV, Y. A., MUSHKALO, I. L. & TOLMACHEV, A. I. 1995. Photorelaxation processes in covalently-linked indocarbocyanine and thiocarbocyanine dyes. *Journal of Physical Chemistry*, 99, 886-893.
- CLEGG, R. M., MURCHIE, A. I. & LILLEY, D. M. 1994. The solution structure of the four-way DNA junction at low-salt conditions: a fluorescence resonance energy transfer analysis. *Biophysical Journal*, 66, 99-109.
- CLEGG, R. M., MURCHIE, A. I., ZECHEL, A., CARLBERG, C., DIEKMANN, S. & LILLEY, D. M. J. 1992. Fluorescence resonance energy transfer analysis of the structure of the four-way DNA junction. *Biochemistry*, 31, 4846-56.
- DICKSON, R. M., CUBITT, A. B., TSIEN, R. Y. & MOERNER, W. E. 1997. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature*, 388, 355-358.
- DOOSE, S., NEUWEILER, H. & SAUER, M. 2009. Fluorescence quenching by photoinduced electron transfer: a reporter for conformational dynamics of macromolecules. *ChemPhysChem*, 10, 1389-98.
- EGGELING, C., FRIES, J. R., BRAND, L., GÜNTHER, R. & SEIDEL, C. A. M. 1998. Monitoring conformational dynamics of a single molecule by selective fluorescence spectroscopy. *Proceedings of the National Academy of Sciences*, 95, 1556-1561.
- FLANNERY, M. R. 1982. Analytical solution of the Debye-Smoluchowski equation for geminate and homogeneous recombination and for fluorescence quenching. *Physical Review A*, 25, 3403-3406.
- FÖRSTER, T. 1948. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik*, 437, 55-75.
- FÖRSTER, T. 1949. Versuche Zum Zwischenmolekularen Übergang Von Elektronenanregungsenergie. *Angewandte Chemie*, 61, 301-301.
- GAIDUK, A., KUHNEMUTH, R., FELEKYAN, S., ANTONIK, M., BECKER, W., KUDRYAVTSEV, V., SANDHAGEN, C. & SEIDEL, C. A. M. 2007. Fluorescence detection with high time resolution: From optical microscopy to simultaneous force and fluorescence spectroscopy. *Microscopy Research and Technique*, 70, 433-441.
- GENDRON, P. O., AVALTRONI, F. & WILKINSON, K. J. 2008. Diffusion coefficients of several rhodamine derivatives as determined by pulsed field gradient-nuclear magnetic resonance and fluorescence correlation spectroscopy. *Journal of Fluorescence*, 18, 1093-1101.
- GOPICH, I. V. & SZABO, A. 2003. Single-macromolecule fluorescence resonance energy transfer and free-energy profiles. *Journal of Physical Chemistry B*, 107, 5058-5063.
- GRAY, H. B. & WINKLER, J. R. 1996. Electron transfer in proteins. *Annual Review of Biochemistry*, 65, 537-61.
- GRYCZYNSKI, I., WICZK, W., JOHNSON, M. L., CHEUNG, H. C., WANG, C. K. & LAKOWICZ, J. R. 1988. Resolution of end-to-end distance distributions of flexible molecules using quenching-induced variations of the Förster distance for fluorescence energy transfer. *Biophysical Journal*, 54, 577-586.
- HAAS, E., WILCHEK, M., KATCHALSKI-KATZIR, E. & STEINBERG, I. Z. 1975. Distribution of end-to-end distances of oligopeptides in solution as estimated by energy transfer. *Proceedings of the National Academy of Sciences of the United States of America*, 72, 1807-1811.
- HELL, S. W. & WICHMANN, J. 1994. Breaking the Diffraction Resolution Limit by Stimulated-Emission - Stimulated-Emission-Depletion Fluorescence Microscopy. *Optics Letters*, 19, 780-782.
- HOFIG, H., GABBA, M., POBLETE, S., KEMPE, D. & FITTER, J. 2014. Inter-dye distance distributions studied by a combination of single-molecule FRET-filtered lifetime

- measurements and a weighted accessible volume (wAV) Algorithm. *Molecules*, 19, 19269-19291.
- HUANG, K. C. & WHITE, R. J. 2013. Random Walk on a Leash: A Simple Single-Molecule Diffusion Model for Surface-Tethered Redox Molecules with Flexible Linkers. *Journal of the American Chemical Society*, 135, 12808-12817.
- HWANG, H. & MYONG, S. 2014. Protein induced fluorescence enhancement (PIFE) for probing protein-nucleic acid interactions. *Chemical Society Reviews*, 43, 1221-9.
- JACOB, M. H., DSOUZA, R. N., GHOSH, I., NOROUZY, A., SCHWARZLOSE, T. & NAU, W. M. 2013. Diffusion-enhanced Förster resonance energy transfer and the effects of external quenchers and the donor quantum yield. *Journal of Physical Chemistry B*, 117, 185-198.
- JOSHI, N., JOHNSON, M. L., GRZYCZYNSKI, I. & LAKOWICZ, J. R. 1987. Radiation boundary-conditions in collisional quenching of fluorescence - Determination by frequency-domain fluorometry. *Chemical Physics Letters*, 135, 200-207.
- KALININ, S., FELEKYAN, S., VALERI, A. & SEIDEL, C. A. M. 2008. Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis. *Journal of Physical Chemistry B*, 112, 8361-8374.
- KALININ, S., FULLE, S., KÜHNEMUTH, R., SINDBERT, S., FELEKYAN, S., GOHLKE, H. & SEIDEL, C. A. M. 2015. Diffusion with traps: experiment, simulation, and theory to describe fluorophore dynamics in biomolecular FRET.
- KALININ, S., PEULEN, T., SINDBERT, S., ROTHWELL, P. J., BERGER, S., RESTLE, T., GOODY, R. S., GOHLKE, H. & SEIDEL, C. A. M. 2012. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nature Methods*, 9, 1218-25.
- KLEHS, K., SPAHN, C., ENDESFELDER, U., LEE, S. F., FÜRSTENBERG, A. & HEILEMANN, M. 2014. Increasing the brightness of cyanine fluorophores for single-molecule and superresolution imaging. *ChemPhysChem*, 15, 637-641.
- KRSTAJIC, N., POLAND, S., LEVITT, J., WALKER, R., ERDOGAN, A., AMEER-BEG, S. & HENDERSON, R. K. 2015. 0.5 billion events per second time correlated single photon counting using CMOS SPAD arrays. *Optics Letters*, 40, 4305-4308.
- KUDRYAVTSEV, V., SIKOR, M., KALININ, S., MOKRANJAC, D., SEIDEL, C. A. & LAMB, D. C. 2012. Combining MFD and PIE for accurate single-pair Förster resonance energy transfer measurements. *ChemPhysChem*, 13, 1060-78.
- LAPTENOK, S. P., SNELLENBURG, J. J., BUCHERL, C. A., KONRAD, K. R. & BORST, J. W. 2014. Global analysis of FRET-FLIM data in live plant cells. *Methods Mol Biol*, 1076, 481-502.
- LEVITUS, M. & RANJIT, S. 2011. Cyanine dyes in biophysical research: the photophysics of polymethine fluorescent dyes in biomolecular environments. *Quarterly Reviews of Biophysics*, 44, 123-151.
- LIVESEY, A. K. & SKILLING, J. 1985. Maximum-Entropy Theory. *Acta Crystallographica Section A: Foundations of Crystallography*, 41, 113-122.
- LOVELL, S. C., DAVIS, I. W., ARENDALL, W. B., 3RD, DE BAKKER, P. I., WORD, J. M., PRISANT, M. G., RICHARDSON, J. S. & RICHARDSON, D. C. 2003. Structure validation by C $\alpha$  geometry: phi,psi and C $\beta$  deviation. *Proteins*, 50, 437-50.
- MARCUS, R. A. & SUTIN, N. 1985. Electron transfers in chemistry and biology. *Biochimica et Biophysica Acta*, 811, 265-322.
- MAUS, M., COTLET, M., HOFKENS, J., GENSCHE, T., DE SCHRYVER, F. C., SCHAFFER, J. & SEIDEL, C. A. M. 2001. An experimental comparison of the maximum likelihood estimation and nonlinear least squares fluorescence lifetime analysis of single molecules. *Analytical Chemistry*, 73, 2078-2086.

- MCCANN, J. J., ZHENG, L., CHIANTIA, S. & BOWEN, M. E. 2011. Domain orientation in the N-Terminal PDZ tandem from PSD-95 is maintained in the full-length protein. *Structure*, 19, 810-20.
- MCCANN, J. J., ZHENG, L. Q., ROHRBECK, D., FELEKYAN, S., KÜHNEMUTH, R., SUTTON, R. B., SEIDEL, C. A. M. & BOWEN, M. E. 2012. Supertertiary structure of the synaptic MAGuK scaffold proteins is conserved. *Proceedings of the National Academy of Sciences*, 109, 15775-15780.
- MEKLER, V., KORTKHONJIA, E., MUKHOPADHYAY, J., KNIGHT, J., REVYAKIN, A., KAPANIDIS, A. N., NIU, W., EBRIGHT, Y. W., LEVY, R. & EBRIGHT, R. H. 2002. Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell*, 108, 599-614.
- MÖGLICH, A., JODER, K. & KIEFHABER, T. 2006. End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proceedings of the National Academy of Sciences*, 103, 12394-12399.
- MOSER, C. C., KESKE, J. M., WARNCKE, K., FARID, R. S. & DUTTON, P. L. 1992. Nature of biological electron transfer. *Nature*, 355, 796-802.
- MÜLLER, C. B., LOMAN, A., PACHECO, V., KOBERLING, F., WILLBOLD, D., RICHTERING, W. & ENDERLEIN, J. 2008. Precise measurement of diffusion by multi-color dual-focus fluorescence correlation spectroscopy. *EPL*, 83.
- NEUBAUER, H., GAIKO, N., BERGER, S., SCHAFFER, J., EGGELING, C., TUMA, J., VERDIER, L., SEIDEL, C. A. M., GRIESINGER, C. & VOLKMER, A. 2007. Orientational and dynamical heterogeneity of rhodamine 6G terminally attached to a DNA helix revealed by NMR and single-molecule fluorescence spectroscopy. *Journal of the American Chemical Society*, 129, 12746-12755.
- PERIASAMY, N., DORAISWAMY, S., VENKATARAMAN, B. & FLEMING, G. R. 1988. Diffusion controlled reactions - experimental-verification of the time-dependent rate-equation. *Journal of Chemical Physics*, 89, 4799-4806.
- RICE, S. A. 1985. *Diffusion-Limited Reactions*, Amsterdam, Elsevier.
- ROTHWELL, P. J., ALLEN, W. J., SISAMAKIS, E., KALININ, S., FELEKYAN, S., WIDENGREN, J., WAKSMAN, G. & SEIDEL, C. A. 2013. dNTP-dependent conformational transitions in the fingers subdomain of KlenTaq1 DNA polymerase: insights into the role of the "nucleotide-binding" state. *Journal of Biological Chemistry*, 288, 13575-91.
- RUST, M. J., BATES, M. & ZHUANG, X. W. 2006. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, 3, 793-795.
- SAUERWEIN, B., MURPHY, S. & SCHUSTER, G. B. 1992. Dynamics of solute motion - photoisomerization shows linear-dependence on solvent mass. *Journal of the American Chemical Society*, 114, 7920-7922.
- SCHAFFER, J., VOLKMER, A., EGGELING, C., SUBRAMANIAM, V., STRIKER, G. & SEIDEL, C. A. M. 1999. Identification of single molecules in aqueous solution by time-resolved fluorescence anisotropy. *Journal of Physical Chemistry A*, 103, 331-336.
- SEIDEL, C. A. M., SCHULZ, A. & SAUER, M. H. M. 1996. Nucleobase-specific quenching of fluorescent dyes 1. Nucleobase one-electron redox potentials and their correlation with static and dynamic quenching efficiencies. *Journal of Physical Chemistry*, 100, 5541-5553.
- SHEATHER, S. J. & JONES, M. C. 1991. A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Statist. Soc. B*, 53, 683-690.
- SINDBERT, S., KALININ, S., NGUYEN, H., KIENZLER, A., CLIMA, L., BANNWARTH, W., APPEL, B., MÜLLER, S. & SEIDEL, C. A. M. 2011. Accurate distance

- determination of nucleic acids via Förster resonance energy transfer: implications of dye linker length and rigidity. *Journal of the American Chemical Society*, 133, 2463-80.
- SISAMAKIS, E., VALERI, A., KALININ, S., ROTHWELL, P. J. & SEIDEL, C. A. M. 2010. Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods in Enzymology*, 475, 455-514.
- SOMSSICH, M., MA, Q., WEIDTKAMP-PETERS, S., STAHL, Y., FELEKYAN, S., BLECKMANN, A., SEIDEL, C. A. M. & SIMON, R. 2015. Real-time dynamics of peptide ligand-dependent receptor complex formation in planta. *Science Signaling*, 8, ra76.
- STAHL, Y., GRABOWSKI, S., BLECKMANN, A., KÜHNEMUTH, R., WEIDTKAMP-PETERS, S., PINTO, K. G., KIRSCHNER, G. K., SCHMID, J. B., WINK, R. H., HULSEWEDE, A., FELEKYAN, S., SEIDEL, C. A. M. & SIMON, R. 2013. Moderation of Arabidopsis root stemness by CLAVATA1 and ARABIDOPSIS CRINKLY4 receptor kinase complexes. *Current Biology*, 23, 362-71.
- STRAUME, M., FRASIER-CADORET, S. G. & JOHNSON, M. L. 2002. Least-Squares Analysis of Fluorescence Data. *Topics in Fluorescence Spectroscopy*. Springer.
- STRIKER, G., SUBRAMANIAM, V., SEIDEL, C. A. M. & VOLKMER, A. 1999. Photochromicity and fluorescence lifetimes of green fluorescent protein. *Journal of Physical Chemistry B*, 103, 8612-8617.
- STRYER, L. & HAUGLAND, R. P. 1967. Energy transfer: a spectroscopic ruler. *Proceedings of the National Academy of Sciences*, 58, 719-26.
- SUN, Y. S., DAY, R. N. & PERIASAMY, A. 2011. Investigating protein-protein interactions in living cells using fluorescence lifetime imaging microscopy. *Nature Protocols*, 6, 1324-1340.
- SUSTARSIC, M. & KAPANIDIS, A. N. 2015. Taking the ruler to the jungle: single-molecule FRET for understanding biomolecular structure and dynamics in live cells. *Current Opinions in Structural Biology*, 34, 52-59.
- SWAMINATHAN, R., HOANG, C. P. & VERKMAN, A. S. 1997. Photobleaching recovery and anisotropy decay of green fluorescent protein GFP-S65T in solution and cells: Cytoplasmic viscosity probed by green fluorescent protein translational and rotational diffusion. *Biophysical Journal*, 72, 1900-1907.
- VAN DER MEER, B. W., RAYMER, M. A., WAGONER, S. L., HACKNEY, R. L., BEECHEM, J. M. & GRATTON, E. 1993. Designing matrix models for fluorescence energy transfer between moving donors and acceptors. *Biophysical Journal*, 64, 1243-1263.
- VAN DER MEER, W. B., VAN DER MEER, D. M. & VOGEL, S. S. 2014. Optimizing the orientation factor kappa-squared for more accurate FRET measurements. In: MEDINTZ, I. & HILDEBRANDT, N. (eds.) *FRET - Förster Resonance Energy Transfer: From Theory to Applications*. Weinheim, Germany: Wiley-VCH Verlag.
- VOLKMER, A., SUBRAMANIAM, V., BIRCH, D. J. S. & JOVIN, T. M. 2000. One- and two-photon excited fluorescence lifetimes and anisotropy decays of green fluorescent proteins. *Biophysical Journal*, 78, 1589-1598.
- VÖPEL, T., HENGSTENBERG, C. S., PEULEN, T.-O., AJAJ, Y., SEIDEL, C. A. M., HERRMANN, C. & KLARE, J. P. 2014. Triphosphate induced dimerization of human guanylate binding protein 1 involves association of the C-terminal helices: A joint double electron-electron resonance and FRET study. *Biochemistry*, 53, 4590-4600.
- WEIDTKAMP-PETERS, S., FELEKYAN, S., BLECKMANN, A., SIMON, R., BECKER, W., KÜHNEMUTH, R. & SEIDEL, C. A. M. 2009. Multiparameter fluorescence image spectroscopy to study molecular interactions. *Photochemical & Photobiological Sciences*, 8, 470-480.

- WESTLUND, P. O. & WENNERSTROM, H. 1993. Electronic-energy transfer in liquids - the effect of molecular dynamics. *Journal of Chemical Physics*, 99, 6583-6589.
- WIDENGREN, J., KUDRYAVTSEV, V., ANTONIK, M., BERGER, S., GERKEN, M. & SEIDEL, C. A. M. 2006. Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. *Analytical Chemistry*, 78, 2039-2050.
- WIDENGREN, J., SCHWEINBERGER, E., BERGER, S. & SEIDEL, C. A. M. 2001. Two new concepts to measure fluorescence resonance energy transfer via fluorescence correlation spectroscopy: Theory and experimental realizations. *Journal of Physical Chemistry A*, 105, 6851-6866.
- WIDENGREN, J. & SCHWILLE, P. 2000. Characterization of photoinduced isomerization and back-isomerization of the cyanine dye Cy5 by fluorescence correlation spectroscopy. *Journal of Physical Chemistry A*, 104, 6416-6428.
- WOZNIAK, A. K., SCHRÖDER, G. F., GRUBMÜLLER, H., SEIDEL, C. A. M. & OESTERHELT, F. 2008. Single-molecule FRET measures bends and kinks in DNA. *Proceedings of the National Academy of Sciences*, 105, 18337-18342.
- YIRDAW, R. B. & MCHAOURAB, H. S. 2012. Direct observation of T4 lysozyme hinge-bending motion by fluorescence correlation spectroscopy. *Biophysical Journal*, 103, 1525-1536.

## Chapter B - Experimental reference samples for accurate time-resolved FRET analysis Handling positional effects of fluorescent dyes in bis-functional proteins

**Carola S. Hengstenberg<sup>†\*</sup>, Thomas-Otavio Peulen<sup>§\*</sup>, Jakub Kubiak<sup>§</sup>, Markus Richert<sup>§</sup>, James J. McCann<sup>%</sup>, Mark Bowen<sup>%</sup>, Christian Herrmann<sup>†</sup>, Claus A.M. Seidel<sup>§</sup>**

<sup>†</sup>Physical Chemistry I, Faculty of Chemistry and Biochemistry, Ruhr-University Bochum, Universitätsstr. 150, 44780 Bochum, Germany <sup>§</sup>Chair for Molecular Physical Chemistry, Heinrich-Heine University, Universitätsstr. 1, 40225 Düsseldorf, Germany, <sup>%</sup>Department of Physiology & Biophysics Stony Brook University Basic Science Tower 5-123 Stony Brook, New York

\*Contributed equally

### 1. Introduction

Förster Resonance Energy Transfer (FRET) describes the distance and orientation dependent dipolar coupling between fluorophores (Förster, 1948'). By exploiting the distance dependency FRET-measurements serve as spectroscopic ruler (Stryer and Haugland, 1967, Stryer, 1978). As FRET-measurements can be performed in solution at room temperature they are ideally suited to study structural properties of biological macromolecules at near physiological conditions and in living cells FRET is (Vöpel et al., 2014, Sustarsic and Kapanidis, 2015, Stahl et al., 2013). Typically, a distinct fluorophore pair, a donor (D) and an acceptor fluorophore (A), is introduced into the biomolecule to resolved DA-distances by the energy transferred from the D to the A. Recent methodological advances allows to determine accurate structural models including their uncertainties (Kalinin et al., 2012). This is accomplished by multiple FRET-measurement using distinct DA-pairs and by explicitly considering experimental uncertainties, geometrical properties of the fluorophores and the fluorescence properties of the dyes coupled to a biomolecule (Sindbert et al., 2011, Kalinin et al., 2012, Andrecka et al., 2009).

DA-distances quantified are measured by changes of the steady-state intensity or by analyzing time-resolved fluorescence decays. In intensity based experiments the FRET-process is quantified by the FRET-efficiency while time-resolved experiments can be described by the FRET-induced donor decay (see chapter A). Notably, the analysis of both experiment types relies on reference samples. To determine distance by fluorescence intensities the recorded signal has to be corrected for spectral cross-talk, instrument specific detection efficiencies and the quantum yield of the dyes (Rothwell et al., 2003). To separate the effect of FRET from other deexcitation pathways of the excited fluorescent state in time-resolved methods the fluorescence intensity decay of the donor in absence of the acceptor is required.

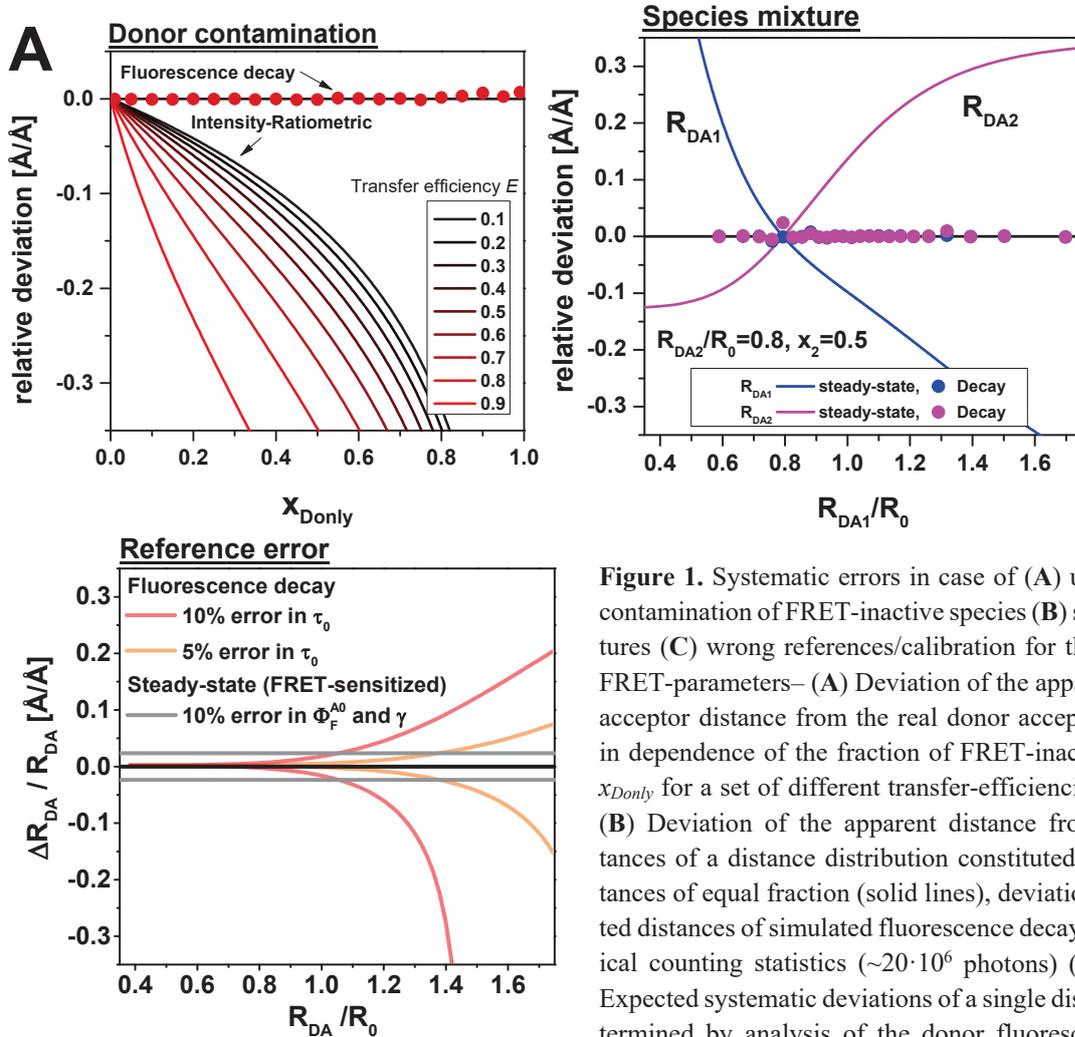
If solely the intensities are analyzed, sample heterogeneities introduce uncertainties on the determined DA-distance, as the intensities are quantity averaged over sample heterogeneities.

Sample heterogeneities are for instance caused FRET-inactive molecules or conformational heterogeneity of the FRET-sample. Thus, even though species mixtures can be partly addressed in ensemble measurements, i.e. FRET-inactive species by the degree of labeling (Clegg et al., 1992), distinct FRET-species are only resolved by the FRET-efficiency under single-molecule conditions for instance by histogram analysis (Antonik et al., 2006). The effect of FRET-inactive molecules on the DA-distance recovered by the FRET-efficiency is shown in Figure 1A. If the fraction of FRET-inactive molecules,  $x_{DOnly}$ , is not considered the fluorescence intensities recover only an apparent DA-distances deviating from the correct DA-distance. In Figure 1A such deviation is displayed in dependence of  $x_{DOnly}$ . In contrast to fluorescence intensity based measurements time-resolved fluorescence measurements consider the waiting time distribution between excitation and the emission of a fluorescent photon. By analyzing experimental fluorescence decay curves by multicomponent models DA-distances can be resolved, FRET-inactive species are discriminated and their fraction is quantified by  $x_{DOnly}$  (Möglich et al., 2006)(see chapter A). Thus, the deviations between the DA-distance and the recovered DA-distances are small (Figure 1A).

The temporal resolution is limited by the count-rate of the experiment. This may limit the spatial resolution determined by the average FRET-efficiency, if the molecules observed under single-molecule conditions undergo fast conformational dynamics (Kalinin et al., 2010, Gopich and Szabo, 2003). The effect of mixing two distinct FRET-states with two distinct DA-distance on deviation of apparent DA-distance is illustrated in Figure 1B. Time-resolved fluorescence measurements resolve both DA-distance. This drastically minimizes the error compared to the average FRET-efficiency (Figure 1B).

In steady-state and time-resolved measurements an improper calibration or reference sample introduces systematic deviations of the recovered DA-distances (see Chapter A). However, time-resolved fluorescence measurements are less robust, in particular for long distances, compared to the distances derived by fluorescence intensities, as reference errors may result in large systematic errors. This is illustrated in Figure 1C where the relative error of a single DA-distance determined by fluorescence intensities and the fluorescence decays are plotted in dependence of the respective DA-distance. In both cases if the relevant calibration parameters deviate by 10 percent from the correct value. The sample specific references in steady-state measurements and time-resolved measurements are the acceptor quantum yield and the fluorescence decay of the donor in absence of the acceptor, respectively. Thus, even though time-resolved

measurements potentially allow resolving species mixtures on calibration free setups, they require reference samples of high quality in particular to resolve long distances. As the fluorescence quantum yields of the and the fluorescence decays are sample dependent and modulated by the local environment of the dyes experimental reference samples of high quality are needed.



**Figure 1.** Systematic errors in case of (A) unaccounted contamination of FRET-inactive species (B) species mixtures (C) wrong references/calibration for the indicated FRET-parameters– (A) Deviation of the apparent donor-acceptor distance from the real donor acceptor distance in dependence of the fraction of FRET-inactive species  $x_{\text{Donly}}$  for a set of different transfer-efficiencies (eq. (4)). (B) Deviation of the apparent distance from two distances of a distance distribution constituted of two distances of equal fraction (solid lines), deviation of the fitted distances of simulated fluorescence decays given typical counting statistics ( $\sim 20 \cdot 10^6$  photons) (circles) (C) Expected systematic deviations of a single distance as determined by analysis of the donor fluorescence decay

(red, equation (8)) or the steady-state transfer efficiency determined by FRET sensitized signal (eq. (1)) (gray, equation (7)) for given inaccurate calibrations or reference samples.

To determine distances by FRET an ideal set of samples consists of a FRET-sample, where each molecule is labeled exactly by a donor and acceptor with precisely known positions, and a donor-reference, with the same composition as the FRET-sample but missing an acceptor. As the acceptor fluorescence properties can be determined by direct excitation of the FRET-sample, for instance by pulsed interleaved excitation (PIE) (Kudryavtsev et al., 2012), a separate acceptor reference is usually not necessary (Kudryavtsev et al., 2012). In bis-functional proteins with two labeling sites the donor may be located at two different positions. If different labeling chemistry position for the donor and acceptor within a protein are used, the donor position is precisely known and reference samples are easily obtained (Brustad et al., 2008, Ratner et al.,

2002). However, if the same labeling chemistry is used for the donor and acceptor without protection of a single site, neither the donor nor the acceptor location are precisely known. Consequently, a multitude of possible species with different acceptor and donor stoichiometry has to be considered. Hence, sample with a single labeling position for a donor is no appropriate reference.

Unfortunately, labeling strategies allowing for ideal FRET-samples are not generally applicable to all proteins or are currently very laborious. For instance, unnatural amino acids (UAAs) require extensive preceding work involving genetic engineering of the *E. coli* expression strain and organic synthesis of the tRNA (Kim et al., 2013, Wan et al., 2012) and often yield incomplete peptides, in particular in case of large proteins. As alternative to UAAs cysteines are introduced via site-directed mutagenesis and the dyes are coupled to the protein in a sequential labeling protocol either under native or unfolding conditions using maleimide thiol coupling chemistry. This elaborate and highly standardized procedure was applied in many studies. Here we demonstrate how to obtain representative reference samples for bis-functional proteins by sequential labeling strategies, limited proteolysis of the FRET-sample and the references by making use of the site-specific fluorescence properties reflected in the fluorescence lifetime (fluorescence decay analysis), the quenching (fluorescence correlation spectroscopy, FCS) and the anisotropy.

## 2. Materials and Methods

### 2.1. Distances by FRET

The steady-state FRET-efficiency,  $E$ , of a sample is determined by the fluorescence quantum yields  $\Phi_F^{D0}$  and  $\Phi_F^{A0}$  of the donor and the acceptor fluorophore, respectively, and  $F_{DD}^{DA}/F_{AD}^{DA}$ , the intensity-ratio of the donor fluorescence and FRET-sensitized acceptor emission:

$$E = \left[ 1 + \frac{\Phi_F^{A0} F_{DD}^{DA}}{\Phi_F^{D0} F_{AD}^{DA}} \right]^{-1} = \left[ 1 + \left( \frac{\langle R_{DA} \rangle_E}{R_0} \right)^6 \right]^{-1} \quad (1)$$

Here,  $R_0$  is Förster-radius and  $\langle R_{DA} \rangle_E$  is the fluorescence averaged DA-distance. Throughout this paper the subscripts indicate the detection and excitation channels: The subscript  $D|D$  represents donor detection  $D|D$  given donor excitation  $D|D$  while  $A|D$  corresponds to acceptor detection  $A|D$  given donor excitation  $A|D$ . The super-script specifies the sample type:  $DA$  represents a FRET-sample while  $D0$  and  $A0$  represents samples containing a donor and acceptor fluorophore at the first labeling position, respectively.

To determine the FRET-efficiency both, the donor fluorescence quantum yield and the acceptor fluorescence quantum yield, have to be known. However, to determine the DA-distance  $\langle R_{DA} \rangle_E$  only the acceptor fluorescence quantum yield  $\Phi_F^{A0}$  is necessary, as the donor fluorescence quantum yield is also included in the definition of  $R_0$ . Therefore, the fluorescence averaged distance  $\langle R_{DA} \rangle_E$  is given by:

$$\langle R_{DA} \rangle_E = R_{0r} \Phi_F^{A0} \left( \gamma \frac{F_{D/D}^{DA}}{F_{A/D}^{DA}} \right)^{1/6} \quad (2)$$

Here  $R_{0r}$  is the reduced Förster-radius ( $R_{0r} = R_0 / \sqrt[6]{\Phi_F^{D0}}$ ) is determined once for a given dye-pair (Rothwell et al., 2003).

If the donor and the acceptor dye are separated by a DA-distance  $R_{DA}$  the fluorescence decay of the donor in absence of FRET is mono exponential and the fluorescence decay in presence of FRET is given by eq. (3).

$$\begin{aligned} f_{D/D}^{DA}(t) &= k_{F,D} \cdot N_0 \cdot e^{-(k_D + k_{RET})t} \\ &= k_{F,D} \cdot N_0 \cdot e^{-k_D \left( \left( \frac{R_0}{R_{DA}} \right)^6 + 1 \right) t} \end{aligned} \quad (3)$$

Here,  $f_{D/D}^{DA}(t)$  is the fluorescence intensity decay,  $1/k_D = \tau_D$  is the lifetime of the donor in absence of FRET,  $k_{F,D}$  is the radiative rate of fluorescence and  $N_0$  is the initial amplitude at time  $t=0$ .

## 2.2. Systematic errors

In Figure 1A/B we show the effect of FRET-inactive species and a mixture of species in the apparent DA-distance  $R_{DA,rec}$ . For a mixture of  $N$  species with fractions  $x_i$  and transfer-efficiencies  $E_i$  the FRET-efficiency  $E$  is given by:

$$E = \sum_{i=1}^N x_i E_i = \sum_{i=1}^N x_i \left( 1 + \left( \frac{R_{DA,i}}{R_0} \right)^6 \right)^{-1} \quad (4)$$

Here,  $R_{DA,i}$  is the distance that corresponds to the transfer-efficiency  $E_i$  and  $R_0$  is the Förster-radius of the dye-pair. Using the FRET-efficiency  $E$  a interpreted apparent DA-distance  $R_{DA,rec}$  is calculated:

$$\left( \frac{R_{DA,rec}}{R_0} \right) = \left( \frac{1}{E} - 1 \right)^{1/6} \quad (5)$$

Overall  $N$  relative deviations have to be considered:

$$\delta_i = \left( \frac{R_{DA,i} - R_{DA,rec}}{R_{DA,i}} \right) \quad (6)$$

In Figure 1A/B the relative deviations are plotted in dependence either of  $x_{Donly}$ , the fraction of FRET-inactive species, (Figure 1A) or in dependence of a second distance (Figure 1B). By direct excitation of the error introduced by species acceptor lacking an acceptor (Figure 1A) can be minimized (Kudryavtsev et al., 2012, Clegg et al., 1992).

In Figure 1C we show the effect of inappropriate acceptor reference quantum yield  $\Phi_F^{A0}$  and inaccurately calibrated setups calibration on the relative error of the FRET-averaged DA-distance  $\langle R_{DA} \rangle_E$ . This error is obtained by simple error propagation using equation (2). The main error sources for  $\langle R_{DA} \rangle_E$  are systematic errors caused by the the acceptor quantum yield  $\Phi_F^{A0}$  and  $\gamma = g_G/g_R$  the detection efficiency ratio of the donor and acceptor channel:

$$\left( \frac{\Delta \langle R_{DA} \rangle_E}{\langle R_{DA} \rangle_E} \right)^2 = \frac{1}{36} \left( \left( \frac{\Delta \gamma'}{\gamma'} \right)^2 + \left( \frac{\Delta \Phi_F^{A0}}{\Phi_F^{A0}} \right)^2 \right) \quad (7)$$

The error of the FRET-averaged donor-acceptor distance  $\langle R_{DA} \rangle_E$  scales linear with the distance and depends only weakly on the acceptor quantum yield  $\Phi_F^{A0}$  and the detection-efficiency ratio.

Using equation (3) the effect of an inaccurately determined de-excitation rate of the donor in absence of FRET can be estimated. If  $k_D$  deviates by  $\pm \Delta k_D$  from the correct  $k_D$  the recovered distance is given by:

$$R_{DA,\pm} = R_0 \left( \left( \frac{R_0}{R_{DA}} \right)^6 \pm \frac{\Delta k_D}{k_D} \right)^{-1/6} \quad (8)$$

In other words if the reference rate constant  $k_D$  is determined with an accuracy of  $\Delta k_D$  the donor-acceptor distance is expected to be the range of  $R_{DA,-}$  to  $R_{DA,+}$ .

### 2.3. Accessible-volume simulations

The dyes were couple to the protein by long linkers. Hence, the spatial distribution of the fluorophores has to be considered. The spatial dye distributions were modeled by the accessible-volume (AV) approach according to (Cai et al., 2007, Muschielok et al., 2008, Sindbert et al., 2011). The AV-approach uses a geometric search algorithm to determine all dye positions within the linker-length from the attachment point which do not cause steric clashes with the macromolecule. The dyes were approximated by ellipsoids. The center of each ellipsoid was connected to its attachment point by a flexible linkage of a length  $L_{link}$ . In all cases the  $C_{\beta}$ -atoms

of the amino acids were used as attachment points. The linker-length is given by the longest distance from the attachment point ( $C_{\beta}$ -atom of the cysteine) to the center of the dye. It includes the reactive group, a spacer and the internal linker of the dye. Both Alexa Fluor 488 C5 maleimide (Alexa488) and Alexa Fluor 647 C2 maleimide (Alexa647) were modeled with a linker width of 4.5 Å. As linker-length  $L_{\text{link}}$  of 20.5 Å and 22 Å were used, for Alexa488 and Alexa647, respectively. The radii of the ellipsoid ( $R_{\text{dye1}}$ ,  $R_{\text{dye2}}$  and  $R_{\text{dye3}}$ ) to model the dye were determined by the dye dimensions. Alexa488 was modeled using radii of 5.0 Å, 4.5 Å and 1.5 Å. Alexa647 was modeled using radii of 11.0 Å, 4.7 Å and 1.5 Å. To study the effect of the linker-length on the symmetry, additionally the fluorophore pair BodipyFL C1 iodacetamine (Bodipy) and Alexa647 was simulated. Here a linker-length of 10.8 Å and linker-width of 4.5 Å was used while the dye was simulated by radii of 4.5 Å, 3.2 Å and 0.9 Å.

To determine the effect of the labeling symmetry on the DA-distance distribution (Figure 2) a set of 5592 protein structures with at least 360 amino acids in the chain, a minimum resolution of 1.8 Å and no unidentified amino acids was selected from the protein databank using the software PDBselect (Griep and Hobohm, 2010). For each crystal structure at least 180 random labeling pairs were chosen at random. Next, for each labeling pair the acceptor and donor AVs of the pair DA (the donor is located at the first amino acid), and the pair AD (the donor is located at the second amino acid) were calculated. For the DA and AD-pair the average and the width of the DA-distance distribution was calculated. If one of the two amino acids was buried and inaccessible by the dye the labeling pair was discarded.

#### **2.4. Recombinant protein expression and purification**

The cysteine mutations were introduced into a human guanylate binding protein 1 (hGBP1) construct in a pQE80L vector (Qiagen, Germany), after removal of all nine native cysteines (hGBP1 Cys9), following the instructions of the QuikChange site-directed mutagenesis kit (Stratagene, USA). The success of the mutagenesis was verified by DNA sequencing with a 3130xl sequencer (Applied Biosystems, USA). Into the cysteine free hGBP1 construct either one or two new cysteines were introduced at various positions of interest (N18C, Q254C, Q344C, T481C, A496C, Q525C, 540C, Q577C). The hGBP1 variants used in this study were expressed in *Escherichia coli* BL21(DE3) and purified following the protocol described previously (Praefcke et al., 1999). For the affinity chromatography Cobalt-NTA-Superflow was used instead of Nickel-NTA-Superflow. No glycerol was added to any buffer as it did not make any detectable differences. The storage buffer did not contain any DTT or DTE as it would interfere with the following labeling reactions. The protein concentration was determined by absorption

at 280nm according to the protocol of Gill&Hippel using an extinction coefficient of  $45400 \text{ M}^{-1} \text{ cm}^{-1}$  (Gill and Vonhippel, 1989). The site-directed mutagenesis and purification of the postsynaptic density protein PSD-95 was performed as previously published (McCann et al., 2012).

## **2.5. Labeling**

### *hGBP1*

The hGBP1 labeling was performed in two subsequent labeling steps, if not otherwise stated following the acceptor-first approach (see Fig. 3). First, the protein was incubated with a deficit amount of Alexa647 maleimide C2 (Alexa647) (Invitrogen, Germany). To start the first labeling reaction, protein solution of 100-300  $\mu\text{M}$  concentration in labeling buffer containing 50mM Tris-HCl (pH 7.4), 5mM  $\text{MgCl}_2$ , 250mM NaCl was gently mixed with 1.5-fold molar excess of Alexa647. After incubation for 1 hour on ice, the unbound dye was removed using a HiPrep 26/20 S25 Desalting column (GE Healthcare, Germany) with a flow rate of 0.5ml/min. This first labeling step generates four different protein species: doubly labeled protein (AA), the single labeled proteins (A0 and 0A) and unlabeled protein (00). Based on the charge difference introduced by the coupled dyes, these species were separated by anion exchange chromatography (ResourceQ column, GE Healthcare, Germany) using a salt gradient (0-500mM NaCl over 120ml at a pH of 7.4, flow rate of 2.0 ml/min). The peaks in the elugram were analyzed for their degree of labeling (DOL) by UV/Vis spectroscopy measuring their absorption at a wavelength of 280 nm and the 651 nm. In the next step the single-labeled fraction (A0 and 0A) with the highest amount of protein was labeled by a 4-fold molar excess of Alexa488 C5 maleimide (Alexa488). After labeling the uncoupled dye separated as described for the first labeling step and the final DOL for both dyes was determined. The DOL was usually in a range of 70-100% for each dye. The DOL were determined by absorption using  $71.000 \text{ M}^{-1} \text{ cm}^{-1}$  as extinction coefficient for Alexa488 and  $265.000 \text{ M}^{-1} \text{ cm}^{-1}$  for Alexa647, respectively. Finally, the labeled proteins were aliquoted in storage a buffer containing 50mM Tris-HCl (pH 7.9), 5mM  $\text{MgCl}_2$ , 2mM DTT, shock-frozen in liquid nitrogen and stored at  $-80^\circ\text{C}$ . A typical elugrams of the donor-first and acceptor first labeling are shown in Figure 3.

### *PSD-95*

The labeling of the postsynaptic density protein PSD-95 was performed by the donor-first approach (see Figure 3) in a buffer containing 50mM Tris-HCl (pH 7.4), 300mM NaCl and 0.5 mM EDTA. The PSD-95 concentration in the labeling buffer was adjusted to 6  $\mu\text{M}$ . Afterwards a deficit amount of Alexa488 C5 maleimide was added to the labeling buffer (0.4 dye molecules

per cysteine) and incubated for 3 hours at 4°C. Following the labeling by the donor dye and purification using a NAP-5 pre-packed columns (GE Healthcare) a fraction of the sample was saved as donor reference. After collection of the donor sample the acceptor dye (Alexa 647 C2 maleimide) was added in excess (2.0 dye molecules per cysteine) to the sample. Finally, the sample was incubated for 3 hours at 4°C and the labeled FRET samples were purified with NAP-5 pre-packed columns (GE Healthcare).

## **2.6. Enzymatic proteolysis**

The distribution of D and A among the two possible labeling sites of hGBP1 was determined by limited enzymatic proteolysis. The resulting band patterns of the protein fragments of the double labeled FRET constructs were compared to the band patterns of proteins labeled with single fluorophores at known sites. The proteolysis was performed at protein concentrations of the 1 mg/ml using a trypsin (Sigma, Germany) concentration of 0.005 mg/ml. After defined incubation times samples were taken and analyzed by SDS-PAGE with 12.5% polyacrylamide. After electrolysis the gel was bathed in a fixing solution (50% v/v H<sub>2</sub>O, 40% v/v EtOH, 10% v/v acetic acid) for 15 min. Next, the fluorescent protein bands of the donor and acceptor detection-channel were visualized in a BioSpectrum Imaging System (UVP, UK) excited by a BioLite MultiSpectral Source (UVP, UK). Finally, the gel was stained with Coomassie dye and imaged. For visual comparison the band patterns of the donor, the acceptor and the Coomassie staining were overlaid..

## **2.7. MFD-measurements**

The MFD-measurements on hGBP1 were performed using a polarized, active-mode-locked argon-ion laser (Sabre®, Coherent) operating at a repetition rate of 73.5 MHz and a wavelength of 496.5 nm for excitation. The PSD-95 measurements were performed using a 495 nm diode laser (PicoQuant) operating at a repetition rate of 64 MHz. In both cases the laser light with a power of 120 μW at the objective was focused into a dilute solution by an corrected water immersion objective (Olympus, 60X, 1.2 NA). The single-molecule analysis PDA and seTCSPC (sub-ensemble time-correlated single photon counting) were performed at concentration of ~50 pM to ensure that only distinct molecules are detected. The FCS-experiments were performed at higher concentrations yielding autocorrelation amplitudes of ~1.5 which corresponds to ~2 molecules in the observation volume. In both cases the signal emitted by the sample was collected through the same objective and spatially filtered by a 100 μm pinhole, to define an effective confocal detection volume. The collected light was divided by a polarizing beam splitter into parallel and perpendicular and subsequently into two different colors (“green” and “red”).

After passing through band pass filters (AHF, HQ 520/35 and HQ 720/150) the light was detected by two “green” (either  $\tau$ -SPADs, PicoQuant, Germany (PSD-95), or MPD-SPADs, Micro Photon Devices, Italy (hGBP1)) and two “red” detectors (APD SPCM-AQR-14, Perkin Elmer, Germany). For each photon the time between the laser pulse and detection of the photon, the time since the last photon, the polarization, and the detection window (“green” or “red”) was recorded by a personal computer using two time-correlated single photon counting (TCSPC) boards (SPC 132, Becker & Hickel, Berlin). To generate FCS curves the recorded photon streams of the parallel and perpendicular detector were cross-correlated. The non-fluorescent background (1-2 kHz) was discriminated as previously described (Eggeling et al., 2001) before generating sub-ensemble fluorescence decay histograms and single molecule anisotropy histograms (Eggeling et al., 2001). Additionally, for each detected single-molecule the mean fluorescence lifetime was determined by a maximum likelihood estimator (Maus et al., 2001).

## 2.8. Anisotropy PDA

Using the D0 and 0D-samples of a FRET-sample as a reference the fraction of DA molecules  $x^{DA}$  was determined. Here, exploit differences in anisotropy among the donor samples to determine the composition of the FRET-samples. In the analysis, it has to be considered, that the steady-state anisotropy  $r_s$  depends on the fluorescence lifetime and the mobility of the fluorophore: with decreasing lifetime the steady state anisotropy  $r_s$  increases. This dependence is described by the Perrin-equation related the rotational correlation time of the fluorophore, the fluorescence lifetime,  $\tau$ , and the steady-state anisotropy,  $r_s$ . To describe the time-resolved fluorescence decays (see Table S4) two rotational correlation times were necessary. Thus, the Perrin-equation consists of two terms:

$$r_s(\tau) = r_0 \left( \frac{1 - x_{slow}}{1 + \frac{\tau}{\rho_{fast}}} + \frac{x_{slow}}{1 + \frac{\tau}{\rho_{slow}}} \right) \quad (9)$$

Here  $\tau$  is the fluorescence lifetime given by the fluorescence averaged lifetime,  $x_{slow}$  is the fraction of slowly rotating dye and  $\rho_{fast}$  and  $\rho_{slow}$  are the fast and the slow rotational correlation times, respectively. Following the Perrin-equation of a D0 and the 0D sample the single-molecule MFD-histograms to assign of the FRET-samples by visually (see Figure 5C).

These differences were quantified by analyzing the anisotropy histograms  $P(r_s)$  of the FRET-sample (a mixture of DA and AD) and the two donor references (D0 and 0D) by photon distribution analysis (PDA) (Kalinin et al., 2007). Here the anisotropies of the references (D0 and

0D) were corrected by the Perrin-equation to fluorescence lifetime of FRET-population to yield the species fraction  $x^{DA}$  of DA molecules. This analysis was performed in four steps. First, The single-molecule anisotropy histograms of D0 and 0D reference samples were fitted individually, usually by two steady state anisotropies  $r_{S,1}, r_{S,2}$ :

$$P(r_S) = x_1 \cdot r_{S,1} + x_2 \cdot r_{S,2} \quad (10)$$

These results are summarized in Table S3. Second, A Perrin-equation for each anisotropy state was calculated using the fitted rotational times. Third, for each anisotropy state and the steady-state anisotropy at the fluorescence averaged lifetime as given by the FRET-sample was calculated. Finally, the anisotropy histogram of the FRET-sample was fitted by a linear combination of the reference samples patterns (corrected by the fluorescence lifetime) to determine the species fraction  $x^{DA}$  of DA molecules and the error of  $x^{DA}$  is estimated by the F-distribution. The results of this analysis are summarized in Table S1 and Table S2 in the supplement.

## 2.9. FCS

Often, the autocorrelation function  $G(t_c)$  of the D0 and the 0D sample visually distinguishable (compare Figure 5A). Here, exploit these differences, using pure D0 and 0D samples as a references, to determine the species fraction  $x^{D0}$  of D0 molecules in an unknown mixture of D0 and 0D molecules. The fluctuations of the donor fluorescence intensities studied by FCS and recorded by a confocal setup as described above, are caused by the diffusion of the molecules and quenching of the donor fluorescence. In absence of an acceptor the donor is quenched solely by its local environment.

To determine  $x^{D0}$  by the correlation curves, we characterized the site-specific fluctuation of the fluorescence intensity by FCS-measurements of the two donor references (D0 and 0D) obtained by labeled single cysteine constructs. The normalized autocorrelation functions of the donor-labeled single cysteine constructs were described by:

$$G(t_c) = 1 + \frac{1}{N_{eff}} \cdot G_{diff}(t_c) \cdot G_B(t_c) \quad (11)$$

Herein is  $G_{diff}(t_c) = (1 + t_c / t_{Diff})^{-1} (1 + s^2 t_c / t_{Diff})^{-1/2}$  a normalized diffusion term for a Gaussian shaped detection volume. The shape parameter  $s$  and the diffusion time  $t_{Diff}$  are characteristic for the detection volume and the diffusion coefficient of the observed particle, respectively (Aragon and Pecora, 1976).  $G_B$  is a function describing empirically the quenching, resulting in bunching processes of the autocorrelation function, and  $N_{eff}$  is the effective number of molecules. The

effective number of molecules  $N_{eff}$  is a weighted average determined by the number of molecules  $N_i$  and their respective brightness  $B_i$ :

$$N_{eff} = \frac{\sum_i N_i B_i^2}{(\sum_i N_i B_i)^2} \quad (12)$$

In the analysis, we assume that the molecules are described by the same diffusion coefficient. For PSD-95 the pattern of the bunching-terms  $G_B(t_c)$  was empirically described by a combination of a single- and a stretched exponential function  $G_B(t_c)$  with the amplitudes  $d_1$  and  $d_2$ , the correlation times  $t_1$  and  $t_2$   $G_B(t_c) = (1 - d_1 + d_1 \cdot \exp(-t_c/t_1) - d_2 + d_2 \cdot \exp(-(t_c/t_2)^\gamma))$  and the stretching exponent  $\gamma$ .

Next, we use the patterns to recover the fraction  $x^{D0}$  of by fitting the D0/0D mixture by a linear combination:

$$G_{Mix}(t_c) = 1 + \frac{1}{N_{eff}} \cdot G_{diff}(t_c) \cdot [A_1 \cdot G_{B,1}(t_c) + A_2 \cdot G_{B,2}(t_c)] \quad (13)$$

The amplitudes  $A_1$  and  $A_2$  were determined by a global fit of all three experimental autocorrelation curves: the two autocorrelation curves of the single-cysteine donor labeled constructs (D0 and 0D) and the donor labeled FRET-construct. The species fractions were calculated using the relative brightness of the single labeled cysteine constructs. The brightness were determined by separate TCSPC measurements (Kim et al., 2005). For two species with brightness  $B_1$ ,  $B_2$  and species fractions  $x_1$  and  $x_2$  where  $x_2 = 1 - x_1$  the FCS-amplitudes  $A_1$ ,  $A_2$  relate to the species fraction by:

$$x_1 = \left( \frac{A_2}{A_1} \left( \frac{B_1}{B_2} \right)^2 + 1 \right)^{-1} \quad (14)$$

The statistical errors were estimated by sampling the fitting parameters with by a Monte-Carlo scheme (Goodman and Weare, 2010, Foreman-Mackey et al., 2013) and using the F-distribution with a 68% confidence level for a given the minimum determined  $\chi^2$  (Goodman and Weare, 2010, Foreman-Mackey et al., 2013).

## 2.10. Fluorescence decay analysis

The FRET-sample is composed of a mixture of DA- and AD-molecules. In DA-molecules the donor is located at the first in AD-molecules the donor is located at the second labeling position. To determine the initially unknown composition of the FRET-sample (a DA/AD mixture) the D0 and 0D donor reference-samples of single-cysteine variants serve as a reference. In the FRET-sample the unknown distribution of the fluorophores among the two labeling sites, given by the species fraction  $x^{DA}$ , was determined by fluorescence decay analysis in 4 steps. First,

single-molecule measurements of the FRET-samples and the donor references were performed. Second, the fluorescence photons were separated from the background by a threshold criterion. Third, using the fluorescent photons (in average  $5 \cdot 10^6$  to  $10 \cdot 10^6$  fluorescence photons per measurement) fluorescence decay histograms were constructed. Finally, all three samples, both donor-samples (D0 and 0D) and the FRET-sample (a DA and AD mixture), were globally analyzed to yield the fraction  $x^{DA}$ . For each sample a parallel and a perpendicular fluorescence decay was measured. Thus, overall six donor fluorescence decay curves were globally fitted.

The fluorescence decays of the donor sample D0 in the parallel  $f_{D|D_{\parallel}}^{D0}(t)$  and the perpendicular detection channel  $f_{D|D_{\perp}}^{D0}(t)$  are given by (Bialik et al., 1998):

$$\begin{aligned} f_{D|D_{\parallel}}^{D0}(t) &= \frac{1}{3} \cdot f_{D|D}^{D0}(t) \cdot (1 + 2 \cdot r^{D0}(t)) \\ f_{D|D_{\perp}}^{D0}(t) &= \frac{1}{3} \cdot f_{D|D}^{D0}(t) \cdot (1 - r^{D0}(t)) \end{aligned} \quad (15)$$

Here  $f_{D|D}^{D0}(t)$  is the fluorescence decay and  $r^{D0}(t)$  is the anisotropy decay of the donor located at the first labeling site. In the analysis, we assume that quenching by FRET and the depolarization of the dyes are independent. Therefore, the fluorescence decay of the donor in presence of the acceptor of the DA-sample for  $f_{D|D_{\parallel}}^{DA}(t)$  the parallel and  $f_{D|D_{\perp}}^{DA}(t)$  the perpendicular detection channel are given by:

$$\begin{aligned} f_{D|D_{\parallel}}^{DA}(t) &= \frac{1}{3} \cdot f_{D|D}^{DA}(t) \cdot (1 + 2 \cdot r^{D0}(t)) \\ f_{D|D_{\perp}}^{DA}(t) &= \frac{1}{3} \cdot f_{D|D}^{DA}(t) \cdot (1 - r^{D0}(t)) \end{aligned} \quad (16)$$

The donor fluorescence decay in presence of FRET can be factorized into the FRET-induced donor-decay  $\varepsilon_D(t)$  and the fluorescence decay of the donor in absence of FRET (see chapter A). Hence, the parallel and perpendicular fluorescence decays of the DA sample are given by:

$$\begin{aligned} f_{D|D_{\parallel}}^{DA}(t) &= \varepsilon_{\parallel}^{DA}(t) \cdot f_{D|D_{\parallel}}^{D0}(t) = \varepsilon_D^{DA}(t) \cdot f_{D|D}^{D0}(t) \cdot \frac{1}{3} \cdot (1 + 2 \cdot r^{D0}(t)) \\ f_{D|D_{\perp}}^{DA}(t) &= \varepsilon_{\perp}^{DA}(t) \cdot f_{D|D_{\perp}}^{D0}(t) = \varepsilon_D^{DA}(t) \cdot f_{D|D}^{D0}(t) \cdot \frac{1}{3} \cdot (1 - r^{D0}(t)) \end{aligned} \quad (17)$$

This highlights, that the polarization resolved FRET-induced donor decay  $\varepsilon_{\perp}^{DA}(t)$  and  $\varepsilon_{\parallel}^{DA}(t)$  can be utilized to identify correlations between the FRET-decay  $\varepsilon_D(t)$  and the anisotropy.

If the species fraction of the DA-molecule is given by  $x^{DA}$ , the recorded fluorescence decays

$f_{D|D_{\parallel}}^{\{DA\}}$  and  $f_{D|D_{\perp}}^{\{DA\}}$  of a species mixture is given by:

$$\begin{aligned} f_{D|D\parallel}^{\{DA\}}(t) &= x^{DA} \cdot \varepsilon^{DA}(t) \cdot f_{D|D\parallel}^{D0}(t) + (1 - x^{DA}) \cdot \varepsilon^{AD}(t) \cdot f_{D|D\parallel}^{0D}(t) \\ f_{D|D\perp}^{\{DA\}}(t) &= x^{DA} \cdot \varepsilon^{DA}(t) \cdot f_{D|D\perp}^{D0}(t) + (1 - x^{DA}) \cdot \varepsilon^{AD}(t) \cdot f_{D|D\perp}^{0D}(t) \end{aligned} \quad (18)$$

Next, we approximate the two FRET-induced donor decays  $\varepsilon^{DA}(t)$  and  $\varepsilon^{AD}(t)$  by a joint FRET-induced decay  $\varepsilon_D(t)$ . This approximation was justified by AV-simulations as shown below in Figure 2. Using this simplification, the fluorescence decays are given by:

$$\begin{aligned} f_{D|D\parallel}^{\{DA\}}(t) &= \varepsilon(t) \cdot \left( x^{DA} \cdot f_{D|D\parallel}^{D0}(t) + (1 - x^{DA}) \cdot f_{D|D\parallel}^{0D}(t) \right) \\ f_{D|D\perp}^{\{DA\}}(t) &= \varepsilon(t) \cdot \left( x^{DA} \cdot f_{D|D\perp}^{D0}(t) + (1 - x^{DA}) \cdot f_{D|D\perp}^{0D}(t) \right) \end{aligned} \quad (19)$$

The FRET-induced donor decay, the fluorescence decays and the anisotropy-decays were formally fitted bi-exponential model functions. The statistical errors of  $x^{DA}$  was estimated by support-plane analysis (Straume et al., 1991) and correspond to a confidence level of 68%(Straume et al., 1991). Correction factors as polarization dependent detection efficiencies and anisotropy mixing in the objective (Koshioka et al., 1995) were considered but were omitted above for clarity(Koshioka et al., 1995).

### 3. Results and Discussion

#### 3.1. Position dependencies of mobile dyes

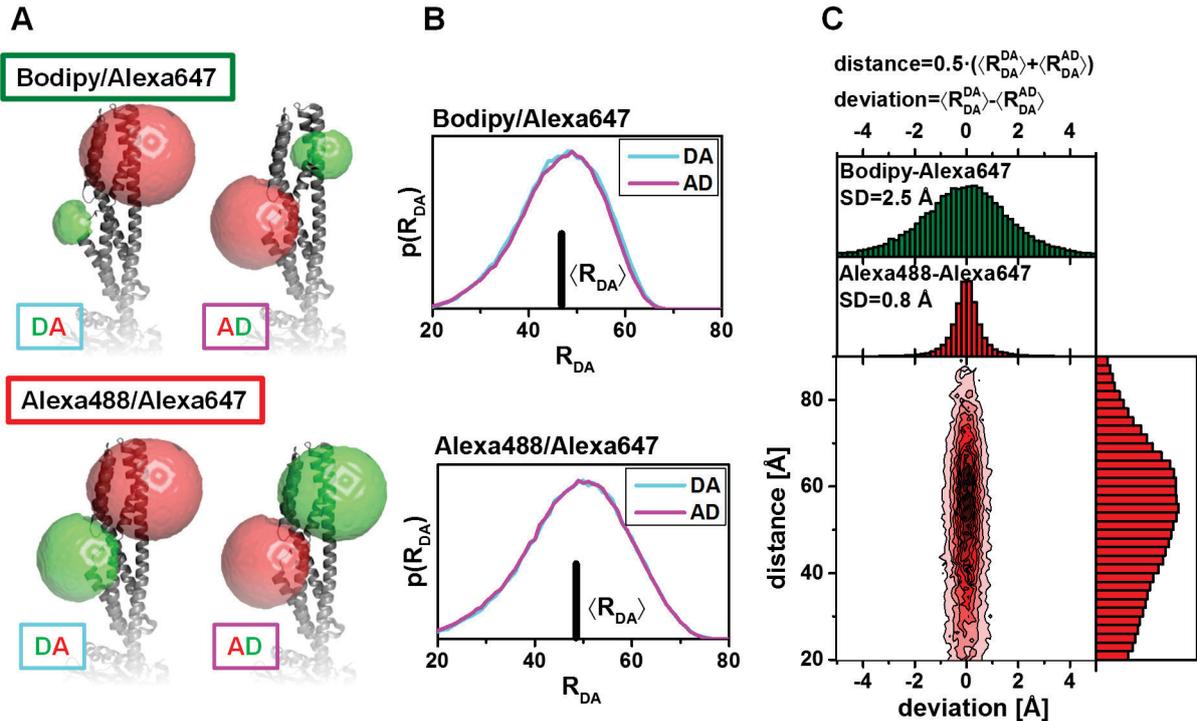
The used dyes were flexible coupled to the proteins. We interpreted the FRET induced donor-decays  $\varepsilon_D(t)$  in terms of distance distributions. As the dyes rotate fast compared to the rate constant of energy transfer and the dye linker diffusion is slow compared to the fluorescence lifetime, the FRET-induced donor decay (FRET-decay) of the DA-molecule and the AD-molecules can be directly related to DA-distance distributions (see chapter A). For a DA-molecules with a DA-distance distribution  $p^{DA}(R_{DA})$  the FRET-induced decay  $\varepsilon^{DA}(t)$  is given by:

$$\varepsilon^{DA}(t) = \int_0^{\infty} p^{DA}(R_{DA}) \cdot \exp\left(-k_{F,D} \cdot \left(\frac{R_{0,r}}{R_{DA}}\right)^6 \cdot t\right) dR_{DA} \quad (20)$$

Here,  $k_{F,D}$  is the radiative rate constant of fluorescence and  $R_{0,r}$  is the reduced Förster radius of the dye-pair. In the analysis of the fluorescence decays (eq. 19) we determine the fraction of DA-molecules,  $x^{DA}$ , assuming that the FRET-induced donor decay  $\varepsilon^{DA}(t)$  of the DA-molecules equals the FRET-induced donor decay  $\varepsilon^{AD}(t)$  of the AD-molecule. Following eq. 20 this assumption is correct if  $p^{DA}(R_{DA}) = p^{AD}(R_{DA})$ . Generally, this is an approximation as D and A have a different dimensions and affinities towards the biomolecule. This may result in distinct distance distributions and thus FRET-induced donor decays of the two species.

To validate if this approximation, we tested for differences of the distance distributions  $p^{DA}(R_{DA})$  and  $p^{AD}(R_{DA})$  by the AV-simulations. All experiments were performed with the dye pair

Alexa488-Alexa647. This dye-pair has a comparable linker-length. Therefore, comparable accessible volumes are obtained (see Figure 2A). To test for the effect linker-length, we additionally simulated the dye pair Bodipy-Alexa647. In this DA-pair the donor fluorophore (Bodipy) has a significantly shorter linker. This results in a smaller AV (see Figure 2A). For the FRET-pair Q344C/A496C of protein hGBP1 the distance distributions of the DA-molecule and the AD-molecule are in perfect agreement for both fluorophore pairs (see Figure 2B).



**Figure 2.** Effect of labeling symmetry on the expected distance distributions evaluated by the accessible volume approach. (A) Accessible volumes of Alexa488/Alexa647- and BODIPY/Alexa647-dye pairs attached to the amino acids Q344C/A496C of a hGBP1 protein structure (PDB-ID: 1DG3) (B) resulting distance distributions and mean distances  $\langle R_{DA} \rangle$ . (C) Comparison of both possible average distances for a set of proteins. The average distances are plotted vs. the deviation between the distances. These calculations are shown for a random set of large protein structures (more than 360 amino acids) for the fluorophore pair Alexa488/Alexa647 (shown in red). The histograms on the side and the top are the projections of the respective axes. For the pair Bodipy/Alexa647 only the deviation histogram is shown (green).

To test if  $\varepsilon^{DA}(t)$  and  $\varepsilon^{AD}(t)$  can be approximation by a single decay we performed multiple AVs calculations of on random proteins structures with minimum chain length of 360 amino acids (see materials and methods). For the fluorophore pair Alexa488/Alexa647 the average mean of the mean distance is plotted vs. the difference of the mean distances. The distribution of the mean distance shows that our simulations cover the typical distance range of FRET is completely. Therefore, we did not perform a reweighting of the samples as described in chapter B. Next we use the differences between both mean DA-distances to quantify the goodness of our approximation. These results are shown in form of histograms in Figure 2C. As it can be seen

the fluorophore pair Alexa488/Alexa647 we expect deviations of approximately 0.8 Å deviation of 2.5 Å are anticipated for the fluorophore pair Bodipy/Alexa647. These simulations corroborate our data analysis to determine the species donor distribution among the two labeling sites as we assumed that the FRET-decay is independent on the position of the donor.

### 3.2. Labeling

#### *FRET-sample*

The aim of the labeling is to obtain a FRET-sample with minimum amount of contaminations and to provide donor and acceptor references. Generally, for a protein with two reactive groups nine distinct molecule species have to be considered: molecules lacking a dye (00), molecules without acceptor (D0, 0D and DD), FRET-molecules (DA and AD) and finally molecules only labeled by acceptors (A0, 0A and AA). An uncontaminated FRET-sample consists of DA and AD. Ideally the fraction of either the DA or the AD the species is zero. However, typically the FRET-sample contains a mixture of DA and AD and is contaminated by 0D-, D0-, DD-molecules and A0-, 0A- and AA-molecules which constitute the so-called donor-only and acceptor-only fraction, respectively. The unlabeled species 00 and the acceptor-only fraction are irrelevant for FRET-measurements, if the spectral crosstalk from the acceptor- to the donor-channel is a negligible. However, as the acceptor fluorescence of the FRET-species is detected together with the acceptor-only species a high acceptor-only fraction prohibits the accurate determination of the acceptor quantum yield by ensemble measurements. This limitation can be overcome by single-molecule techniques as PIE-measurements where FRET-molecules can be separated from the acceptor-only fraction (Kudryavtsev et al., 2012).

#### *One-step random labeling*

In random labeling all nine protein species may be formed. Their actual fraction varies and is hard to predict beforehand. However, if it is assumed that both labeling sites are equally reactive and react completely with all available dyes, the species fractions can be estimated by their counting statistics. The probability of a species to consist in total of  $n_D$  donors and  $n_A$  acceptors and  $n_U$  unlabeled sites can be determined by the multinomial distribution taking two samples:

$$p(n_D, n_A, n_U) = \frac{2}{n_D! n_A! n_U!} \cdot p_D^{n_D} p_A^{n_A} p_U^{n_U} \quad \text{with } n_D + n_A + n_U = 2 \quad (21)$$

Here  $p_D$ ,  $p_A$  and  $p_U$  are the probabilities of finding a donor, acceptor or an unlabeled-binding site on a protein. Per protein two binding sites are available. Hence, if in total less dye is used than binding sites are available not all sites will be occupied. For instance, if 0.1 donors and 0.4

acceptors per site are used the probabilities of a finding a donor ( $p_D$ ), an acceptor ( $p_A$ ) and a free binding site ( $p_U$ ) are 0.04, 0.16 and 0.8, respectively. If more dyes than binding sites are used all binding sites are covered. Hence, the probability of an unlabeled site is zero. Under these conditions the probability of a donor  $p_D$  and acceptor  $p_A$  are given by their relative species fractions. For instance, if 0.5 donor and 2.0 acceptors are used, per site the amount of total dye exceeds the number of binding sites. Therefore, the probability of a free site ( $p_U$ ) is zero while the probabilities of a donor ( $p_D$ ) and acceptor ( $p_A$ ) are 0.2 and 0.8, respectively.

Using the multinomial distribution (eq. 21) one can rationalize concentration effects on the final species distribution and plan a single-step labeling experiment that allows for a donor- and acceptor-reference in a single PIE-experiment. Experimentally, the DD, D0 and 0D species are indistinguishable. Hence, the fraction of DD has to be minimized by a sub-stoichiometric amount of donor per labeling sites. At the same time, the donor-only fraction, given by the ratio of the FRET-molecules (DA, AD) to donor-only molecules (D0, 0D, DD) should be small. This can be accomplished by using a larger acceptor- than donor-concentration. For instance, if 0.1 donors and 0.7 acceptors per labeling-site are used, a donor-fraction of  $\sim 0.25$  is expected. This donor-fraction will be contaminated by 20% DD-molecules. Such labeling strategy comes at the cost of a large acceptor fraction which comprises 70% of the sample.

### *Two-step random labeling*

Ensemble techniques require a separate donor-reference to reliably resolve low-FRET species. An accurate donor reference can be obtained by a two-step approach as depicted in Figure 3. Here, similarly as in the one-step approach the DD contamination of the donor-reference can be minimized by sub-stoichiometric labeling. We estimate based on the donor cysteine ratio that 25% of the donor-reference in the PSD-95 samples are labeled by two donors contaminates, while 75% are labeled by a single donor. Furthermore, we estimate using equation 21 a donor-only fraction of 12% in the FRET-samples. Strictly, equation 21 does not apply for a two-step labeling. Still, this result is in good agreement with the experimentally average donor-only fraction of  $(16 \pm 4)\%$  of the six PSD-95 samples.

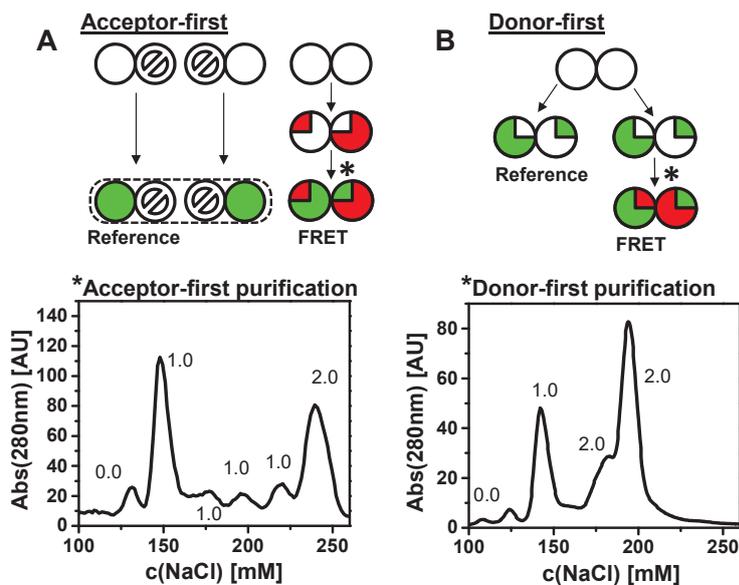
### *Two-step chromatography*

To reduce the acceptor (A0, 0A, AA) and donor (D0, 0D, DD) contaminations we sequentially label the FRET-sample. By additional chromatographic purification steps (see Figure 3) we select and identify the desired FRET-species. In our study we used anion exchange chromatography similar as previously reported (Lerner et al., 2013). To separate the species, we made use

## B - Experimental reference samples for accurate time-resolved FRET analysis

of the negative charge introduced by of Alexa647 and Alexa488 which alters the protein affinity to the stationary phase of the ion exchange matrix.

In the first step, we either labeled the protein by the donor- (donor-first) or by the acceptor-fluorophore (acceptor-first). Both schemes are outlined in Figure 3. The acceptor-first approach is more laborious, as it requires three protein preparations as a donor-reference is not directly obtained. On the other hand, Alexa647 introduces two negative charges as opposed to one by Alexa488 resulting in in a better peak separation (see Figure 3). In total, we tested 11 different cysteine pairs being distributed over the whole hGBP1 protein. As shown for N18C/Q577C the first labeling step can be successful accomplished either with the donor Alexa488 or the acceptor Alexa647 in the first place. By selectively labeling either singly labeled donor or acceptor species we were able to reduce the donor-only fraction to at least 5% and in average to  $(8\pm 5)\%$  being significantly smaller than anticipated by a random labeling.



**Figure 3.** Workflow of the individual steps in sequential labeling with chromatographic purification: (A) acceptor-first vs. (B) donor-first labeling. The flow chart illustrates the preparation steps starting from the unlabeled proteins going to the donor-reference samples and the FRET-sample. Binding sites are shown as circles. Labeled binding sites are colored according to the attached dye (donor - green, acceptor - red). The distribution of the dyes among the two sites is illustrated by partly filled circles. In the acceptor-first approach this distribution is not reflected by the donor reference samples and has to be determined. The chromatographic separation by anion exchange chromatography after the

first labeling step shown by a star (\*). On the bottom a elugram following the first labeling step of the hGBP1-variant N18C/Q577C is shown. The numbers on top of the peaks correspond to the degree of labeling (DOL) of the individual peaks as determined by UV-Vis spectroscopy.

As illustrated in Figure 3 no donor-reference with the correct fraction  $x^{D0}$  is obtained by the acceptor-first approach. Hence, the only way to obtain a donor-reference is to measure the D0- and the 0D-sample separately and determine their relative fraction in the FRET-sample. Therefore, acceptor-first approaches are inferior to donor-first approaches and should only be used, if demanded by of biochemical or preparative reasons. To demonstrate that the more laborious acceptor-first approach is nevertheless feasible, we used a set of different methods capable of

determining the fraction  $x^{D0}$  in proteins hGBP1 and PSD-95. Here, the simplest being an enzymatic digestion of the FRET labeled molecules and the single labeled reference samples followed by an analysis with SDS-PAGE. For quantitative analysis we employed spectroscopic methods as: time-correlated single photon counting (TCSPC), anisotropy photon distribution analysis (PDA) and fluorescence correlation spectroscopy (FCS).

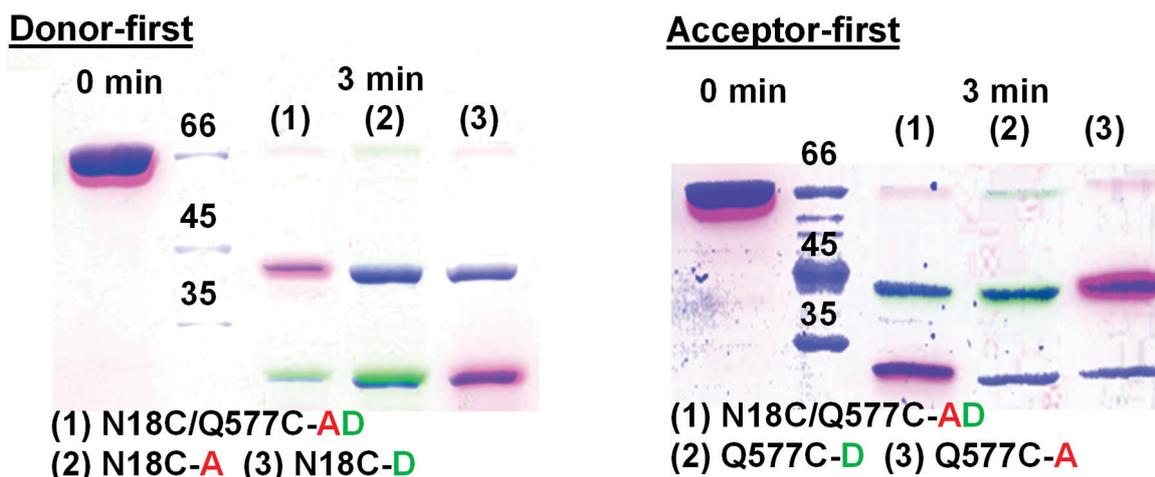
### **3.3. Determination of dye locations**

#### **3.3.1 Enzymatic proteolysis**

The distribution of the donor in double-labeled FRET samples was determined qualitatively by proteolytic digestion of the FRET-sample using the labeled single cysteine variants as a reference. The labeled single-cysteine as well as double-cysteine constructs were subjected to short simultaneous protease digest using commercially available trypsin. Following this digestion, the fragments were analyzed by denaturing SDS-PAGE. For assigning the dyes to either one or the other cysteine, the band pattern of the single-labeled and double-labeled constructs were compared. As previously reported (Rothwell et al., 2013), this approach most yielded unambiguous results. The protein used in this study, hGBP1, is a 67kDa protein consisting of a globular GTPase domain and long expanded helical domain (Prakash et al., 2000).

After short incubation trypsin digested hGBP1 mainly into two fragments of approximately 30 and 40kDa in size. The crystal structure and the primary sequence of hGBP1 suggest that the first cleavage happens at position K252 separating the globular GTPase domain from the helical domain. Therefore, the assignment for the variants with cysteines distributed among these two first fragments such as N18C/Q344C, N18C/V540C and N18C/Q577C was straight forward. By comparing the band patterns of the FRET-sample and the singly-labeled reference constructs either the donor or acceptor could be assigned to one of the fragments. This is shown for N18C/Q577C in Figure 4.

Furthermore, it can be deduced that irrespectively of the used dye (either Alexa488 or Alexa647) the amino acid Q577C is more reactive compared to the amino acid N18C, as in the donor-first approach the labeling is reversed Therefore, the fluorophores can be directed to different sites. Determining the distribution of the donor among two labeling sites by proteolysis has its limitations and is not always applicable. In case of hGBP1, the assignment by digestion and SDS-PAGE analysis was not possible in double cysteine variant where the cysteines were spatially “too close” such as in the variants Q254C/Q344C or T481C/Q525C (data not shown).



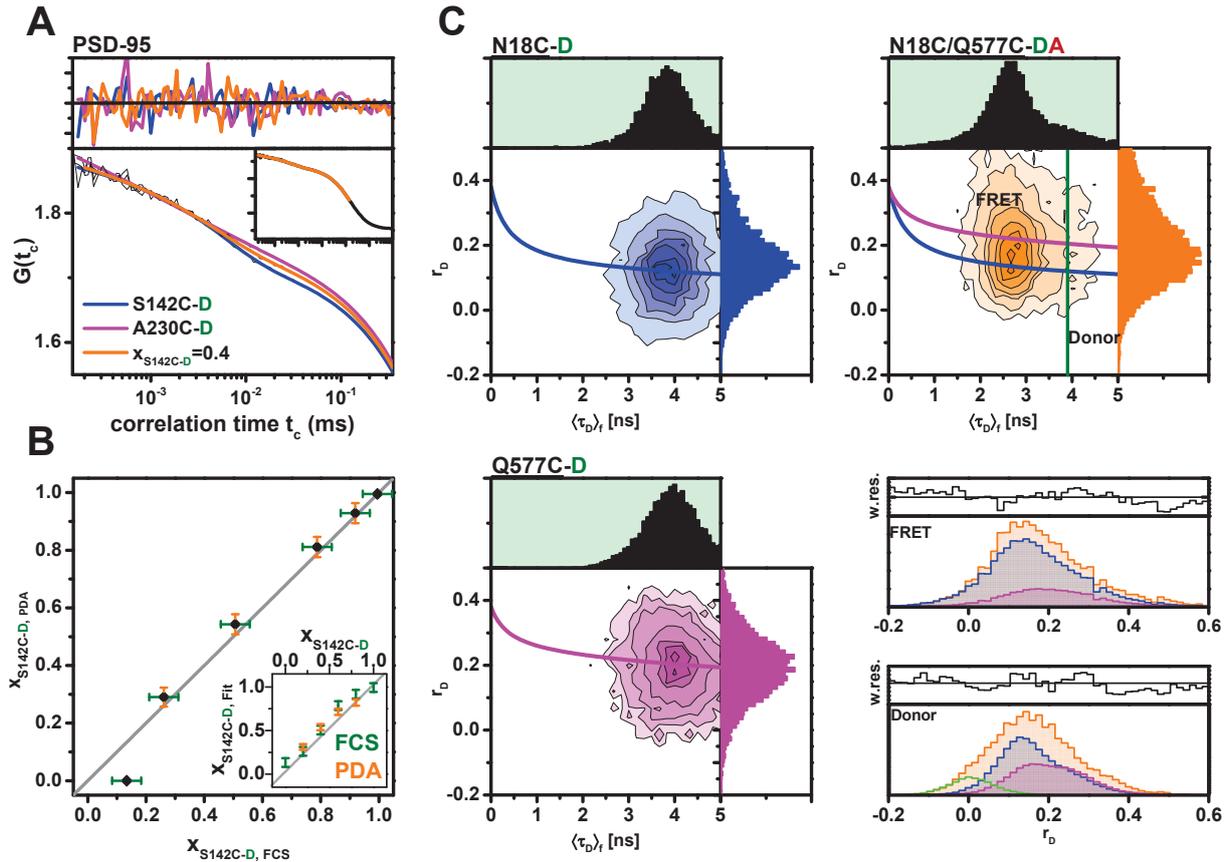
**Figure 4.** SDS-PAGE Gels of sequentially labeled hGBP1 N18C/Q577C, N18C and Q577C labeled by the donor and acceptor fluorophore. N18C/Q577C was either labeled by the donor or acceptor first. For each gel the fluorescence donor (green) and the acceptor signal (magenta) was recorded and overlaid with the coomassie stained protein bands (blue).

### 3.3.2 FCS, Anisotropy

Before analyzing the unknown protein samples, we validated and cross-validates the anisotropy PDA- and the FCS-analysis method using known mixtures of the two donor labeled molecules (PSD-95 S142C and PSD-95 A230C) (see Figure 5C). In FCS, the timescales and magnitude of dark states of the dyes are determined by correlation of the time-dependent fluorescence intensities. The local protein environment alters the quenching processes. Hence, the autocorrelation curves reflect dynamic processes of the fluorophore and internal processes of the biomolecule at the specific position of the fluorophores. By anisotropy PDA the fraction of donor bound to the first and the second binding site is determined. This reports on the site specific dye mobility and thus, via references samples, on the labeling site. Different species fractions of the molecules PSD-95 S142C and PSD-95 A230C were analyzed by PDA and FCS as described in the method section above. For calibration the species fractions were determined by steady-state fluorescence intensities and the fluorescence quantum yields of the samples. The relative fluorescence quantum yields of the fluorescent species were determined by separate TCSPC-measurements. As illustrated in Figure 5B, the statistical errors in this validation are small compared to the systematic deviations between the results obtained by PDA and FCS. These systematic deviations are presumably caused by differences of the molecular brightness of the two species. The PSD-95 species labeled on position S142 are roughly 16% brighter as compared to species labeled on position A230C resulting in a systematic deviation towards the species S142C.

We analyzed the composition of the PSD-95 FRET samples before acceptor labeling by global analysis of FCS-autocorrelation curves of the donor-only labeled samples of the single-cysteine and the double-cysteine constructs of PSD-95. These results agree with the results found by the

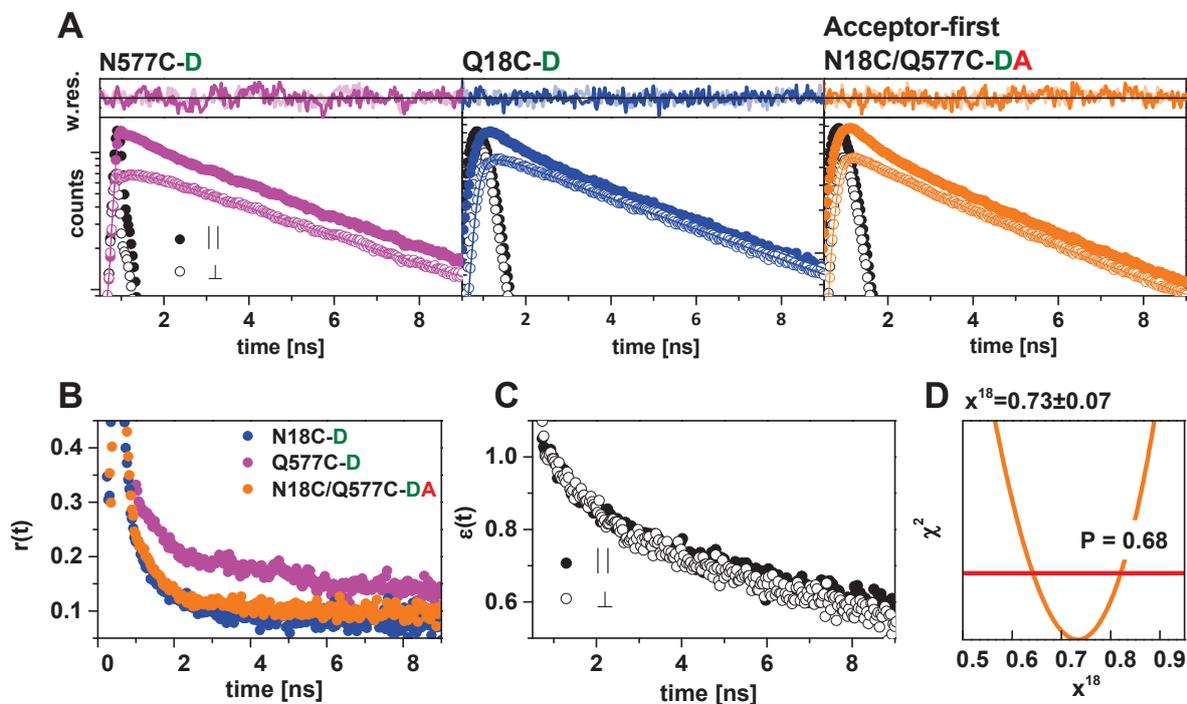
proteolytic digestion. If a major species was identified it always corresponded to the result of the limited proteolysis. For instance, the gel analysis of the acceptor-first labeled hGBP1 FRET-sample N18C/Q577C, suggested that the donor-fluorophore is located on position N18C. Similarly, the analysis by PDA revealed that roughly in 77% the donor is located at position N18C while in 23% it is located at position Q577C (see Figure 5C). This is in agreement with the recorded gel (see Figure 4) which indicate that the donor might be is partially located at position Q577C.



**Figure 5.** FCS and PDA method. (A, B) Two donor labeled PSD-95 single-cysteine variants (S142C and A230C) were mixed at different species fractions (A) normalized FCS-curves a S142C, A230C mixture ( $x_{S142C}=0.4$ ). The full autocorrelation curve of the donor labeled S142C is shown as inset. (B) Fitted fractions of S142C in a S142C, A230C mixture determined by PDA and FCS. The species-fractions vs the recovered species-fractions are shown as an inset. The statistical error of the method is indicated by the error bars at 68% confidence level. (C) Burst-wise histograms and PDA analysis of single-molecule MFD data of Alexa488 labeled N18C and Q577C hGBP1 variant and a acceptor-first labeled hGBP1 N18C/Q577C variant - the steady-state anisotropy  $r_D$  of individual single-molecule bursts are plotted versus the fluorescence averaged lifetime  $\langle \tau_D \rangle_f$  of the donor fluorophore in a histogram. The Perrin equation of the single cysteine variants are shown as blue (N18C,  $\rho_1=0.2$  ns and  $\rho_2=17$  ns,  $b_1=0.27$ ,  $b_2=0.11$ ) and (purple, Q577C  $\rho_1=0.3$  ns  $\rho_2=8.3$  ns,  $b_1=0.24$ ,  $b_2=0.14$ ) lines. The histogram of FRET-sample was described by a linear combination of the histogram shapes of the single-cysteine samples corrected by the Perrin-equation and fitted by PDA. Based on the lifetime the FRET-sample was separated into a FRET- ( $2.5 \text{ ns} < \langle \tau_D \rangle_f < 3.0 \text{ ns}$ ) and donor-only fraction ( $\langle \tau_D \rangle_f > 3.5 \text{ ns}$ ). In the FRET-sample the donor is mainly located at position N18C (FRET molecules  $x_i=0.77$ , donor molecules  $x_i=0.60$ ). The donor-only fraction additionally required a state describing unbound fluorophores (12%, light green) with an anisotropy fixed to the steady-state anisotropy of Alexa488.

### 3.3.3 Fluorescence decay analysis

To determine the location of the donor fluorophore within the FRET samples via the fluorescence decay, six polarization resolved decays (two of the FRET-sample and four of the references) were globally fitted. The fitting results for the hGBP1 double cysteine variant N18C/Q577C labeled by the acceptor-first and the corresponding single-cysteine variant labeled by a donor are shown in Figure 6. A visual inspection of the time-resolved anisotropy decays already reveals, that the FRET-sample consist of a mixture, as none of the individual donor anisotropy decays describes the FRET-samples. Following the anisotropy decays shown in Figure 6B, the donor is mainly located at position N18C. The non-exponential decay of  $\varepsilon_D(t)$  indicates presence of multiple FRET species. As the FRET-decay  $\varepsilon_D(t)$  shows no significant polarization effects, we conclude that our polarization independent FRET-decay model (equation 14) is applicable for the studied samples. As we are mainly interested in the species fraction,  $x^{D0}$ , which quantifies the distribution of the donor among the two labeling sites and structural interpretations are beyond the scope of this paper, we used a formal bi-exponential model with two FRET-rate constants to describe the FRET-induced donor decay. For the hGBP1-sample N18C/Q577C labeled by the acceptor first approach we find that the donor is mainly located at the position N18C ( $x^{18}=0.73\pm 0.07$ ) being in agreement with the limited proteolysis and the PDA-analysis of the respective same sample.

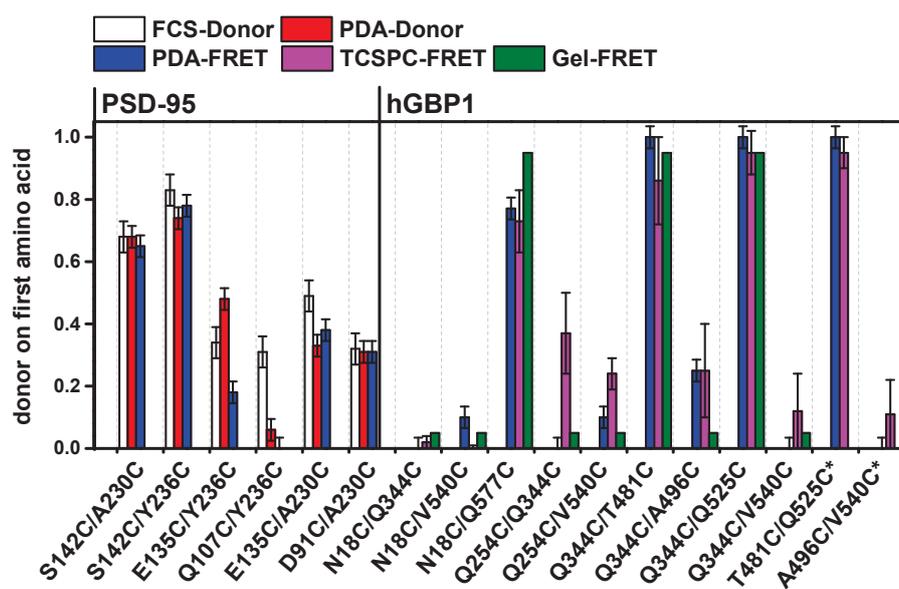


**Figure 6.** Global fluorescence decay analysis of sub-ensemble single-molecule data - (a) polarized donor fluorescence decays of the donor labeled hGBP1- N18C, Q577C and the acceptor-first labeled FRET hGBP1-variant N18C/Q577C (b) anisotropy decays of the FRET sample and the donor-samples (c) polarization dependent FRET-decays  $\varepsilon(t)$  (d) confidence of global fit of the species fraction of donors located at N18C ( $0.73\pm 0.07$ ) in the FRET-sample for a confidence level of 68%.

### **3.3.4 Method comparison**

All methods to determine the distribution of the donor among the two labeling sites provided comparable results (see Figure 7). Among all applied techniques limited proteolysis of the labeled protein with subsequent gel-imaging is the simplest approach, as it can be applied without access to advanced spectroscopic tools. However, it requires the largest sample amounts and only provides qualitative results. Furthermore, it is limited to protein constructs with proteolytic cleavage sites between the two labeling positions. Therefore, in the hGBP1 samples T481C/Q525C and A496C/V540C, the assignment was problematic. To quantify the distribution of the donor among the two labeling sites we made use of the site-specific fluorescence properties of the dyes. The properties FRET-samples were compared to donor reference samples covering in the nanosecond (TCSPC), microsecond (FCS) and millisecond regime (PDA). By this approach the location of the dyes could be resolved.

The presented FCS analysis is restricted to protein samples prepared by the donor-first approach and requires differences in the shape of the bunching terms. However, no sophisticated single-molecule setup is needed. However, this method requires samples free of unreacted dye molecules, as they interfere with the bunching pattern. Contrary to the simple FCS-approach, the single-molecule histogram analysis by anisotropy PDA can be applied to FRET-samples which may be contaminated by unbound dye. In contrast to the PDA analysis which requires single-molecule detection, the fluorescence decay analysis may be applied to single-molecule and ensemble measurements. Even though, the number of free fitting parameters in the FCS-approach is higher compared to the anisotropy PDA, ultimately only a single the parameter (the species fraction) is determined. Both methods recover the species fraction, given the experimental errors of (compare Figure 5C and Figure 7), both seem to be biased towards the brighter species, even though the amplitudes were corrected for the brightness of the respective species (Kim et al., 2005). By inspecting the fluorescence decays shapes, namely FRET-induced donor decay  $\varepsilon_D(t)$  and the anisotropy decay  $r(t)$ , qualitative results may be obtained by visual inspection while the precisely know counting statistics delivers quantitative results.



**Figure 7.** Summary of the analysis results - The fraction of donor located at the first labeling site in the protein sequence is shown for all measured and analyzed PSD-95 and hGBP1-samples. The error bars correspond to the statistical errors at a confidence level of 68%. In case of the gel-analysis for better visual comparison fractions of 0.95 and 0.05 were used in case the donor is located or not located at the first labeling site. (\*) no conclusive result by limited proteolysis

#### 4. Conclusion

Summing up, we recommend if possible first to label by the donor fluorophore. This way as the donor-reference which is compulsory for the analysis of fluorescence decays is easily obtained without need of additional protein samples. If the FRET-sample has to be labeled first by the acceptor two additional donor-references are needed and distribution of the two possible donor species among the labeling sites has to be determined as presented above. In some cases, we were able to steer the dyes to specific labeling sites by sequential labeling and purification steps. As previously reported such an approach allows for highly homogenous samples with respect to the distribution of the fluorophores among the labeling sites (Rothwell et al., 2013). For dyes where the accessible-volume approach holds the distribution of the dyes among the labeling sites is only of minor importance. In particular, if the FRET-pairs such as Alexa488/Alexa647 with a comparable linker-length are used. Thus, the precise composition of the FRET-sample does not have to be determined.

## 5. Supplement

**Table S1.** r-PDA and TCSPC (lifetime and anisotropy decay) labeling distribution

hGBP1-Sample	Labeling-order	Digest	PDA				TCSPC		
		major species	$\langle\tau_{D(A)}\rangle_f$	$r_s$	$\chi_r^2$	major species	fraction	major species	fraction
N18C/Q344C	A,D	Q344C	12	0.16	1.35	Q344C	1.0	Q344C	0.98±0.02 <sup>e</sup>
N18C/V540C	A,D	V540C	2.75	0.17	3.53	V540C	0.90	V540C	1.00±0.01 <sup>e</sup>
N18C/Q577C	A, D	N18C	2.75	0.17	2.35	N18C	0.77	N18C	0.73±0.10 <sup>e</sup>
Q254C/Q344C	A,D	Q344C	3.67	0.12	5.85	Q344C	1.0	Q344C	0.63±0.13 <sup>e</sup>
Q254C/V540C	A,D	V540C	2.81	0.25	4.44	V540C	0.9	V540C	0.76±0.05 <sup>e</sup>
Q344C/T481C	A,D	Q344C	2.51	0.16	1.11	Q344C	1.0	Q344C	0.86±0.14 <sup>e</sup>
Q344C/A496C	A,D	A496C	1.98	0.15	1.00	A496C	0.75	A496C	0.75±0.25 <sup>a,e</sup>
Q344C/Q525C	A,D	Q344C	1.80	0.18	0.76	Q344C	1.0	Q344C	0.95±0.07 <sup>e</sup>
Q344C/V540C	A,D	V540C	2.54	0.25	0.89	V540C	1.0	V540C	0.88±0.12 <sup>e</sup>
T481C/Q525C	A,D	n.c.d	3.55	0.10	3.87	T481C	1.0	T481C	0.95±0.05 <sup>a,e</sup>
A496C/V540C	A,D	n.c.d	3.36	0.17	1.15	V540C	1.0	V540C	0.89±0.11 <sup>e</sup>

<sup>a</sup> Parallel channels, <sup>b</sup>see paper, <sup>c</sup>lifetime and anisotropy properties too similar for seTCSPC, <sup>d</sup> unclear <sup>e</sup>global subensemble (vv, vh) TCSPC of FRET-population with FRET-rate fit using known anisotropies of both possible donor-position, error at 68% confidence level.<sup>g</sup>FCS

**Table S2.** PSD-95 and FCS PDA-analysis

Sample	Labeling-order	$\langle\tau_{D(A)}\rangle_f$	$r_s$	PDA-FRET		FCS-Donor
				$\chi_r^2$	major species (fraction)	major species (fraction)
S142C/A230C	D, A	2.81	0.19	1.16	S142C (0.65)	
S142C/A230C	D	3.74	0.16	1.18	S142C (0.68)	S142C (0.68±0.05) <sup>g</sup>
S142C/Y236C	D, A	2.94	0.17	1.06	Y236C (0.78)	
S142C/Y236C	D	3.67	0.15	0.93	S142C (0.74)	Y236C (0.83±0.05) <sup>g</sup>
E135C/Y236C	D, A	3.00	0.17	0.88	Y236C (0.82)	
E135C/Y236C	D	3.66	0.16	1.04	Y236C (0.52)	Y236C (0.66±0.05) <sup>g</sup>
Q107C/Y236C	D, A	2.524	0.17	1.01	Y236C (1.0)	
Q107C/Y236C	D	3.76	0.17	0.83	Y236C (0.94)	Y236C (0.69±0.05) <sup>g</sup>
E135C/A230C	D, A	2.94	0.20	1.03	A230C (0.62)	
E135C/A230C	D	3.67	0.19	1.12	A230C (0.67)	A230C (0.51±0.05) <sup>g</sup>
D91C/A230C	D, A	2.07	0.24	1.75	D91C (0.69)	
D91C/A230C	D	3.51	0.22	0.93	A230C (0.69)	91 (0.68±0.05) <sup>g</sup>

**Table S3.** Steady-state anisotropies of the donor references as determined by PDA

Variant Vi	Anisotropy						
	$x_1$	$r_1$	$x_2$	$r_2$	$\chi_r^2$	$r_s$	
hGBP1	N18C	0.739	0.128	0.274	0.278	1.26	0.17
	Q254C	0.705	0.255	0.295	0.330	0.98	0.28
	Q344C	0.858	0.116	0.142	0.257	1.23	0.14
	T481C	0.817	0.084	0.183	0.206	1.52	0.11
	A496C	0.696	0.110	0.304	0.250	1.45	0.15
	Q525C	0.806	0.154	0.194	0.303	0.81	0.18
	V540C	0.388	0.163	0.612	0.273	1.16	0.23
	Q577C	0.511	0.144	0.489	0.252	0.76	0.20
	D91C	0.216	0.121	0.784	0.249	0.90	0.22
	Q107C	0.839	0.146	0.161	0.303	1.12	0.17
PSD-95	E135C	0.776	0.148	0.224	0.228	0.62	0.17
	S142C	0.839	0.132	0.161	0.216	1.00	0.15
	A230C	0.182	0.117	0.818	0.217	1.10	0.20
	Y236C	0.797	0.146	0.203	0.232	1.05	0.16

**Table S4.** Donor lifetimes and rotational correlation times

Variant	$x_1$	$\tau_1$ [ns]	$x_2$	$\tau_2$ [ns]	$b_1$	$\rho_1$ [ns]	$b_2$	$\rho_2$ [ns]	
hGBP-1	N18C	0.82	4.15	0.18	1.35	0.27	0.19	0.11	17.41
	Q254C	0.69	3.60	0.31	0.53	0.26	0.16	0.12	16.85
	Q344C	0.94	3.78	0.06	1.00	0.27	0.28	0.11	17.77
	T481C	0.93	3.78	0.07	0.37	0.29	0.17	0.09	1.35
	A496C	0.84	3.68	0.16	0.32	0.13	0.26	0.25	11.75
	Q525C	0.80	3.51	0.20	0.66	0.29	0.26	0.09	2.12
	V540C	0.85	3.87	0.15	1.50	0.19	0.20	0.19	8.70
	Q577C	0.91	4.15	0.09	1.49	0.24	0.30	0.14	8.34
	D91C	0.66	3.66	0.34	0.41	0.13	0.53	0.25	21.9
PSD-95	Q107C	0.66	3.76	0.34	0.27	0.22	0.39	0.16	18.7
	E135C	0.77	3.89	0.23	0.27	0.20	0.28	0.18	15.8
	S142C	0.73	3.96	0.27	0.22	0.22	0.44	0.16	14.6
	A230C	0.74	3.73	0.26	0.46	0.17	0.46	0.21	29.4
	Y236C	0.71	3.74	0.29	0.35	0.20	0.47	0.18	16.2

## References

- ANDRECKA, J., TREUTLEIN, B., ARCUSA, M. A. I., MUSCHIELOK, A., LEWIS, R., CHEUNG, A. C. M., CRAMER, P. & MICHAELIS, J. 2009. Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex. *Nucleic Acids Research*, 37, 5803-5809.
- ANTONIK, M., FELEKYAN, S., GAIDUK, A. & SEIDEL, C. A. M. 2006. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *Journal of Physical Chemistry B*, 110, 6970-8.
- ARAGON, S. R. & PECORA, R. 1976. Fluorescence Correlation Spectroscopy as a Probe of Molecular-Dynamics. *Journal of Chemical Physics*, 64, 1791-1803.
- BIALIK, C. N., WOLF, B., RACHOFISKY, E. L., ROSS, J. B. A. & LAWS, W. R. 1998. Dynamics of biomolecules: Assignment of local motions by fluorescence anisotropy decay. *Biophysical Journal*, 75, 2564-2573.
- BRUSTAD, E. M., LEMKE, E. A., SCHULTZ, P. G. & DENIZ, A. A. 2008. A General and Efficient Method for the Site-Specific Dual-Labeling of Proteins for Single Molecule Fluorescence Resonance Energy Transfer. *Journal of the American Chemical Society*, 130, 17664-+.
- CAI, Q., KUSNETZOW, A. K., HIDEK, K., PRICE, E. A., HAWORTH, I. S. & QIN, P. Z. 2007. Nanometer distance measurements in RNA using site-directed spin labeling. *Biophys J*, 93, 2110-7.
- CLEGG, R. M., MURCHIE, A. I., ZECHEL, A., CARLBERG, C., DIEKMANN, S. & LILLEY, D. M. J. 1992. Fluorescence resonance energy transfer analysis of the structure of the four-way DNA junction. *Biochemistry*, 31, 4846-56.
- EGGELING, C., BERGER, S., BRAND, L., FRIES, J. R., SCHAFFER, J., VOLKMER, A. & SEIDEL, C. A. M. 2001. Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *Journal of Biotechnology*, 86, 163-180.
- FOREMAN-MACKEY, D., HOGG, D. W., LANG, D. & GOODMAN, J. 2013. emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*, 125, 306-312.
- FÖRSTER, T. 1948. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik*, 437, 55-75.
- GILL, S. C. & VONHIPPEL, P. H. 1989. Calculation of Protein Extinction Coefficients from Amino-Acid Sequence Data. *Analytical Biochemistry*, 182, 319-326.
- GOODMAN, J. & WEARE, J. 2010. Ensemble Samplers with Affine Invariance. *Communications in Applied Mathematics and Computational Science*, 5, 65-80.

- GOPICH, I. V. & SZABO, A. 2003. Single-macromolecule fluorescence resonance energy transfer and free-energy profiles. *Journal of Physical Chemistry B*, 107, 5058-5063.
- GRIEP, S. & HOBBOHM, U. 2010. PDBselect 1992-2009 and PDBfilter-select. *Nucleic Acids Research*, 38, D318-D319.
- KALININ, S., FELEKYAN, S., ANTONIK, M. & SEIDEL, C. A. M. 2007. Probability distribution analysis of single-molecule fluorescence anisotropy and resonance energy transfer. *Journal of Physical Chemistry B*, 111, 10253-10262.
- KALININ, S., PEULEN, T., SINDBERT, S., ROTHWELL, P. J., BERGER, S., RESTLE, T., GOODY, R. S., GOHLKE, H. & SEIDEL, C. A. M. 2012. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat Methods*, 9, 1218-25.
- KALININ, S., VALERI, A., ANTONIK, M., FELEKYAN, S. & SEIDEL, C. A. M. 2010. Detection of structural dynamics by FRET: a photon distribution and fluorescence lifetime analysis of systems with multiple states. *J Phys Chem B*, 114, 7983-95.
- KIM, J., SEO, M. H., LEE, S., CHO, K., YANG, A., WOO, K., KIM, H. S. & PARK, H. S. 2013. Simple and Efficient Strategy for Site-Specific Dual Labeling of Proteins for Single-Molecule Fluorescence Resonance Energy Transfer Analysis. *Analytical Chemistry*, 85, 1468-1474.
- KIM, S. A., HEINZE, K. G., BACIA, K., WAXHAM, M. N. & SCHWILLE, P. 2005. Two-Photon Cross-Correlation Analysis of Intracellular Reactions with Variable Stoichiometry. *Biophysical Journal*, 88, 4319-4336.
- KOSHIOKA, M., SASAKI, K. & MASUHARA, H. 1995. Time-Dependent Fluorescence Depolarization Analysis in 3-Dimensional Microspectroscopy. *Applied Spectroscopy*, 49, 224-228.
- KUDRYAVTSEV, V., SIKOR, M., KALININ, S., MOKRANJAC, D., SEIDEL, C. A. & LAMB, D. C. 2012. Combining MFD and PIE for accurate single-pair Forster resonance energy transfer measurements. *ChemPhysChem*, 13, 1060-78.
- LERNER, E., HILZENRAT, G., AMIR, D., TAUBER, E., GARINI, Y. & HAAS, E. 2013. Preparation of homogeneous samples of double-labelled protein suitable for single-molecule FRET measurements. *Analytical and Bioanalytical Chemistry*, 405, 5983-5991.
- MAUS, M., ROUSSEAU, E., COTLET, M., SCHWEITZER, G., HOFKENS, J., VAN DER, A., DE SCHRYVER, F. C. & KRUEGER, A. 2001. New picosecond laser system for easy tunability over the whole ultraviolet/visible/near infrared wavelength range based on flexible harmonic generation and optical parametric oscillation. *Review of Scientific Instruments*, 72, 36-40.
- MCCANN, J. J., ZHENG, L. Q., ROHRBECK, D., FELEKYAN, S., KUHNEMUTH, R., SUTTON, R. B., SEIDEL, C. A. M. & BOWEN, M. E. 2012. Supertertiary structure of the synaptic MAGuK scaffold proteins is conserved. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 15775-15780.
- MÖGLICH, A., JODER, K. & KIEFHABER, T. 2006. End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proceedings of the National Academy of Sciences*, 103, 12394-12399.
- MUSCHIELOK, A., ANDRECKA, J., JAWHARI, A., BRÜCKNER, F., CRAMER, P. & MICHAELIS, J. 2008. A nano-positioning system for macromolecular structural analysis. *Nat Methods*, 5, 965-71.
- PRAEFCKE, G. J. K., GEYER, M., SCHWEMMLE, M., KALBITZER, H. R. & HERRMANN, C. 1999. Nucleotide-binding characteristics of human guanylate-binding protein 1 (hGBP1) and identification of the third GTP-binding motif. *Journal of Molecular Biology*, 292, 321-332.

- PRAKASH, B., RENAULT, L., PRAEFCKE, G. J., HERRMANN, C. & WITTINGHOFER, A. 2000. Triphosphate structure of guanylate-binding protein 1 and implications for nucleotide binding and GTPase mechanism. *EMBO Journal*, 19, 4555-64.
- RATNER, V., KAHANA, E., EICHLER, M. & HAAS, E. 2002. A general strategy for site-specific double labeling of globular proteins for kinetic FRET studies. *Bioconjug Chem*, 13, 1163-70.
- ROTHWELL, P. J., ALLEN, W. J., SISAMAKIS, E., KALININ, S., FELEKYAN, S., WIDENGREN, J., WAKSMAN, G. & SEIDEL, C. A. M. 2013. dNTP-dependent Conformational Transitions in the Fingers Subdomain of KlenTaq1 DNA Polymerase: insights into the role of the "nucleotide-binding" state. *J Biol Chem*, 288, 13575-91.
- ROTHWELL, P. J., BERGER, S., KENSCH, O., FELEKYAN, S., ANTONIK, M., WÖHRL, B. M., RESTLE, T., GOODY, R. S. & SEIDEL, C. A. M. 2003. Multi-parameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 Reverse Transcriptase:primer/template complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 1655-1660.
- SINDBERT, S., KALININ, S., NGUYEN, H., KIENZLER, A., CLIMA, L., BANNWARTH, W., APPEL, B., MÜLLER, S. & SEIDEL, C. A. M. 2011. Accurate distance determination of nucleic acids via Förster resonance energy transfer: implications of dye linker length and rigidity. *J Am Chem Soc*, 133, 2463-80.
- STAHL, Y., GRABOWSKI, S., BLECKMANN, A., KUHNEMUTH, R., WEIDTKAMP-PETERS, S., PINTO, K. G., KIRSCHNER, G. K., SCHMID, J. B., WINK, R. H., HULSEWEDE, A., FELEKYAN, S., SEIDEL, C. A. & SIMON, R. 2013. Moderation of Arabidopsis root stemness by CLAVATA1 and ARABIDOPSIS CRINKLY4 receptor kinase Complexes. *Current Biology*, 23, 362-71.
- STRAUME, M., FRASIER-CADORET, S. G. & JOHNSON, M. L. 1991. Least-Squares Analysis of Fluorescence Data. In: LAKOWICZ, J. R. (ed.) *Topics in Fluorescence Spectroscopy, Volume 2: Principles*. New York: Plenum Press.
- STRYER, L. 1978. Fluorescence energy transfer as a spectroscopic ruler. *Annu Rev Biochem*, 47, 819-46.
- STRYER, L. & HAUGLAND, R. P. 1967. Energy transfer: a spectroscopic ruler. *Proc Natl Acad Sci U S A*, 58, 719-26.
- SUSTARSIC, M. & KAPANIDIS, A. N. 2015. Taking the ruler to the jungle: single-molecule FRET for understanding biomolecular structure and dynamics in live cells. *Current Opinions in Structural Biology*, 34, 52-59.
- VÖPEL, T., HENGSTENBERG, C. S., PEULEN, T.-O., AJAJ, Y., SEIDEL, C. A. M., HERRMANN, C. & KLARE, J. P. 2014. Triphosphate induced dimerization of human guanylate binding protein 1 involves association of the C-terminal helices: A joint double electron-electron resonance and FRET study. *Biochemistry*, 53, 4590-4600.
- WAN, W. W., WANG, Y.-S. & LIU, W. R. 2012. Genetically Encoding Bioorthogonal Functional Groups for Site-selective Protein Labeling. *Organic Chemistry: Current Research*, 1, 1000-e111.

## Chapter C - A generalized description of multidimensional single-molecule FRET histograms: Theoretical treatment of FRET parameters of multistate dynamics systems

Hugo Sanabria<sup>1,2\*</sup>, Thomas-Otavio Peulen<sup>1\*</sup>, Stanislav Kalinin<sup>1</sup>, Suren Felekyan<sup>1</sup>, Claus A.M. Seidel<sup>1</sup>

<sup>1</sup>Institut für Physikalische Chemie, Lehrstuhl für Molekulare Physikalische Chemie, Heinrich Heine Universität, Düsseldorf, Germany, <sup>2</sup>Department of Physics and Astronomy, Clemson University, Clemson USA

\*Contributed equally

### 1 Introduction

Single-molecule fluorescence spectroscopy has the ability to “track” the state of labeled biomolecules by measuring a set of fluorescence observables over time (e.g. absorption and fluorescence  $F(\lambda_A, \lambda_F)$  (Tamarat et al., 2000), fluorescence brightness, fluorescence quantum yields  $\Phi_F$  (Chen et al., 1999, Fries et al., 1998, Kask et al., 1999), fluorescence lifetimes  $\tau$  (Tellinghuisen et al., 1994, Zander and Drexhage, 1996) and the anisotropy  $r$  (Ha et al., 1999, Schaffer et al., 1999)). Multiparameter Fluorescence Detection (MFD) measures these parameters simultaneously and thus solves most problems commonly associated with quantitative single-molecule (sm) experiments (Kühnemuth and Seidel, 2001, Felekyan et al., Felekyan et al., 2012, Eggeling et al., 1998, Eggeling et al., 2001, Rothwell et al., 2003, Rothwell et al., 2013). Typically, single-molecule MFD experiments use Förster Resonance Energy Transfer (FRET) to interrogate the time evolution of the conformational state of biomolecules (Weiss, 1999, Weiss, 2000, Felekyan et al.). Thus, single-molecule FRET (smFRET) in combination with MFD has been extensively applied to study the conformational transitions of nucleic acids, nucleic binding proteins and proteins (Sisamakos et al., 2010, Felekyan et al., 2012, Widengren et al.) and, more recently, they were used to determine near atomistic structural models of biomolecules (Kalinin, 2012). FRET-indicators e.g. FRET efficiency ( $E$ ), proximity ratio ( $PR$ ), fluorescence intensity ratio of donor over acceptor ( $F_D/F_A$ ) and fluorescence lifetimes as the averaged donor fluorescence lifetime per single-molecule burst ( $\langle\tau_{D(A)}\rangle_F$ ) are used to select species-specific parameters such as the interdye distance. Understanding the mathematical and physical relationship among typical FRET-indicators in single-molecule experiments is a topic of extensive research (Gopich and Szabo, 2003, Gopich and Szabo, 2005, Gopich and Szabo, 2007, Gopich and Szabo, 2012, Antonik et al., 2006, Kalinin et al., 2007, Kalinin et al., 2010a, Nir et al., 2006). The relationship between these FRET-indicators are referred to as “FRET-lines” which were introduced to relate FRET observables such as the steady-state transfer efficiency  $E$  and the fluorescence averaged lifetime of a donor in presence of an acceptor  $\langle\tau_{D(A)}\rangle_F$ .

They were applied to explain experimental results using multi-dimensional MFD histograms (Rothwell et al., 2003, Antonik et al., 2006, Kalinin et al., 2007, Kalinin et al., 2010a, Sindbert et al., 2011). Hidden in the FRET-lines is the possibility to understand the mechanisms behind biomolecular dynamics and interactions. More recently, Szabo's group presented a theoretical approach to treat these FRET-lines using photon statistics; but only simplified cases were considered (Gopich and Szabo, 2012). Thus, in this paper, we include in the FRET-lines the effects of the dye linker dynamics, detection efficiencies, quantum yields of fluorophores, multi-exponential properties of dyes, exchange rates in complex dynamic networks and presented a simplified framework for their calculation.

### 1.1 Theoretical background: Förster theory

The aim in FRET-experiments, is to measure the distance between carefully selected dyes. As the inter-dye distance is not a direct observable, it can only be deduced by the rate constant of energy transfer from a donor (D) to an acceptor (A) fluorophore. As the rate of energy transfer, due to the dipole-dipole coupling of the electronic states of the fluorophores, depends strongly on the distance between fluorophores, the rate constant of energy transfer ( $k_{RET}$ ) depends on donor-acceptor distance  $R_{DA}$  as well as on their mutual dipole orientation reflected by the orientation-factor  $\kappa^2$ . This is described by the Förster relationship (Förster, 1948):

$$k_{RET} = \kappa^2 \cdot k_{F,D} \cdot \left( \frac{R_{0J}}{R_{DA}} \right)^6 \quad (1)$$

Here  $R_{0J}$  is the reduced Förster-radius specific for each DA-pair, and  $k_{F,D}$  is the radiative rate constant of the donor fluorophore. The reduced Förster-radius can be determined using the following relation:

$$R_{0J} = \sqrt[6]{9 / (128 \cdot \pi^5 \cdot N_A \cdot n^4) \cdot J} \quad \text{with} \quad J = \int f_D(\lambda) \cdot \varepsilon_A(\lambda) \cdot \lambda^4 \cdot d\lambda \quad (2)$$

where  $N_A$  is the Avogadro's constant,  $n$  is the refractive index of the medium and  $J$  is the overlap integral between  $f_D(\lambda)$ , the donor emission spectrum and  $\varepsilon_A(\lambda)$ , the acceptor absorption spectrum.

For a single species with a fixed inter-dye distance,  $k_{RET}$  relates to the yield of the FRET-process ( $E$ ), which transfers energy from D to A. This yield  $E$  is commonly known as the FRET-efficiency. The FRET-efficiency is the fraction of energy that is transferred per donor excitation event. Thus, it is given by the ratio of the FRET-rate constant ( $k_{RET}$ ) over the sum of all de-excitation rate constants:

$$E = \frac{k_{RET}}{k_D + k_{RET}} \quad (3)$$

Here,  $k_D$  is the deactivation rate constant of the donor from its excited fluorescence state in the absence of FRET. This relationship is only valid in the absence of other de-excitation processes.

If a single doubly-labeled biomolecule exchanges between different structural states due to conformational dynamics, excreting various inter-dye distances, the fluorescence of multiple FRET states has to be considered. Each of the resulting FRET-state is associated to its own FRET-rate constant and its corresponding FRET-efficiency. Considering that the fluorescence intensity per species is additive, we first describe the fluorescence of a single species and extend this concept to include multiple species.

## 1.2 FRET efficiency of single- and multiple- species

### *Single species*

For a single doubly-labeled molecule where the donor fluorophore is excited by a single photon at time  $t = 0$ , the expected population of the fluorescent donor state, in the absence ( $n_{D|D}^{D0}(t)$ ) and in presence ( $n_{D|D}^{DA}(t)$ ) of an acceptor, the expected population of the fluorescent acceptor state excited through the donor by energy transfer ( $n_{A|D}^{DA}(t)$ ), and the expected population of a directly excited acceptor ( $n_{A|A}^{A0}(t)$ ) are given by:

$$\begin{aligned} n_{D|D}^{D0}(t) &= e^{-k_D t} \\ n_{D|D}^{DA}(t) &= e^{-(k_{RET} + k_D)t} = n_{D|D}^{D0}(t) \cdot e^{-k_{RET} t} \\ n_{A|D}^{DA}(t) &= \frac{k_{RET}}{k_{RET} + k_D - k_A} \cdot (e^{-k_A t} - e^{-(k_{RET} + k_D)t}) \\ n_{A|A}^{A0}(t) &= e^{-k_A t} \end{aligned} \quad (4)$$

Here,  $k_D$ ,  $k_{RET}$  and  $k_A$  are the rate constants of the donor deactivation (in the absence of an acceptor), by Resonance Energy Transfer (RET) and the acceptor deactivation, respectively. To identify the excitation and emission process we used the notations  $n_{D|D}^{D0}$ ,  $n_{D|D}^{DA}$ , and  $n_{D|D}^{D0}$  to represent donor excitation ( $n_{D|D}^{D0}$ ) and donor emission ( $n_{D|D}^{D0}$ ) in presence ( $n_{D|D}^{DA}$ ) and absence ( $n_{D|D}^{D0}$ ) of the acceptor. Thus, for acceptor sensitized emission we have  $n_{A|D}^{DA}$ .

The depopulation from the excited state in eq. (4) can be rewritten in terms of the donor fluorescence lifetimes in the absence ( $\tau_{D(0)} = 1/k_D$ ) and in the presence of the acceptor ( $\tau_{D(A)} = 1/(k_D + k_{RET})$ ), and acceptor fluorescence lifetime ( $\tau_A = 1/k_A$ ) as:

$$\begin{aligned}
 n_{D|D}^{D0}(t) &= e^{-\frac{t}{\tau_{D(0)}}} \\
 n_{D|D}^{DA}(t) &= e^{-\frac{t}{\tau_{D(A)}}} \\
 n_{A|D}^{DA}(t) &= \frac{\tau_A \cdot (\tau_{D(0)} - \tau_{D(A)})}{\tau_{D(0)}\tau_A - \tau_{D(A)}\tau_A - \tau_{D(A)}\tau_{D(0)}} \cdot \left( e^{-\frac{t}{\tau_A}} - e^{-\frac{t}{\tau_{D(A)}}} \right) \\
 n_{A|A}^{A0}(t) &= e^{-\frac{t}{\tau_{A(0)}}}
 \end{aligned} \tag{5}$$

The fluorescence intensity represents each depopulation event accumulated over time. For a given fluorophore (either donor or acceptor), the fluorescence intensity is given by the product of the radiative rate ( $k_F$ ) of the respective fluorophore, the fluorescence quantum yield ( $\Phi_F$ ) and its time-dependent population of the excited state as:

$$\frac{f(t)}{\Phi_F} = N_0 \cdot k_F \cdot n(t) \tag{6}$$

where  $N_0$  is the number of labeled molecules. Therefore, the donor fluorescence intensity decays, in absence  $f_{D|D}^{D0}(t)$  and in the presence of acceptor  $f_{A|D}^{DA}(t)$ , are directly proportional to the probability of finding the donor in its excited state. In eq. (6), we introduced the fluorescence quantum yield ( $\Phi_F$ ), which for donor and acceptor fluorophores are denoted as  $\Phi_F^{A0}$  and  $\Phi_F^{D0}$ , respectively. The fluorescence quantum yield is the fraction of emitted photons per each depopulation event. It represents the first correction of the fluorescence emission, which considered that fluorophores are not ideal emitters.

The next step is to relate the time dependent fluorescence emission to the steady-state or integrated fluorescence which could be used in measuring the FRET efficiency and thus the rate constant of the energy transfer. For this, we introduce for convenience the FRET-induced donor-decay  $\varepsilon_D(t)$  as factorization of the fluorescence in presence ( $f_{D|D}^{DA}$ ) and absence ( $f_{D|D}^{D0}$ ) of the acceptor:

$$f_{D|D}^{DA}(t) = \varepsilon_D(t) \cdot f_{D|D}^{D0}(t) \tag{7}$$

If the FRET-induced donor decay is independent of other deactivation processes,  $\varepsilon_D(t)$  is solely a function of the FRET-rate constants and the steady-state transfer efficiency ( $E$ ) is given by:

$$E = 1 - \frac{\int \varepsilon_D(t) \cdot f_{D|D}^{D0}(t) \cdot dt}{\int f_{D|D}^{D0}(t) \cdot dt} = 1 - \frac{F_{D|D}^{DA}}{F_{D|D}^{D0}} \tag{8}$$

Taking the integral above for a single FRET-rate constant ( $\varepsilon_D(t) = \exp(-k_{RET} \cdot t)$ ) reveals, that the FRET-efficiency ( $E$ ) is directly related to the fluorescence lifetime of the donor in presence ( $\tau_{D(A)}$ ) and absence ( $\tau_{D(0)}$ ) of FRET:

$$E = 1 - \frac{\tau_{D(A)}}{\tau_{D(0)}} \quad (9)$$

Above we used the definition of the steady-state fluorescence emission as

$$\begin{aligned} \frac{F_{D/D}^{D0}}{\Phi_F^{D0}} &= \int f_{D/D}^{D0}(t) \cdot dt \\ \frac{F_{D/D}^{DA}}{\Phi_F^{D0}} &= \int \varepsilon_D(t) \cdot f_{D/D}^{D0}(t) \cdot dt \end{aligned} \quad (10)$$

In terms of the FRET-sensitized acceptor emission, it is possible to rewrite eq. (8) as

$$E = \frac{F_{D/D}^{D0} - F_{D/D}^{DA}}{F_{D/D}^{D0}} = \frac{F_{A/D}^{DA} / \Phi_F^{A0}}{F_{A/D}^{DA} / \Phi_F^{A0} + F_{D/D}^{DA} / \Phi_F^{D0}} \quad (11)$$

where it can be seen that the FRET-efficiency represents the ratio of the acceptor excitations number through FRET mechanism over the total number of donor excitations. It can be easily shown that eqs. (3, 8, 11) are equivalent. Comparing the same equation, it was identified that FRET-efficiency can be measured by intensity based and time-resolved methods.

### ***Mixture of multiple species***

For a single species eq. 9 and eq. 11 relate the FRET-efficiency to the fluorescence intensities and fluorescence lifetimes. As samples are composed of species mixtures, these equations are often inappropriate. For instance, the sample may be contaminated by a small residual fraction of molecules labeled only by a donor fluorophore. In more complicated cases the molecule of interest may exchange among multiple conformations, where each conformation results in a different FRET-rate constant.

For heterogeneous mixtures, it is important to have in mind that the fluorescence intensities are additive and weighted by their respective a species fractions ( $x_i$ ). Furthermore, the integral of the fluorescence decay  $f(t)$  represents the total number of fluorescent photons emitted by all species. Considering this, we extend eq. (6) for a mixture of labeled species ( $i$ ) as:

$$f(t) = N_0 \cdot k_F \cdot \sum_i x_i \cdot \Phi_{F,i} \cdot n_i(t) \quad (12)$$

Using this relationship, it is possible to define various averaged lifetimes:

$$\begin{aligned}
 \frac{\tau_{F,D}}{\Phi_{F,D} \cdot N_0} \int f_{D|D}^{D0}(t) \cdot dt &= \langle \tau_{D(0)} \rangle_x \\
 \frac{\tau_{F,D}}{\Phi_{F,D} \cdot N_0} \int f_{D|D}^{DA}(t) \cdot dt &= \langle \tau_{D(A)} \rangle_x \\
 \frac{\tau_{F,D}}{\Phi_{F,D} \cdot N_0} \int f_{A|D}^{DA}(t) \cdot dt &= \langle \tau_{D(0)} \rangle_x - \langle \tau_{D(A)} \rangle_x
 \end{aligned} \tag{13}$$

Here the brackets  $\langle \dots \rangle_x$  denote the time average weighted by the species fractions. Also, we have assumed that each donor emission process shares the same fluorescence quantum yield. For a multi-exponential fluorescence decays mixture the integrals shown in eq. (13) are given by an species fraction weighted average fluorescence lifetime ( $\langle \tau \rangle_x = \sum_i x_i \cdot \tau_i$ ).

### Experimental observables

#### Fluorescence intensities

Experimentally, the FRET-efficiency is determined by the donor and acceptor fluorescence intensities ( $F_D$  and  $F_A$ ). These intensities are obtained by the total detected signal in two spectral windows. The donor signal ( $S_G$ ) is typically detected over a “green” wavelength window and the acceptor signal ( $S_R$ ) over a “red” wavelength window. It is important to consider that the used detectors for the “green” and “red” detection channel are non-ideal and have their own detection efficiency, for green ( $g_G$ ) and red ( $g_R$ ), respectively. These detection efficiencies are the second layer of corrections. In addition, to the fluorescent photons background photons (e.g.  $B_G$  for green and  $B_R$  for red channels) contribute to the signal in each detection channel. With these considerations, the measured signals are related to the fluorescence intensities by:

$$F_D \equiv F_{D|D}^{DA} = \frac{S_G - \langle B_G \rangle}{g_G} \tag{14}$$

$$F_A \equiv F_{A|D}^{DA} = \frac{S_R - \alpha F_G - \langle B_R \rangle}{g_R} \tag{15}$$

$$\frac{F_D}{F_A} \equiv \frac{F_{D|D}^{DA}}{F_{A|D}^{DA}} = \frac{g_R}{g_G} \cdot \frac{S_G - \langle B_G \rangle}{S_R - \alpha F_G - \langle B_R \rangle} \tag{16}$$

For simplicity, we omitted the explicit notation and the brackets  $\langle \dots \rangle$  to indicate time averages.

#### Time-resolved fluorescence

In time-resolved experiments the fluorescence intensity decay of the donor in presence of an acceptor  $f_{D|D}^{DA}(t)$  is monitored via a fluorescence decay histogram to reveal, in combination with the fluorescence decay in absence of an acceptor  $f_{D|D}^{D0}(t)$ , information on FRET via  $\varepsilon_D(t)$  the

FRET-induced donor decay (eq. 7). If the depopulation of the excited fluorescent state of the donor by FRET is independent of other pathways  $\varepsilon_D(t)$  directly relates to  $p(R_{DA})$ , the donor acceptor distance distribution:

$$\varepsilon_D(t) = \int p(R_{DA}) \cdot \exp\left(-k_F \cdot \left(\frac{R_0}{R_{DA}}\right)^6\right) \quad (17)$$

In ensemble (cuvette) experiments, the number of detected photons is sufficiently high to recover the  $p(R_{DA})$  even for heterogeneous mixtures. However, in smFRET-experiments molecules the time a freely single diffusing molecule is observed is very brief time (in the order of milliseconds) and only a limited number of photons are detected (typically up to a few hundred). Thus, the recorded fluorescence decay histograms, e.g. the fluorescence decay of the donor  $f_{D|D}^{DA}(t)$ , are very noisy and only the average time between excitation and fluorescence is determined in addition to the FRET-efficiency.

Under ideal conditions, i.e. excitation of the sample by a  $\delta$ -pulse at time  $t = 0$ , this average time (the average fluorescence intensity weighted lifetime  $\langle\tau_{D(A)}\rangle_F$ ) is given by the fluorescence intensity weighted average:

$$\langle\tau_{D(A)}\rangle_F = \frac{1}{N_0 \cdot k_{F,D}} \cdot \int t \cdot f_{D|D}^{(D,A)}(t) \cdot dt \quad (18)$$

If the fluorescence decay is multi-exponential the fluorescence intensity weighted lifetime  $\langle\tau_{D(A)}\rangle_F$  is given by an averaged sum of the fluorescence lifetimes of the individual species:

$$\langle\tau_{D(A)}\rangle_F = \frac{\sum_i x_i \cdot \tau_i^2}{\sum_i x_i \cdot \tau_i} = \frac{\sum_i x_i \cdot \tau_i^2}{\langle\tau_{D(A)}\rangle_x} \quad (19)$$

Practically, to consider the instrument response function and the background photons explicitly,  $\langle\tau_{D(A)}\rangle_F$  is determined by maximum likelihood estimators (MLE) (Maus et al., 2001).

#### *Moments accessible by smFRET*

To sum up, in smFRET-experiments typically the average time between excitation and detection of fluorescence (the fluorescence weighted average lifetime  $\langle\tau_{D(A)}\rangle_F$ ) and the FRET-efficiency ( $E$ ) are determined. The FRET-efficiency is proportional to  $\langle\tau_{D(A)}\rangle_x$ , the species fraction weighted lifetime of the donor:

$$E = 1 - \frac{\int \varepsilon_D(t) \cdot f_{D|D}^{D0}(t) \cdot dt}{\int f_{D|D}^{D0}(t) \cdot dt} = 1 - \frac{F_{D|D}^{DA}}{F_{D|D}^{D0}} = 1 - \frac{\langle\tau_{D(A)}\rangle_x}{\langle\tau_{D(0)}\rangle_x} \quad (20)$$

It follows that for every detected single-molecule event the first two moments on the lifetime distribution can be determined. These moments are: i) the species weighted average lifetime and ii) the fluorescence weighted average lifetime (or mean arrival time). The species weighted lifetime,  $\langle \tau_{D(A)} \rangle_x$ , is accessible by the integrated signals (or transfer efficiency) while the fluorescence weighted lifetime,  $\langle \tau_{D(A)} \rangle_F$ , is directly determined by the mean arrival time. For a multi-exponential fluorescence decay of the donor both are given by the following sums:

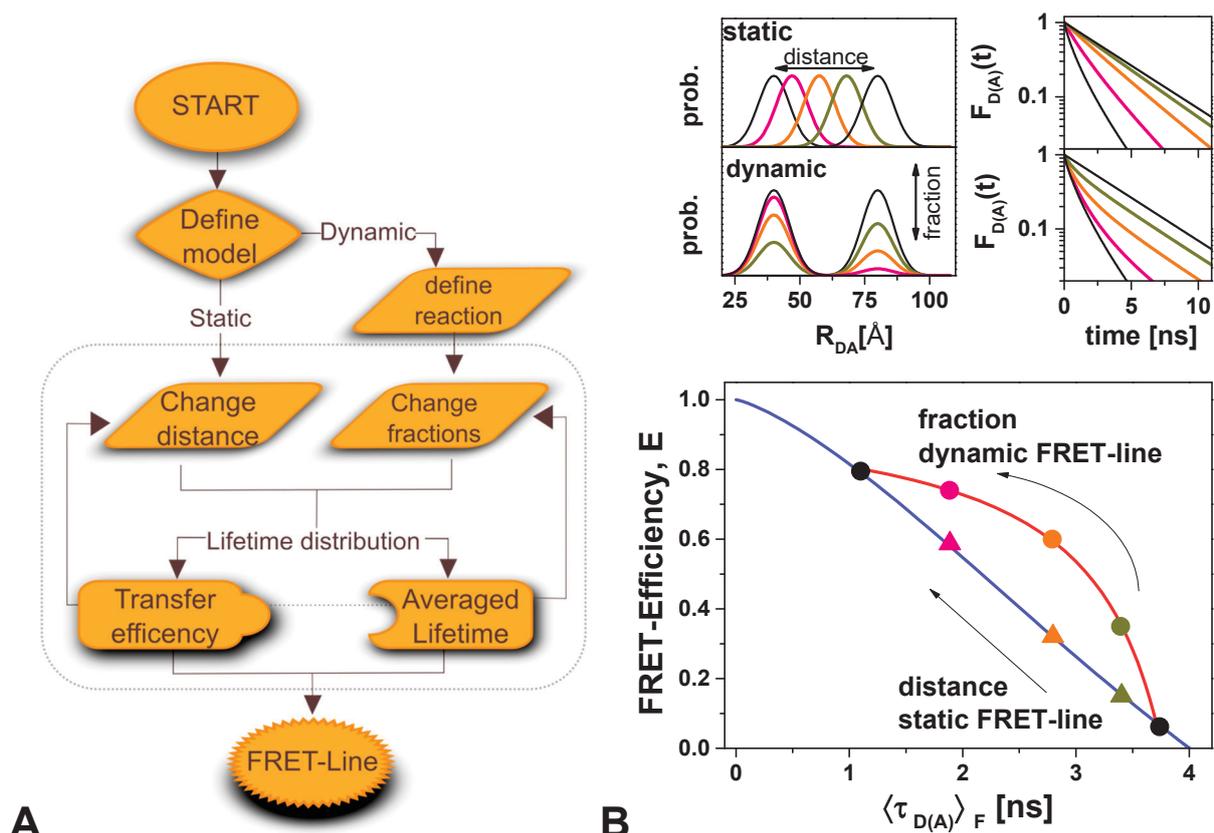
$$\begin{aligned} \langle \tau_{D(A)} \rangle_x &= \sum_i x_i \cdot \tau_i = \sum_i \tau_{D(A)}^{(i)}(R_{DA}) \cdot x^{(i)}(R_{DA}), \\ \langle \tau_{D(A)} \rangle_F \cdot \langle \tau_{D(A)} \rangle_x &= \sum_i x_i \cdot \tau_i^2 = \sum_i x^{(i)}(R_{DA}) \cdot \left( \tau_{D(A)}^{(i)}(R_{DA}) \right)^2, \end{aligned} \quad (21)$$

Using these sums, which can be approximated by integrals, a FRET-lines can be calculated to relate species weighted fluorescence intensities (e.g. steady-state transfer efficiency  $E$  or donor-acceptor fluorescence ratio  $F_D/F_A$ ) to the fluorescence averaged lifetimes. This allows for an easier interpretation of experimental results and serves as a visual guide. A FRET-line is defined by as a parametric curve, which can be thought of as projection a varied model parameter to observable experimental parameters. For instance, the DA-distance  $R_{DA}$  is changed and for every  $R_{DA}$  the fluorescence weighted average lifetime  $\langle \tau_{D(A)} \rangle_F$  and the FRET-efficiency  $E$  are calculated. This relates changes in  $R_{DA}$  to changes of  $\langle \tau_{D(A)} \rangle_F$  and the FRET-efficiency  $E$  gives rise to a line in the  $E$  vs  $\langle \tau_{D(A)} \rangle_F$  - plane.

## 2 FRET-lines

All parameterized changes of  $\varepsilon_D(t)$  define a FRET-line. However, two parameterizations of  $\varepsilon_D(t)$  are particularly useful and lead to FRET-lines coined “static” and “dynamic” FRET-lines. To compute a FRET-line, an underlying model function of the FRET-rate constants distribution or  $p(R_{DA})$  is required. Static FRET-lines are generated by changing a parameters representative for the DA-distance (see Figure 1) and are historically the oldest (Rothwell et al., 2003). Dynamic FRET-lines are obtained by models composed of multiple distinct states (at least two). Changes of the relative species fractions of the states lead to FRET-lines which describe the exchange among the respective states (Figure 1). More details on models for calculating FRET-lines are presented in the following sections.

As future reference, we present in Fig. 1 a work flow diagram which rationalizes how to obtain various FRET-lines. The first step is to select the proper model for the FRET-line. There are two options: a) calculate the static FRET-line, which corresponds to the “static” relation of the fluorescence weighted average lifetime with respect to the intensity based FRET indicator and, b) calculate dynamic FRET-line: this step describes the case where there is dynamic interconversion of species within the duration of a single-molecule event. The next step is to use the defined model of the FRET-rate constants distribution or  $p(R_{DA})$  in order to calculate the corresponding FRET indicators (e.g.  $F_D/F_A$  or  $E$ ). Figure 1B shows an example on how the static and dynamic FRET-lines are calculated for the distribution of distances observed on the top left. In both cases, static and dynamic FRET-lines are shown. In the top right of Fig. 1B, the distribution of DA-distances and corresponding fluorescence intensity decays are presented, color coded for each case. Therefore, each point on the  $E$  vs  $\langle \tau_{D(A)} \rangle_F$  - plane, shown below in Figure 1B (triangles and circles for static and dynamic FRET-lines, respectively), are obtained using the corresponding distributions on top. The static FRET-line follows the blue line and the dynamic FRET-line starts and ends at two different points on the static FRET-line.



**Figure 1.** Generation of FRET-lines (A) workflow performed in calculation of static and dynamic FRET-lines. (B) Examples of FRET-lines determined by the presented workflow for the static and the dynamic case (lower panel). To illustrate the effect of the dynamic-shift the distance distribution and the resulting time-resolved fluorescence intensity decays are illustrated for a few examples in the upper and middle panel. The distances and species fractions of the displayed decays were chosen to result in identical fluorescence weighted average lifetimes. In the dynamic model the two limiting states given by the mean distances of 40 and 80 Å. The species fractions of the 40 Å state are 0.0, 0.39, 0.71, 0.92 and 1.0. The mean distances of the static FRET-line are 40, 47, 58, 68 and 80 Å. The resulting transfer efficiencies of the displayed distributions are highlighted on the dynamic and static FRET-lines by triangles (static) and circles (dynamic). In all cases a Förster-radius of  $R_0=52$  Å and a donor-lifetime of  $\tau_0 = 4$  ns and Gaussian distributed donor-acceptor distances with a width of 6 Å were used.

## 2.1 FRET-lines for static host-molecules

So far, the origins of the FRET-rate constant distribution  $p(k_{RET})$  were not discussed. Mainly  $p(k_{RET})$  arises from the fact that dyes are typically linked to the molecule under study via long flexible linkers, especially for those used in single-molecule experiments. The resulting spatial probability distributions of the dyes leading to a distribution of DA-distances  $p(R_{DA})$  which, in turn generates a FRET-rate constants distribution  $p(k_{RET})$  mentioned throughout. However, before we consider the influence of the linker effect we first describe the original static FRET lines without any corrections.

### 2.1.2 Absence of dye-linker dynamics

Based on the introduction of FRET observables ( $F_D$ ,  $F_A$ ,  $F_D/F_A$ ,  $E$ ,  $\langle\tau_{D(A)}\rangle_x$  or  $\langle\tau_{D(A)}\rangle_F$ ) for a single-FRET species, it is possible to combine eqs. (9) and (11) to relate time-resolved and intensity based observables as:

$$\frac{F_D}{F_A} = \frac{\Phi_F^{D0}}{\Phi_F^{A0}} \cdot \left( \frac{\tau_{D(0)}}{\tau_{D(A)}} - 1 \right)^{-1},$$

$$E = \frac{1}{1 + \frac{\Phi_F^{A0}}{\Phi_F^{D0}} \cdot \frac{F_D}{F_A}} = 1 - \frac{\tau_{D(A)}}{\tau_{D(0)}}.$$
(22)

These two equations are the simplest forms of a static FRET-lines. To calculate these FRET-lines, we implicitly “varied” the FRET-rate between zero and infinity to generate a curve that relates all possible values of the FRET-efficiency  $E$ , or the intensity ration  $F_D/F_A$  to the all “possible” donor fluorescence lifetimes,  $\tau_{D(A)}$ .

In an such experiment, only one FRET-rate is expected, which arises from a single donor lifetime; thus, the species weighted and the fluorescence weighted average lifetimes are equal,  $\tau_{D(A)} = \langle\tau_{D(A)}\rangle_x = \langle\tau_{D(A)}\rangle_F$ . In order to distinguish static populations in a heterogeneous mixture the FRET-lines (eq. (22)) are used as reference lines in MFD histograms, where the FRET-efficiency  $E$  or the green-red fluorescence intensity ratio  $F_D/F_A$  are plotted versus the fluorescence weighted average donor lifetime  $\langle\tau_{D(A)}\rangle_F$ . Several applications of these simple rules have been demonstrated over the years (Widengren et al., 2006, Sisamakias et al., 2010).

### 2.1.3 Presence of dye-linker dynamics

In single-molecule experiments, a DA-labeled molecule with a single donor lifetime  $\tau_{D(0)}$  and a given FRET-rate constant, will be represented by a point lying on the static FRET-line. A mixture of DA labeled samples with different DA-distances results in various points along the static FRET-line. In real experiments, broadening in both dimensions is observed. This is partly because of shot noise and contribution due to the acceptor photophysics (Kalinin et al., 2007). Recent experimental evidence has shown that eq. (22) does not fully describe MFD/smFRET experiments. One of the sources of discrepancies is due to the use of large chemical linkers that couple the fluorophore to the biomolecules (Sindbert et al., 2011). It was shown that ensemble fluorescence intensity decays are best described by a distribution of lifetimes that represent the DA-distance distributions originating from the mobility of the donor and acceptor dyes (Sindbert et al., 2011). Understanding the spatial probability distribution, which can be characterized by accessible volumes and localization of dyes are current topics of intensive research

(Sindbert et al., 2011, Andrecka et al., 2009, Muschielok and Michaelis, 2011). More recently, there has been evidence of the translational diffusion coefficient of the dyes is in the order of 5 to 10 Å<sup>2</sup>/ns while their rotational correlation time is in the order of 100<sup>th</sup> of picoseconds (Kalinin et al., 2015). This slow translational diffusion assures that the FRET-rate constants distribution  $p(k_{RET})$  can be well approximated by the distance distribution  $p(R_{DA})$  (Kalinin et al., 2007), which is sampled during the acquisition time. Hence, a distribution of FRET-efficiencies and donor fluorescence lifetimes is expected (Haas et al., 1975).

To account for this effect, we assume that the spatial distribution of the donor- and acceptor-dyes follows a normal distribution centered at the respective mean dye positions; then, all the DA-distance distribution is distributed according to the non-central chi-squared distribution. However, as typical dye-linker length ( $\sim 20$  Å) and average DA-distance are larger than 35 Å we approximate the non-central chi-squared distribution by a normal distribution with standard deviation  $\sigma_{DA}$  and mean inter-dye distance  $\bar{R}_{DA}$ . Thus, the FRET-induced donor decay  $\varepsilon_D(t)$  is given by:

$$\varepsilon_D(t, \bar{R}_{DA}) = \int p(R_{DA}, \bar{R}_{DA}, \sigma_{R_{DA}}) \cdot \exp\left(-\kappa_{DA}^2 \cdot k_F \cdot (R_0/R_{DA})^6 \cdot t\right) dR_{DA}, \text{ with}$$

$$p(R_{DA}) = \frac{1}{\sqrt{2\pi} \cdot \sigma_{R_{DA}}} \exp\left(-\frac{(R_{DA} - \bar{R}_{DA})^2}{2 \cdot \sigma_{R_{DA}}^2}\right). \quad (23)$$

To compute a FRET-line which includes the dye-linker distribution, we vary the mean DA-distance ( $\bar{R}_{DA}$ ) and calculate the fluorescence averaged lifetime and the FRET-efficiency. Typically, we consider a fixed standard deviation  $\sigma_{DA} \sim 6$  Å that satisfies benchmarking experiments (Kalinin et al., 2010b, Sindbert et al., 2011). The effect of dye linker distributions on the relation between the transfer-efficiency and the fluorescence weighted average lifetime was first described by Kalinin et al. There, as the integrals (eq. 21) are difficult to solve analytically, the relation between  $E$  and  $\langle \tau_{D(A)} \rangle_F$  was solved numerically (Kalinin et al., 2010b).

Given tabulated integrals of eq. (21), the simplest way to relate the fluorescence weighted average lifetimes  $\langle \tau_{D(A)} \rangle_F$  to the species weighted average lifetimes  $\langle \tau_{D(A)} \rangle_x$  is and use an empirical polynomial function with coefficients  $A^{(i)}$  that relates them. This polynomial can be reused for similar systems as long as the variance of the distance distribution does not change. The polynomial function has the form of:

$$\langle \tau_{D(A)} \rangle_x = \sum_{i=0}^n A^{(i)} \cdot \langle \tau_{D(A)} \rangle_F^i \quad (24)$$

Typically, a third order polynomial function satisfies the relationship between  $\langle \tau_{D(A)} \rangle_F$  and  $\langle \tau_{D(A)} \rangle_x$ . Several examples of the  $A^{(i)}$  coefficients for various  $\sigma_{R_{DA}}$  are shown in the appendix (Table A1).

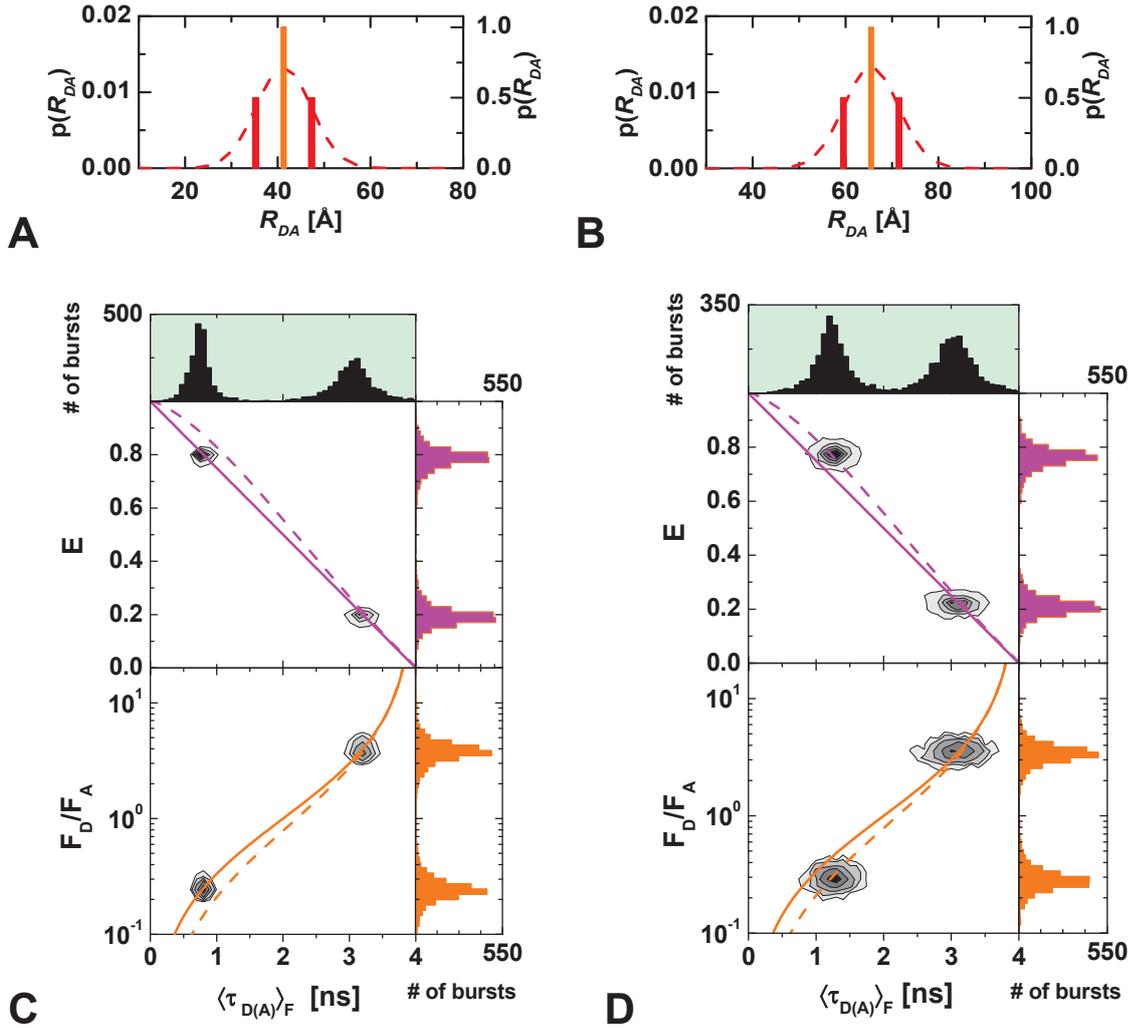
Finally, the corrected FRET-line that includes the dynamics of the dye linkers is modified from Eq. (22) to include the internal averaging on the fluorescence lifetimes as

$$\frac{F_D}{F_A} = \frac{\Phi_F^{(D0)}}{\Phi_F^{(A0)}} \cdot \left( \frac{\tau_{D(0)}}{\langle \tau_{D(A)} \rangle_x} - 1 \right)^{-1} = \frac{\Phi_F^{(D0)}}{\Phi_F^{(A0)}} \cdot \left( \frac{\tau_{D(0)}}{\sum_{i=0} A^{(i)} \cdot \langle \tau_{D(A)} \rangle_F^i} - 1 \right)^{-1}, \quad (25)$$

$$E = \frac{1}{1 + \frac{\Phi_F^{(A0)}}{\Phi_F^{(D0)}} \cdot \frac{F_D}{F_A}} = 1 - \frac{\langle \tau_{D(A)} \rangle_x}{\tau_{D(0)}} = 1 - \frac{\sum_{i=0} A^{(i)} \cdot \langle \tau_{D(A)} \rangle_F^i}{\tau_D^{(D,0)}}.$$

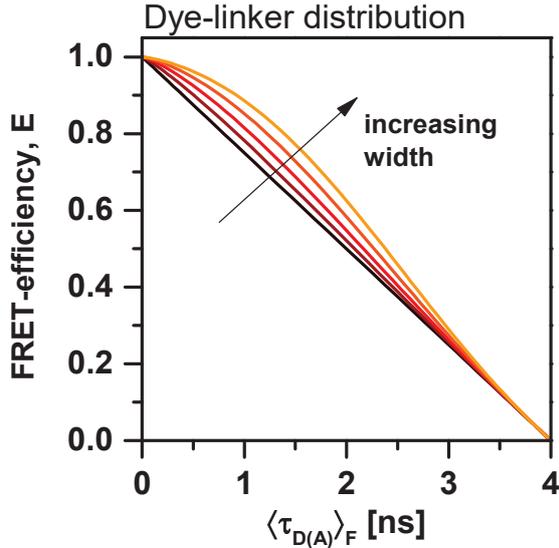
An important note to consider is that the standard deviation of the DA-distance distribution is not to be confused with the broadening often observed in smFRET histograms due to blinking of the acceptor fluorophore (Kalinin et al., 2010a). Acceptor effects can be observed by Photon Distribution Analysis (PDA) (Kalinin et al., 2007). Broadening caused by dye-linker dynamics, observed in the nanosecond regime by TCSPC, is usually not visible in FRET-efficiency histograms, as the dynamics is fast compared to the burst integration time ( $\sim 1$ ms).

To illustrate the effect of the linker dynamics we performed Monte Carlo simulations to generate smFRET data. Two sets of data were generated. *i*) single states with fixed  $R_{DA}$  (Orange bar plot on Fig. 2A and 2B for two different FRET species) and *ii*) two discrete fast interconverting states that approximate a Gaussian distribution  $p(R_{DA}) = \bar{R}_{DA} \pm \sigma_{R_{DA}}$ . The simulations highlight that distribution of distances shifts the populations away from the static FRET-line (eq. (22); solid line on Fig 2C and 2D). Furthermore, the effect of this distribution marked by the linker-corrected FRET-line (dashed lines in Figures 2C and 2D) is more obvious at higher FRET efficiencies. In conclusion, the new static FRET-lines can be used to estimate the variance of the distribution of distances ( $\sigma_{R_{DA}}$ ) for static populations.



**Figure 2.** Monte Carlo simulation of smFRET data of two static subpopulations with  $E_1 = 0.2$  ( $R_{DA} = 41.3 \text{ \AA}$ ) and  $E_2 = 0.8$  ( $R_{DA} = 65.5 \text{ \AA}$ ) with and without linker dynamics. (A-B) Show distance distribution of states, orange bars represent no linker dynamics. To simulate the linker dynamics and the Gaussian distribution of distances (Dashed red line) each population was split into two subpopulation with  $p(R_{DA}) = R_{DA} \pm \sigma_{R_{DA}}$  that dynamically exchange ( $k_{ab}=k_{ba} = 1000 \text{ ms}^{-1}$ ). The standard deviation was  $\sigma_{R_{DA}} = 6 \text{ \AA}$ . (C-D) Two dimensional MFD histograms of FRET indicators  $E$  vs.  $\langle \tau_{D(A)} \rangle_F$  and  $F_D/F_A$  vs.  $\langle \tau_{D(A)} \rangle_F$  for the simulations without and with linker dynamics, respectively. Solid lines represent the FRET lines without the linker correction and dashed lines represent the FRET-lines corrected for linker dynamics with  $\sigma_{DA}$  and  $\tau_{D(0)} = 4.0 \text{ ns}$  as done in Eq. (25) with coefficients  $A^{(0)} = 0.0432$ ,  $A^{(1)} = 0.4934$ ,  $A^{(2)} = 0.2794$ ,  $A^{(3)} = 0.0376$ . One-dimensional histograms are projected to the top and to the right of the MFD histograms. The parameters for the simulations were; the Förster-radius  $R_0 = 52 \text{ \AA}$ , fluorescence quantum yield of acceptor and donor  $\Phi_F^{(D0)} = \Phi_F^{(A0)} = 1$ , the detection efficiency ratio  $g_G/g_R = 1$ ,  $\langle B_G \rangle = 1.5 \text{ kHz}$ ,  $\langle B_R \rangle = 0.7 \text{ kHz}$ ,  $t_{diff} = 1.5 \text{ ms}$ , and crosstalk  $\alpha = 0.0$ .

Additionally, to exemplify how the relationship between the species weighted average lifetimes  $\langle \tau_{D(A)} \rangle_x$  and the fluorescence weighted average lifetimes  $\langle \tau_{D(A)} \rangle_F$  vary depending on the width of the dye-linker distribution, we show in Fig. 3 different lines as we increase the width of the distribution. It is obvious that the major effect is observed when the fluorescence lifetimes are short or when we expect high FRET efficiencies.



**Figure 3.** Effect of the dye linker distribution in the static FRET-line, the sigma of the distribution increases the curvature in the relationship between the species weighted average lifetime  $\langle \tau_{D(A)} \rangle_x$  and the fluorescence weighted average lifetime  $\langle \tau_{D(A)} \rangle_F$  ( $\sigma_{RDA} = 0.1, 4, 6, 8$  and  $10 \text{ \AA}$  from black to orange). The idealized static FRET-line case with no linker effects is displayed in dashed light gray.

Although, the name “static” FRET line suggests non dynamic FRET populations, it turns out that the addition of the linker dynamics corrects for the dynamic motion of the sampling of the fluorophore’s accessible volume at sub microseconds timescales. However, because the dynamics belongs to a single conformational state of the biomolecule we consider it as a static population.

## 2.2 FRET-lines for dynamic host-molecules

Biomolecules are highly dynamic and exchange between different conformations at multiple timescales. Thus, when we carry out experiments, often we do not observe populations following the static FRET lines. To the contrary, often large deviations from the static FRET-line are observed, which could only be accounted to the intrinsic dynamic nature of biomolecules. To explain the dynamic FRET lines, first we first refrain from the use of the linker-dynamic motion. Thereafter, we present how the linker dynamics is considered in dynamic FRET-lines.

### 2.2.1 Absence of dye-linker dynamics

The simplest case, where a dynamic FRET-line is applicable, occurs when a single molecule exchanges between two discrete conformational states. These limiting states are reported by two DA-distances corresponding to two discrete FRET-states. For such a case the FRET-induced donor decay can be written as:

$$\varepsilon(t) = x_a \cdot e^{-k_{RET}^{(1)} t} + (1 - x_a) \cdot e^{-k_{RET}^{(2)} t}. \quad (26)$$

Here,  $k_{RET}^{(1)}$  and  $k_{RET}^{(2)}$  are the FRET-rate constants of the state (1) and (2) and  $x_{(1)}$  is the fraction the molecule spend in the state (1).

Thus, for a single exponential donor in absence of FRET the donor fluorescence decay in presence of FRET, is described by two fluorescence lifetimes:  $\tau_{(1)} = 1/(k_{RET}^{(1)} + k_D)$  and  $\tau_{(2)} = 1/(k_{RET}^{(2)} + k_D)$ . Knowing these lifetimes and  $x_{(1)}$ , the fraction of species (1), the species-weighted and fluorescence-weighted average fluorescence lifetimes are given by:

$$\begin{aligned} \langle \tau_{D(A)} \rangle_x &= x_{(1)} \cdot \tau_{(1)} + (1 - x_{(1)}) \cdot \tau_{(2)} \\ \langle \tau_{D(A)} \rangle_F &= \frac{x_{(1)} \cdot \tau_{(1)}^2 + (1 - x_{(1)}) \cdot \tau_{(2)}^2}{x_{(1)} \cdot \tau_{(1)} + (1 - x_{(1)}) \cdot \tau_{(2)}} \end{aligned} \quad (27)$$

Using these two equations the species weighted average lifetime  $\langle \tau_{D(A)} \rangle_x$  can be expressed as a function of the fluorescence weighted average lifetime  $\langle \tau_{D(A)} \rangle_F$  (Kalinin et al., 2007) and the intensity based indicators ( $E$  and  $F_D/F_A$ ) can be expressed in dependence of the fluorescence weighted average lifetime:

$$\begin{aligned} E &= 1 - \frac{\langle \tau_{D(A)} \rangle_x}{\tau_{D(0)}} = 1 - \frac{1}{\tau_{D(0)}} \cdot \left[ \frac{\tau_{(1)} \cdot \tau_{(2)}}{\tau_{(1)} + \tau_{(2)} - \langle \tau_{D(A)} \rangle_F} \right] \\ \frac{F_D}{F_A} &= \frac{\Phi_F^{D0}}{\Phi_F^{A0}} \cdot \frac{\langle \tau_{D(A)} \rangle_x}{\tau_{D(0)} - \langle \tau_{D(A)} \rangle_x} = \frac{\Phi_F^{D0}}{\Phi_F^{A0}} \cdot \frac{1}{\tau_{D(0)}} \cdot \left[ \frac{\tau_{(1)} \cdot \tau_{(2)}}{(\tau_{(1)} + \tau_{(2)} - \langle \tau_{D(A)} \rangle_F) - \tau_{(1)} \cdot \tau_{(2)} / \tau_{D(0)}} \right] \end{aligned} \quad (28)$$

These two equations represent the dynamic FRET-line corrected for the quantum yield. Although, these equations are handy to use, the basic rationale of the dynamic FRET-line is obfuscated as the species fraction  $x_{(1)}$  does not appear explicitly; keep in mind that this dynamic FRET-line is based on the ‘‘thought experiment’’ of a varying relative species fraction. Thus, this set of equations is only valid if  $\langle \tau_{D(A)} \rangle_F$  is in the interval  $[\tau_{(1)}, \tau_{(2)}]$  and for the simplest case where states (1) and (2) are two static FRET-populations without linker dynamics.

### 2.2.2 Presence of dye-linker dynamics

To include the effect of linker dynamics, the DA-distance distribution  $p(R_{DA})$  caused by the dye linkers needs to be included for the description of each state. Again, we approximate  $p(R_{DA})$  by normal distributions. However, here it is considered for each state separately. Each distance distribution  $p(R_{DA})$  is defined by the mean DA-distance  $\bar{R}_{DA}$  and the corresponding width ( $\sigma_{RDA}$ ). Therefore, the FRET-induced decay can be written as:

$$\begin{aligned} \varepsilon_D(t, x_1) &= \int (x_1 \cdot p_1(R_{DA}) + (1 - x_1) \cdot p_2(R_{DA})) \cdot e^{-k_{RET}(R_{DA}) \cdot t} \cdot dR_{DA}, \\ p_n(R_{DA}) &= \frac{1}{\sqrt{2\pi} \cdot 2 \cdot \sigma_{RDA}} \exp\left( \frac{-(R_{DA} - \bar{R}_{DA,n})^2}{2 \cdot \sigma_{RDA}^2} \right), \end{aligned} \quad (29)$$

where  $n=1, 2$  are the limiting states. As can be expected, not analytical solution is possible for the species weighted ( $\langle\tau_{D(A)}\rangle_x$ ) and fluorescence weighted ( $\langle\tau_{D(A)}\rangle_F$ ) average fluorescence lifetimes. Thus, we solve these integrals by numerically to derive a curve on the  $E$  vs.  $\langle\tau_{D(A)}\rangle_F$  or  $F_D/F_A$  vs.  $\langle\tau_{D(A)}\rangle_F$  plane, which depends on the population fraction and not on  $\bar{R}_{DA}$  the average distance.

The simplest implementation consists off a similar approach as already presented for the static-FRET-lines, which includes the calculation of multiple species and fluorescence weighted average lifetimes and an empirical polynomial. This polynomial conversion function from the fluorescence weighted to the species weighted average lifetime can be directly applied to eq. (25). Of course the polynomial coefficients in the dynamic case differ from the coefficients found in the static FRET-lines.

### 2.2.3 Dynamic FRET-lines and relation to the species fractions

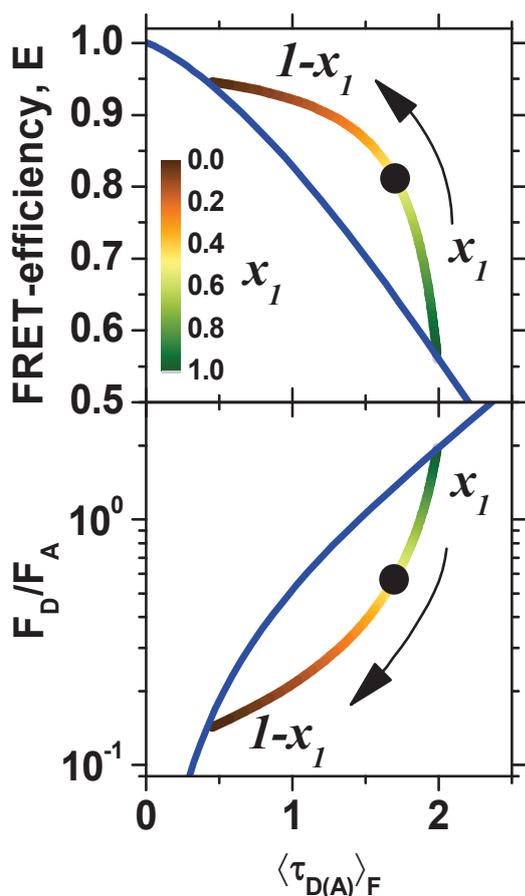
The dynamic FRET-line describes the exchange between two states and relates the FRET-efficiency the species average and the fluorescence average lifetimes. Thus, the location by which a single-molecule is represented in the  $E$  vs  $\langle\tau_{D(A)}\rangle_F$ -plane informs on the species fraction and in an experiment where two conformers dynamically interconvert, with a reaction rate constants faster than the integration time (several milliseconds), the maximum point of the population distribution reveals the equilibrium distribution. Therefore, by a dynamic FRET-line it is possible to obtain the equilibrium constant. For two interconverting species the population fraction is given by:

$$x_{(1)} = \frac{\tau_{(2)} \cdot (\langle\tau_{D(A)}\rangle_F - \tau_{(2)})}{(\tau_{(1)} - \tau_{(2)}) \cdot (\tau_{(1)} + \tau_{(2)} - \langle\tau_{D(A)}\rangle_F)} \quad (30)$$

where  $x_1$  is the fraction of the species with lifetime  $\tau_1$ . Therefore, knowing the end points of the reactions one can extract the fraction  $x_{(1)}$  and the equilibrium constant  $K$ . This last one is defined as:

$$K = \frac{x_1}{(1 - x_1)} \quad (31)$$

On Figure 4 the location of the solid circle shows the species fraction  $x_1$  starting from the longer lifetime and going in the direction of the shorter lifetime. The population fraction  $x_{(1)}$  ranges on values from 0 to 1, where 0 would mean the absence of population (1), and (1) is the absence of population (2). This fraction is illustrated in Figure 4 by the color scale.



**Figure 4.** Equilibrium constant in dynamic FRET-lines. Blue static linker corrected FRET-line ( $\sigma_{\text{RDA}}=6\text{\AA}$ ,  $\tau_{\text{D0}}=4.0$  ns), color-scaled linker corrected dynamic FRET-line with limiting states of fluorescence weighted average lifetimes of 0.45 ns and 2.0 ns. The color-scale gives the species fraction at the respective points on the dynamic FRET-line. The top shows the FRET-lines in the  $E$  vs.  $\langle \tau_{\text{D(A)}} \rangle_F$ . The bottom shows the FRET-lines of the intensity-ratio  $F_D/F_A$  vs.  $\langle \tau_{\text{D(A)}} \rangle_F$ . The black circle corresponds to species-fraction of  $x_1=0.5$  at which both limiting states are equally populated.

## 2.4 Multistate systems

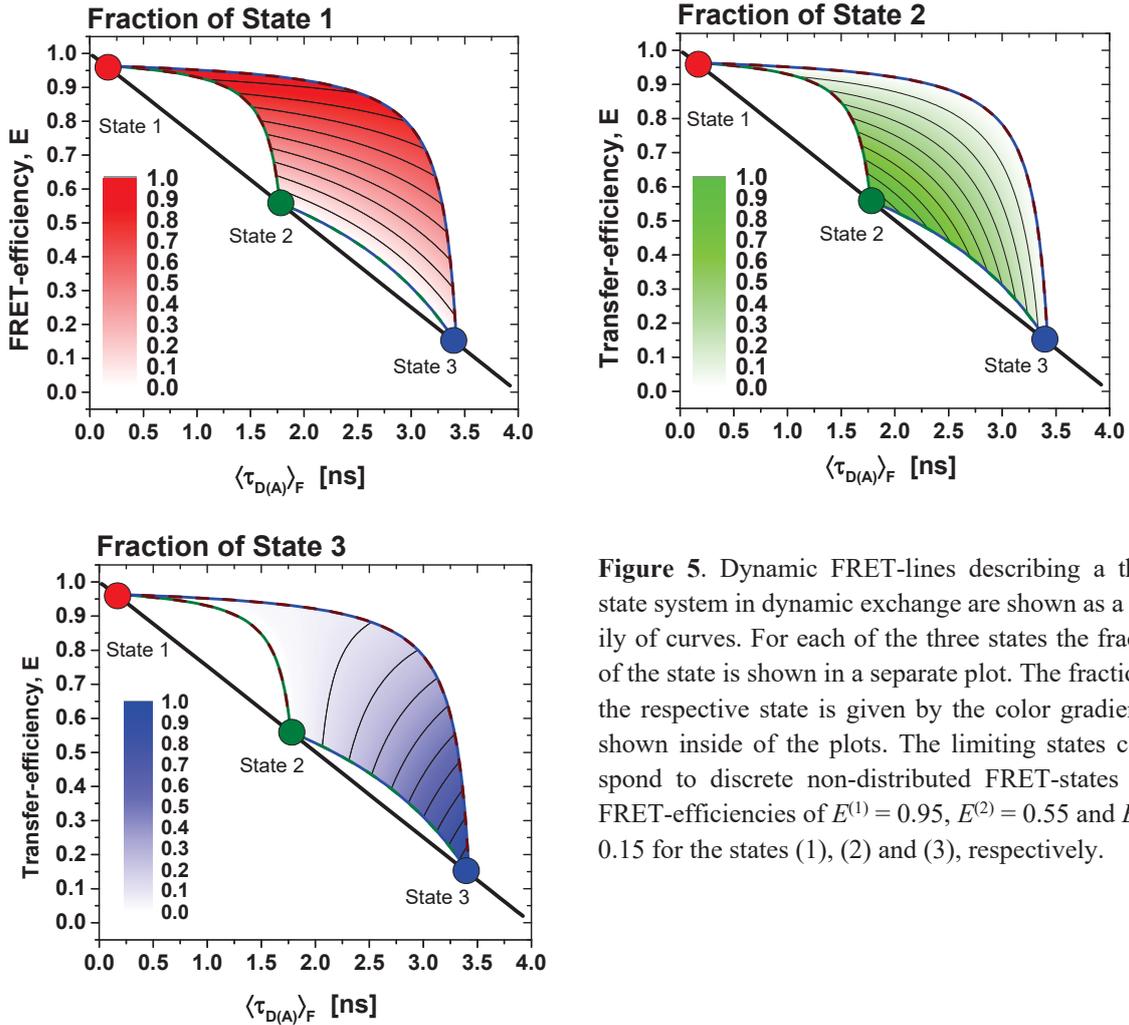
One can expand the principles presented on the previous sections to tackle multiple-state kinetics. Complex kinetic networks show particular signatures on the two dimensional histograms. For example, if the position of the population maximum does not follow the static FRET-line corrected for the linker dynamics or the dynamic FRET-line between two “known” states, then the only conclusion left is that the kinetics scheme is more complex than a two state system. Thus, the underlying distance distribution of the states does not reflect what is expected due to linker dynamics. It is then possible to build a connecting network for complex kinetics. Each exchange is determined by two end points. By connecting all possible ways in which the end points connect it is possible to have a visual inspection on the kinetic mechanisms. Moreover, if the existence of multiple states can be resolved by ensemble Time Correlated Single Photon Counting (eTCSPC) or by sub ensemble TCSPC (seTCSPC) in single-molecule experiments, those states can be used for building the connecting network. Typically, these will map the observed distribution in the MFD histograms.

To describe more complex reaction networks, we describe the FRET-induced donor decay of system composed of  $N$  limiting states, where the donor-acceptor distance distributions of each state ( $i$ ) is given by  $p^{(i)}(R_{DA})$ , by a superposition of all possible states:

$$\varepsilon_D(t) = \int_{R_{DA}=0}^{\infty} \left[ \sum_{i=1}^N x^{(i)} \cdot p^{(i)}(R_{DA}) \right] \cdot e^{-k_{RET}(R_{DA})t} \cdot dR_{DA} \quad \text{with} \quad \sum x^{(i)} = 1. \quad (32)$$

Herein  $x^{(i)}$  stands for the fraction of the state ( $i$ ). By combining the equation above with eq. 20 and eq. 21 the FRET-efficiency and the fluorescence weighted average lifetime  $\langle \tau_{D(A)} \rangle_F$  for an arbitrary mixture of the  $N$  states is obtained.

For such complicated systems obviously simple dynamic FRET-lines, which only consider two states as the ones described above, are inapplicable. To exemplify this, we consider a three-state system in dynamic exchange as the simplest example. To consider which family of curves may describe an experiment conducted on such system one has to recall the origin of the species fractions  $x^{(i)}$ . The idea of the dynamic FRET-line is to describe all molecules in dynamic exchange among distinct states. For a two-state system with the states (1) and (2) the molecule of interest is either in the first or the second state. Thus, during the observation time it may stay a fraction  $x^{(1)}$  in the state (1) and a fraction  $x^{(2)}$  in the state (2). As the fraction  $x^{(2)}$  is linearly dependent on  $x^{(1)}$  ( $x^{(2)} = 1 - x^{(1)}$ ) a line in the  $E$  vs  $\langle \tau_{D(A)} \rangle_F$ -plane is obtained. For a three state system the fractions  $x^{(1)}$ ,  $x^{(2)}$  and  $x^{(3)}$  may vary between zero and one and are only constrained by the normalization condition  $x^{(1)} + x^{(2)} + x^{(3)} = 1$ . Therefore, the system is not described by a single FRET-line but by a family of curves. For a three states system such family is shown in Fig. 5. This figure highlights, that the ‘‘simple’’ two-state FRET-lines are limiting cases for a three-state system. These limiting cases are for instance given by  $x^{(3)} = 0$  for a dynamic line between (1) and (2). Hence, the dynamic FRET-lines define the borders of the curve family. Similarly to the two state system where the position on the dynamic FRET-line defines the species fraction, the position on the  $E$  vs  $\langle \tau_{D(A)} \rangle_F$ -plane defines the species fraction for all states in a three state system as shown in Fig. 5 by the color scale.



**Figure 5.** Dynamic FRET-lines describing a three-state system in dynamic exchange are shown as a family of curves. For each of the three states the fraction of the state is shown in a separate plot. The fraction of the respective state is given by the color gradient as shown inside of the plots. The limiting states correspond to discrete non-distributed FRET-states with FRET-efficiencies of  $E^{(1)} = 0.95$ ,  $E^{(2)} = 0.55$  and  $E^{(3)} = 0.15$  for the states (1), (2) and (3), respectively.

Relating these fractions to dynamic properties of the system, namely the exchange rate constants of the states, for a quantitative analysis of the histograms is matter of ongoing research.

## 2.3 Alternative implementation of FRET-lines

### 2.3.1 Polynomial Approximation

The integrals defined by eq. (21) are difficult to solve analytically. Therefore, the simplest approach to treat these integrals is to create a table which relates the fluorescence averaged lifetimes  $\langle \tau_{D(A)} \rangle_F$  and the respective species averaged lifetimes  $\langle \tau_{D(A)} \rangle_x$ . Using such table an empirical function converting  $\langle \tau_{D(A)} \rangle_F$  to  $\langle \tau_{D(A)} \rangle_x$  can be generated. For two-state systems, it is always possible to generate such a conversion function, as the fluorescence averaged lifetime  $\langle \tau_{D(A)} \rangle_F$  and the species averaged lifetime  $\langle \tau_{D(A)} \rangle_x$  are monotonically dependent on the average inter-dye distance  $\bar{R}_{DA}$  and on the species fraction  $x_I$  for the dynamic FRET-lines. Typically, a third order polynomial with coefficients  $A^{(i)}$  is sufficient to relate  $\langle \tau_{D(A)} \rangle_F$  to  $\langle \tau_{D(A)} \rangle_x$ . As mentioned before, several examples of the  $A^{(i)}$  coefficients for various  $\sigma_{R_{DA}}$ 's are shown in the appendix (Table A1). Although these coefficients apply for the case of the static FRET-lines, a similar approach

can be used to treat dynamic FRET-lines. For static FRET-lines the polynomial coefficients depend on the lifetime distribution of the donor in absence of the acceptor and the dye linker length while for dynamic FRET-lines, the polynomial coefficients depend on the lifetime distribution of the donor in absence of the acceptor, the dye linker length *and* the limiting states. Thus, if the limiting states are changed the dynamic FRET-lines have to be recalculated by numerical integration and new set of polynomial coefficients has to be determined limiting the applicability of the polynomial approach.

### 2.3.2 Quadratic approximation

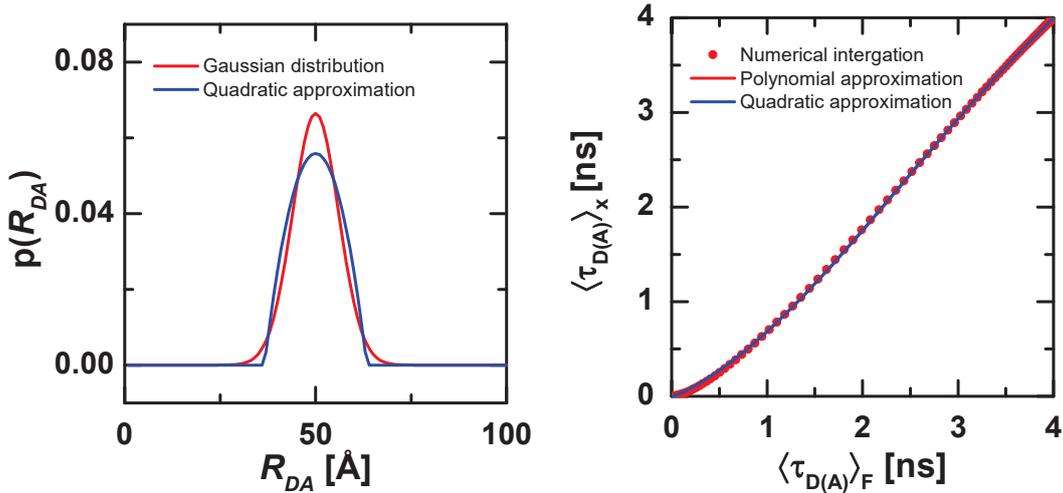
As alternative to the numerical approach, we present an analytical solution of the parametric relation between the species weighted ( $\langle \tau_{D(A)} \rangle_x$ ) and fluorescence weighted ( $\langle \tau_{D(A)} \rangle_F$ ) average fluorescence lifetimes. In order to solve the integral in eq. 21 we describe the DA-distance distribution  $p(R_{DA})$  by an inverse quadratic function, the first order approximation of a normal distribution and express the normalized probability distance distribution  $p(R_{DA})$  with width  $\omega_{DA}$  by:

$$p_Q(R_{DA}) = \begin{cases} \frac{3}{4\sqrt{2} \cdot \omega_{DA}} - \frac{3 \cdot (R_{DA} - \bar{R}_{DA})^2}{8 \cdot \sqrt{2} \cdot \omega_{DA}^3}, & \bar{R}_{DA} - \sqrt{2} \cdot \omega_{DA} < R_{DA} < \bar{R}_{DA} + \sqrt{2} \cdot \omega_{DA} \\ 0 & \text{otherwise} \end{cases} \quad (33)$$

The width  $\omega_{DA}$  directly relates to the standard deviation ( $\sigma_{R_{DA}}$ ) of the Gaussian distance distribution by:

$$\sigma_{R_{DA}}^2 = \frac{2}{5} \omega_{DA}^2 \quad (34)$$

The analytical solution of the species weighted  $\langle \tau_{D(A)} \rangle_x$  and the fluorescence weighted  $\langle \tau_{D(A)} \rangle_F$  average lifetimes for this distribution are presented in the appendix (eqs. (39-42)). In Fig. 6 the distance distribution given by eq. (32) resulting static FRET-line are compared to the numerical integration of eq. (21) of a normal DA-distance distribution.



**Figure 6.** Polynomial approximation of linker corrected static FRET-lines. Comparison of the normal and the inverse quadratic donor-acceptor distance distributions  $p(R_{DA})$  ( $\sigma_{R_{DA}} = 6 \text{ \AA}$ , the quadratic distribution with  $\omega_{DA} = 9.49 \text{ \AA}$ ). Parametric plot  $\langle \tau_{D(A)} \rangle_x$  and  $\langle \tau_{D(A)} \rangle_F$  for comparing various solutions. The overlay of the polynomial approximation for Gaussian with  $\sigma_{DA} = 6 \text{ \AA}$ , and analytical solution of the quadratic distribution with  $\omega_{DA} = 9.49$  shows that they are identical.

Similarly, the dynamic FRET-lines can be solved analytically. Here, each state is represented by a quadratic equation with mean distance  $\bar{R}_{DA,1}$  or  $\bar{R}_{DA,2}$ , for states 1 and 2, respectively. In dynamic exchange, the probability of the mixed state has the form of:

$$p(R_{DA}, x_1) = x_1 \cdot p_1(\omega_{DA}, \bar{R}_{DA,1}) + (1 - x_1) \cdot p_2(\omega_{DA}, \bar{R}_{DA,2}) \quad (35)$$

After integration using the previously determined integrals (eq. (21)), an analytical expression of the species weighted averaged lifetime in dependence of the fluorescence weighted averaged lifetime obtained for the interval defined  $\tau_1 \leq \langle \tau_{D(A)} \rangle_F \leq \tau_2$ , whereas  $\tau_1$  and  $\tau_2$  are the fluorescence weighted averaged lifetimes of the limiting states and  $\tau_1 < \tau_2$ , shows the good agreement between the two models.

### 2.3.3 Alternative representation of FRET lines

Traditionally, experimental data is displayed in terms of FRET-efficiencies and the fluorescence average lifetimes. However, by rearranging eq. (21) and (20) one finds an alternative description by the mean ( $Erw$ ) and the variance ( $Var$ ) of the donor fluorescence lifetime. The mean and the variance are determined by the FRET-efficiency and the fluorescence weighted average lifetime of the donor. Thus, for any point on the  $E$  vs  $\langle \tau_{D(A)} \rangle_F$  plane, the mean lifetime ( $Erw$ ) and the variance ( $Var$ ) of the lifetime can be estimated by:

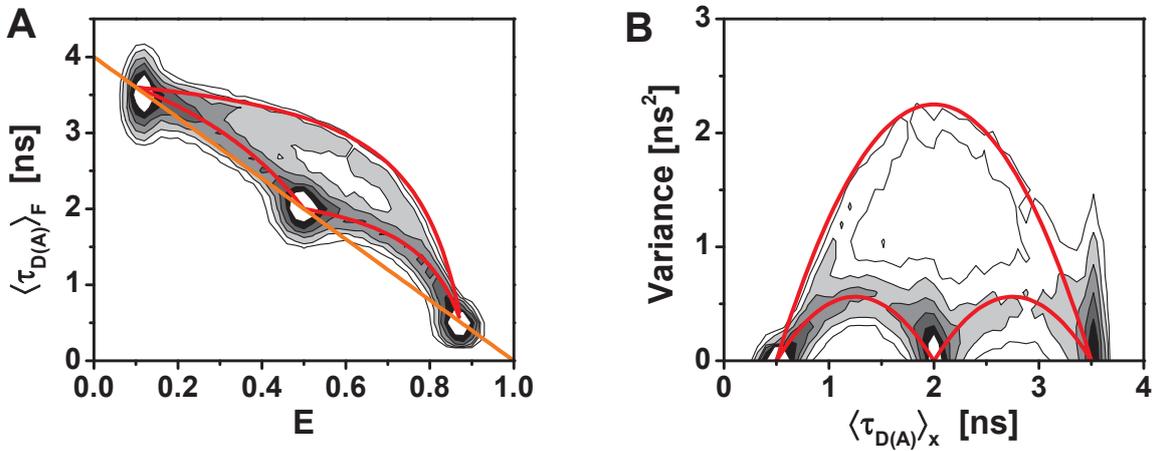
$$Erw(\tau_{D(A)}) = \langle \tau_{D(A)} \rangle_x = (1 - E) \cdot \langle \tau_{D(0)} \rangle_x \quad (36)$$

$$Var(\tau_{D(A)}) = \langle \tau_{D(A)} \rangle_F \cdot \langle \tau_{D(A)} \rangle_x - \left( \langle \tau_{D(A)} \rangle_x \right)^2$$

By this representation the first two moments of the lifetime distribution of the experimental data are presented. This transformation of the experimental data simplifies the analysis and interpretation as in absence of conformational dynamics the variance vanishes. For comparison a mean-variance histogram is shown to the corresponding fluorescence lifetime/FRET-efficiency histogram in Fig. 7 for a 3-state system. In a mean-variance histogram a static FRET-line is a horizontal line centered at zero. In presence of conformational dynamics between two states with the fluorescence lifetimes  $\tau_1$  and  $\tau_2$  the dynamic FRET-line is a semicircle centered around on the average of the two states:

$$\left( \frac{\tau_1 - \tau_2}{2} \right)^2 = Var(\tau_{DA}) + \left( \langle \tau_{DA} \rangle_x - \frac{\tau_1 + \tau_2}{2} \right)^2 \quad (37)$$

Such mean-variance histograms are used to describe of single molecules in patch-clamp experiments and reveal limiting states and exchange dynamics by quantitative analysis (Patlak, 1993). Thus, a similar analysis may be applied to analyze multi-dimensional FRET-histograms quantitatively. For comparison a linear three state system in dynamic exchange is displayed in Fig. 6 by a traditional  $E$  vs  $\langle \tau_{D(A)} \rangle_F$  histogram and a mean-variance histogram.



**Figure 7.** Comparison of (A) fluorescence weighted average lifetime vs. transfer-efficiency histogram and (B) mean-variance vs. lifetime histogram of analyzed data from a simulated single-molecule experiment of a three state system in kinetic exchange. For each single-molecule event the mean and the variance of the fluorescence lifetime were calculated by eq. (35). In (A) and (B) The lifetimes of the limiting states are given by 0.5 ns, 2 ns and 3.5 ns. All states exchange with a rate of  $k=0.05 \text{ ms}^{-1}$ , the donor fluorophore has a single lifetime  $\tau_{D(0)}=4 \text{ ns}$ . In (A) the static-FRET-line was calculated by eq. (22) and the dynamic FRET-lines is given eq. (25). In (B) the dynamic FRET-lines are calculated by eq. (36).

Mean-variance histograms derived using the fluorescence lifetime have to be distinguished from histograms determined by burst-variance analysis (BVA), as in BVA relies on the count-rate to detect kinetics. Thus, in BVA fast exchange may be overlooked, if the exchange kinetics

is faster than the inter-photon time. In mean-variance histograms relying in the fluorescence lifetime this not the case.

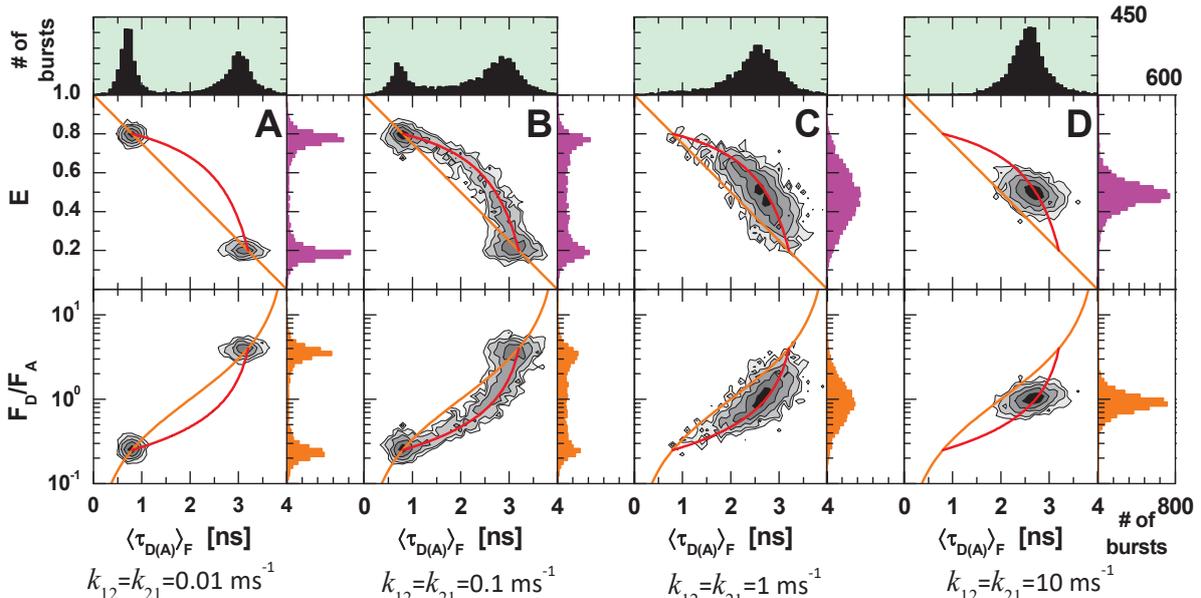
### 2.3.4 FRET Lines and timescale of dynamics

Even though dynamics and time-scales of dynamics have not been discussed so far, FRET-lines are very helpful to understand the time-scale of dynamic inter-conversions. In case of large exchange rate constants compared to the observation time (milliseconds), the state-populations in the single-molecule burst are approaching to the equilibrium state-populations. Here, the position of the measured distribution on 2D plot reflects the equilibrium of the states as mentioned in section 2.2.3. In this case, only a single peak is observed in the FRET-histograms. Keep in mind that, due to different averaging of the fluorescence and species weighted average lifetimes, dynamics results in a shift away from the static FRET-line described by eq. (22).

Contrary to fast processes, slow processes result in a broadening of the observed histograms whereas in the limiting case of an absent dynamics: separate states are described by the static FRET-line (Kalinin et al., 2007, Kalinin et al., 2010a, Gopich and Szabo, 2012, Kubo, 1969). Thus, the shape of MFD-histograms in the  $E$  vs.  $\langle \tau_{D(A)} \rangle_F$  plane give an indication of the time-scale of the dynamics. The more confined it is the faster is the exchange. Slower dynamics, in the range of the observation time, results in broadening of the distributions. We illustrate this by simulations considering a two state system with the states  $S_1$  and  $S_2$  in dynamic exchange:



Here,  $k_{12}$  and  $k_{21}$  are exchange rate constant among the two states ( $k_{12} = k_{21}$ ). For single-molecule simulations, these exchange rate constants were varied in a range from  $0.01 \text{ ms}^{-1}$  to  $10 \text{ ms}^{-1}$ . In Fig. 8 we present the results of the simulations. Note that if  $k_{12} = k_{21} < 1/t_{diff}$  the two subpopulations are clearly identified. That means, that the exchange is so slow that most of the time a molecule is observed it stays in either one or the other state. With increased exchange rate constant, the observed molecules change their state more often while being observed. Thus, instead of two limiting states an elongated mixing peak between them is observed. At higher exchange rate constants most molecules exchange states during the observation time. Thus, only a single peak is detected. Regardless of the magnitude of the rate constants, the distribution of the population lies exactly on top of the dynamic FRET-line.



**Figure 8.** Monte Carlo simulations of smFRET data for a two state system exchanging at various rates with no linker dynamics. (A-D) Two dimensional histograms of  $E$  vs.  $\langle \tau_{D(A)} \rangle_F$  and  $F_D/F_A$  vs.  $\langle \tau_{D(A)} \rangle_F$  from slow ( $k_{12}=k_{21} = 0.01 \text{ ms}^{-1}$ ) to fast dynamic exchange ( $k_{12}=k_{21} = 10 \text{ ms}^{-1}$ ) and slow interconversion, shape and location of the population are indicative of dynamic timescale. Dynamic FRET-line is independent of the dynamics timescale. The parameters for the simulations were: fluorescence quantum yield of acceptor and donor  $\Phi_A = \Phi_{FD(0)} = 1$ . The detection efficiency ratio  $g_G/g_R = 1$ ,  $\langle B_G \rangle = 1.5 \text{ kHz}$ ,  $\langle B_R \rangle = 0.7 \text{ kHz}$  and crosstalk  $\alpha = 0$ .  $E_1 = 0.8$ ,  $E_2 = 0.2$ . For simplicity no linker dynamics was simulated, but correction for the dye linker motion has been presented before. The diffusion time in all simulations was set to  $t_{diff} = 1.5 \text{ ms}$ .

With the identification of dynamic timescales, it is clear that the dynamic FRET-lines represent the likelihood of identifying exchange processes between two limiting states at the sub millisecond timescale, while the “static” FRET lines include the fast motion of the dyes around the labeling site. This motion occurs at sub microsecond timescales. In other word, static and dynamic FRET lines connect the faster dynamics and the slower dynamics.

### 3 Real case and complex FRET-lines

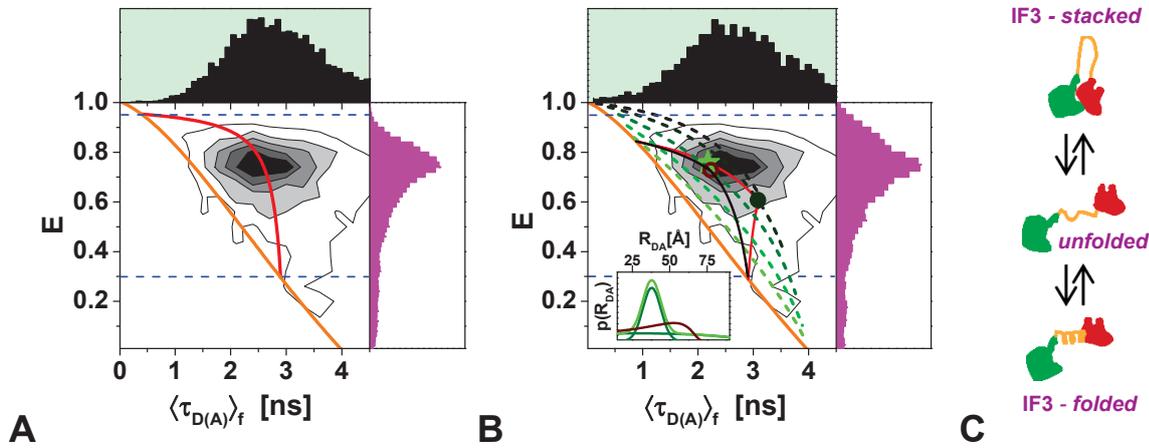
In summary, we presented a simple flow to determine complex biomolecular dynamics by experimental smFRET data. To describe complex real cases, first, it is required to identify the limiting states. As reflected by eq. (21) the underlying DA-distance distribution  $p(R_{DA})$  cannot be recovered solely by a multidimensional MFD-diagram as information regarding  $p(R_{DA})$  (reflected by  $\varepsilon_{<}(t)$ ) is partly lost due to the integration. This can be partly overcome by sub-ensemble TCSPC (seTCSPC) or by combining sm-experiments with molecular simulations. seTCSPC resolves the populations that are sampled at the nanosecond timescale. The width of the  $p(R_{DA})$  of the states can be determined by analyzing of the fluorescence decays by a proper model to obtaining the average distance  $\bar{R}_{DA}$ , and standard deviation  $\sigma_{R_{DA}}$ . Alternatively  $p(R_{DA})$  is determined by combining accessible volume simulations of the fluorophores with an atomistic model of the molecule in question (Sindbert et al., 2011). Once the distribution of the FRET states (

$p(R_{DA})$ ) are obtained, the next step is to resolve the dynamic network or obtain all possible reaction rate constants. This is doable for a two state system and in some simplified cases (non-cyclic) for a 3 state system. Extracting the reaction rate constants for higher order dynamics requires additional tools such as Photon Distribution Analysis (PDA) or filtered Fluorescence Correlation Spectroscopy (fFCS) (Sindbert et al., 2011). With fFCS the relaxation times or eigenvalues of the reaction rate matrix are an experimental observable. Graphical inspection of the 2D histograms would give hints on the complexity of the network and more complex schemes can be solved with help of simulations.

FRET-lines can be applied to test models that help interpret experimental results. We present this case step by step for single-molecule FRET experiments of a two domain translation initiation factor (IF3) in solution without binding partner. The C-terminal domain (CTD) of IF3 binds to the shoulder of the Ribosome's 30S subunit. Contrary, the function and location of the N-terminal domain (NTD) in translation is not known (Kycia et al., 1995, Moreau et al., 1997, Petrelli et al., 2001, Dallas and Noller, 2001, Fabbretti et al., 2007, Julian et al., 2011, McCutcheon et al., 1999). It is important to bear in mind that the structures of the individual domains are known; however, the structure of the full-length is unknown (Biou et al., 1995, Garcia et al., 1995a, Garcia et al., 1995b). The CTD and the NTD are connected by a flexible linker (Moreau et al., 1997), which allows for different arrangements of the NTD and CTD with distance ranging from 45 Å to 75 Å between the centers of mass of the two domains. NMR studies indicated that the two domains of IF3 from *E. coli* move largely independently adding to the observed flexibility. One of the open questions is whether the linker region is unstructured or structured. In thermophile organisms the linker region adopts a helical conformation at room temperature (PDB: 1TIF(Biou et al., 1995)). Given this prior knowledge a two-state or a three-state system might describe the data. In a two state system the two domains are either associated or separated by a structured linker. In a three state system the connecting linkage region might adopt a helical structure or is unstructured.

These models can be tested based on smFRET/MFD experiments (Figure 9). IF3 was labeled with donor and acceptor located at the NTD and CTD (Materials and Methods) to monitor the dynamics of the flexible linker. Using the concept of FRET-lines we are able to identify that multiple states are populated. To reach that conclusion we applied static and dynamic FRET-lines. This last one considers the mixing of the states mentioned above. For simplicity the states describing the compact state (associated domains) and the extended state (helical linking region)

are described by normal distributions. The anticipated state with an unstructured linkage region is described by a worm-like chain distribution.



**Figure 9.** IF3 without binding partner in solution. The transfer-efficiencies of the single-molecule bursts were obtained by correcting the measured green and red signals ( $F_D$  and  $F_A$ ) for background, spectral crosstalk ( $\alpha = 0.01$ ) and ratio of green and red detection efficiencies ( $g_G/g_R = 0.77$ ). The static dye-linker corrected FRET-lines are displayed in orange ( $\sigma_{R_{DA}} = 6 \text{ \AA}$ ). A) 2-state dynamic FRET-line with limiting states determined by geometrical modeling of the folded ( $\bar{R}_{DA}=61 \text{ \AA}$ ) and stacked conformation ( $\bar{R}_{DA}=40 \text{ \AA}$ ), B) the green shaded family of curves represents different realizations of a worm-like chain model with a length of 50 to 200  $\text{\AA}$  at different relative persistence length (0.12, 0.23, 0.34, 0.45, 0.55), the green circle represents the chosen realization of the WLC-model (length 127  $\text{\AA}$ , persistence length 0.23). Given the chosen realization of the WLC-model the dynamic FRET-lines are shown as red-lines to the anticipated folded and the stacked conformations. Given this possible description of the data, the species fractions and the resulting distance distribution (blue curve) at the center of mass can be calculated (green star). An alternative 3-state dynamic FRET-line is shown in dark red. The corresponding distance distributions are shown as inset. C) The schematic model of IF3 and the kinetic model.

To determine the limiting states and their distance distribution we used a geometrical simulation based on steric repulsion of the two domains. Here the helical linker region was either kept helical or unstructured. If the linker region was kept helical, the distance of approximately 60  $\text{\AA}$  is expected. In the case where the linker region was modeled as unstructured the average distance is only 40  $\text{\AA}$ . This distance corresponds to the case where the CTD and the NTD are in contact. In Figure (A) it is clearly visible that the experimental data cannot be described by the two-state model, because the red line misses the distribution observed in the MFD histogram. This is the first indication that more than two states are required to describe the experimental observations. The next step was to consider a three-state system. These states might correspond to: *i*) association of the CTD and the NTD, *ii*) a folded middle-domain and *iii*) an unfolded middle domain. In this unfolded case the distance between the CTD and the NTD can be described by a radial distribution function using a worm-like chain model (Becker et al., 2010). The radial distribution function of a WLC is determined by two parameters: the persistence length of the WLC and the maximum length of the WLC. The maximum length can be estimated

based on the geometry of the molecule while the mean persistence length cannot be easily determined. However, solely based on the FRET-histogram reasonable parameters can be determined using the already known end-points (the stacked conformations and the folded conformation).

Although an additional state is necessary to describe the molecule, the parameters describing this third state, namely the maximum end-to-end distance of the molecule and the mean persistence length, cannot be fully deduced solely based on the 2D-histogram. To test this hypothesis, the transfer-efficiency and the fluorescence weighted average lifetime is calculated for a given set of relative persistence lengths (0 to 0.99) and maximum length (50 to 200 Å). A set of possible distributions of the fluorescence weighted averaged lifetimes and transfer-efficiencies are obtained. Each combination of the persistence length and the maximum-length corresponds to one point in the  $E$  vs.  $\langle \tau_{D(A)} \rangle_F$  plane. These points are illustrated as dashed green lines for given fixed persistence length and varying maximum-length. Given these points in the  $E$  vs.  $\langle \tau_{D(A)} \rangle_F$  plane, we selected one point of reasonable maximum end-to-end length (127 Å) and relative persistence length (0.34) (Green circle in Fig. 9B). From this point we calculated the dynamic FRET-lines leaving from and going to this state from the two other states (the stacked and the folded state). As illustrated by Fig. 9B this yields a better description of the experimental data. Although discrimination among models is based on visual inspection, this approach is sufficient to show that IF3 can be described as a polymer chain like model.

## 4 Conclusions

We have presented several guidelines to properly read MFD histograms using both the static and dynamic FRET-lines. Besides the obvious identification of single subpopulations that are static, 2D histograms are rich in information about the dynamics of FRET labeled biomolecules. Here we summarize few of the graphical interpretations of the 2D histograms. Donor only population is shown at high  $F_D/F_A$  ratio with lifetime  $\sim 4$  ns (donor-only for Alexa 488) and  $E = 0$ . High FRET appears at shorter lifetimes when the fluorescence of acceptor is high ( $F_D/F_A \rightarrow 0$ ,  $E \rightarrow 1$ ). Static FRET states distributions have circular shapes in  $F_D/F_A$  vs.  $\langle \tau_{D(A)} \rangle_F$  (typical semi-logarithmic scale), and their distribution widths are given by shot noise and acceptor photophysics (Kalinin et al., 2010a). Static FRET states follow a theoretical line (static FRET-line) that can account for dye linker mobility and as such they can serve to estimate the width of the distribution of states given by linker dynamics. We also provided an analytical solution of the FRET-lines that demonstrates its robustness and rigorous mathematical treatment. We verified

that the previous empirical description, using a polynomial approximation, is in agreement with the numerical solution of our parametric equations. In all cases, one can obtain the width of the distribution only if the sample is known to be static and not in dynamic exchange between multiple conformations. For a molecule that exchanges conformations at timescales faster than the diffusion time we presented the dynamic FRET-line. Dynamic populations alter the shape of the histograms and its shape depends on the timescale of the dynamics. Moreover, we presented the correction for the dynamic FRET-line which includes the mobility of the linkers. Moreover, FRET-lines can be treated to consider corrections for most of experimental “effects” (e.g. dye linker dynamics, changes in detection efficiencies, variations in quantum yield of fluorophores, and the multi-exponential properties of dyes) and as such they are also very reliable. Finally, we also show how complex kinetic mechanisms contribute and complex distributions can be solved with the help of various FRET-lines.

## 5 Materials and Methods

### 5.1 Multiparameter Fluorescence Detection

MFD for confocal single molecule Förster Resonance Energy Transfer (smFRET) measurements were done using a linearly polarized, active-mode-locked Argon-ion laser (Innova Saber, Coherent, Santa Clara, CA, USA, 496.5 nm, 73.5 MHz,  $\sim 300$  ps) or by a 485 nm diode laser (LDH-D-C 485, Picoquant, Berlin, Germany) operating at 64 MHz. The laser light is focused into the dilute solution ( $< 50$  pM) of labeled molecules by a 60x/1.2 water immersion objective. Each molecule generates a brief burst of fluorescence photons as it traverses the detection volume. This photon-train is divided initially into its parallel and perpendicular components via a polarizing beamsplitter and then into a wavelength ranges below and above 595 nm by using a dichroic beamsplitter (595 DCXR, AHF, Tübingen, Germany). Additionally, red (HQ 720/150 nm for Cy5) and green (HQ 533/46 nm for Alexa 488 and Rh110) bandpass filters (both made by AHF, Tübingen, Germany) in front of the detectors ensure that only fluorescence photons coming from the acceptor and donor molecules are registered. An estimate of the focal geometry is acquired by determining the diffusion correlation time of  $200 \pm 13$   $\mu\text{s}$  for Rhodamine 110 and knowing its diffusion coefficient of  $0.34 \pm 0.03$   $\mu\text{m}^2/\text{ms}$ . Detection is performed using four avalanche photodiodes (SPCM-AQR-14, Laser Components, Germany or alternatively for the green channels PDM050CTC, or  $\tau$ -SPAD-100, both PicoQuant, Berlin, Germany). The signals from all detectors are guided through a passive delay unit and two routers to two synchronized time-correlated single photon counting boards (SPC 132 or SPC 832, Becker and Hickl, Berlin, Germany) connected to a PC. Bursts of fluorescence photons are distinguished from the background of 1-2 kHz by applying certain threshold intensity criteria (Eggeling et al., 2001). Bursts

during which bleaching of the acceptor occurs are excluded from further analysis by applying a criterion regarding the difference in macroscopic times,  $|T_G - T_R| < 0.5$  ms, where  $T_G$  and  $T_R$  are the average macroscopic times in which all photons have been detected in the green and red channels respectively during one burst (Eggeling et al., 2006).

## 5.2 Single molecule simulator (Brownian Dynamics)

Simulations of single-molecule measurements were done via Brownian dynamics. 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

$$\begin{aligned} P(t_i) &= k_i^{-1} \exp(-k_i t_i) \\ P(t_j) &= k_j^{-1} \exp(-k_j t_j) \end{aligned} \quad 39$$

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

## 5.3 WLC Polymer model

A semiflexible macromolecules can generically be described by the worm-like chain models. No closed analytical solution to the radial distribution function of the worm-like chain model is available. Hence, we used an approximation as presented in (Becker et al., 2010):

$$p(R) = \frac{8 \cdot N_W}{\left(1 - \left(\frac{R}{L}\right)^2\right)^{\frac{3}{2}} \cdot \left(2 - \left(\frac{R}{L}\right)^2\right)^3} \cdot \left(\frac{3}{4\pi \cdot \kappa}\right)^{\frac{3}{2}} \cdot e^{-\frac{3\left(\frac{R}{L}\right)^2}{4\kappa \left(1 - \left(\frac{R}{L}\right)^2\right)}}$$

where  $N_W$  is a numerically determined normalization constant,  $\kappa = l_p / L$  - the chain stiffness and  $L$  is the chain length. The distribution function is defined within the range from 0 to  $L$ . Considering the end-to-end distance distribution as donor to acceptor distances distribution in unfolded biomolecules,  $p(R_{DA})$ , dynamic FRET lines could be generated for the cases when macromolecule is in either folded or unfolded state.

## 5.4 IF3 production and labeling

Double-labeled IF3 mutants were prepared by using a native Cys at position 65 in the NTD and introducing a Cys at position E166 at the CTD. The position 65 is less solvent exposed than position E165C allowing for site-specific labeling. The double cysteine mutant of IF3 (100  $\mu$ M)

was first labeled for 30 min at room temperature (RT) with a 20-fold molar excess of Alexa 647-maleimide (Life Technologies, Inc.) in labeling buffer (50 mM Tris-HCl pH 7.1, 200 mM NH<sub>4</sub>Cl, 5% glycerol). The labeling reaction was stopped by addition of 2-mercaptoethanol (6 mM) and purified from the excess of dye using SP FF column (GE healthcare) and step gradient elution with 1 M NH<sub>4</sub>Cl. The eluted protein was preferentially labeled at Cys166 as compared to the labeling of wild-type IF3 under the same conditions (< 5%). To overcome the low reactivity of Cys65, Alexa647-labeled IF3 was dialyzed against labeling buffer containing 2 M urea and subsequently reacted with Alexa 488-maleimide (Life Technologies, Inc) for 2 hours at RT. The reaction was stopped by addition of 2-mercaptoethanol (6 mM) and the protein was purified from the excess of the dye as described above. Spectrometric characterization of the double-labeled IF3 indicated the presence of one molecule of each dye covalently attached to IF3. All complexes were prepared in buffer A (50 mM Tris-HCl, pH 7.5, 70 mM NH<sub>4</sub>Cl, 30 mM KCl, 7 mM MgCl<sub>2</sub>) at room temperature. For MFD measurements IF3 was diluted to a picomolar concentration in buffer A. This assures a final concentration of ~ 0.01 molecules in the observed volume at any given time, which allowed us to analyze individual bursts of each diffusing molecule.

### **Acknowledgement**

We thank Dmytro Rodnin for providing the experimental data of IF3. HS acknowledges support from the Alexander von Humboldt foundation and Clemson University start-up funds. TP thanks the International Helmholtz Research School of Biophysics and Soft Matter (IHRS BioSoft).

## 6 Appendix

### A: Polynomial approximation coefficients

**Table A1.** Polynomial coefficients for the conversion of the species averaged lifetimes to the fluorescence averaged lifetimes of a single exponential donor dye with a lifetime of 4 ns

$\sigma_{R_{DA}}$ [Å]	$A^{(5)}$	$A^{(4)}$	$A^{(3)}$	$A^{(2)}$	$A^{(1)}$
2	0.0015	-0.0173	0.0656	0.9164	-0.0018
4	0.0046	-0.0552	0.2217	0.7011	-0.0048
6	0.0058	-0.0818	0.3742	0.4404	-0.0048
8	0.0025	-0.0711	0.4398	0.2202	-0.0010
10	-0.0058	-0.0165	0.3866	0.0884	0.0045
12	-0.0179	0.0749	0.2211	0.0597	0.0079
14	-0.0326	0.1955	-0.0463	0.1444	0.0015
16	-0.0493	0.3414	-0.4159	0.3667	-0.0321
18	-0.0681	0.5154	-0.9098	0.7819	-0.1290

### B: Quadratic approximation (Analytical Solution)

#### B.1 Species averaged lifetime

$$\langle \tau_{D(A)} \rangle_x = -\frac{\tau_{D(0)}}{32 \cdot \sqrt{2} \cdot \omega_{DA}^3} \cdot H(R_{DA}) \Bigg|_{\bar{R}_{DA} - \sqrt{2}\omega_{DA}}^{\bar{R}_{DA} + \sqrt{2}\omega_{DA}} \quad (40)$$

$$H(R_{DA}) = A + B + C + D + E + F + G$$

where  $A, B, C, D, E, F$  and  $G$  are defined as

$$\begin{aligned} A &= -4R_{DA} \cdot (R_{DA}^2 - 3R_{DA} \cdot \bar{R}_{DA} + 3R^2) \\ B &= -4R_0 \cdot (R_0^2 - \bar{R}_{DA}^2 + 2\omega_{DA}^2) \cdot \tan^{-1}\left(\frac{R_{DA}}{R_0}\right) \\ C &= -2R_0 \cdot (2R_0 \cdot (R_0 - \sqrt{3} \cdot \bar{R}_{DA}) + R^2) \cdot \tan^{-1}\left(\sqrt{3} - \frac{2R_{DA}}{R_0}\right) \\ D &= 2 \cdot R_0 \cdot (2R_0 \cdot (R_0 + \sqrt{3} \cdot \bar{R}_{DA}) + R^2) \cdot \tan^{-1}\left(\sqrt{3} + \frac{2R_{DA}}{R_0}\right) \\ E &= -4R_0^2 \cdot \bar{R}_{DA} \cdot \log(R'^2) \\ F &= R_0 \cdot (2R_0 \cdot \bar{R}_{DA} - \sqrt{3}R^2) \cdot \log(-\sqrt{3}R_0 \cdot \bar{R}_{DA} + R'^2) \\ G &= R_0 \cdot (2R_0 \cdot \bar{R}_{DA} + \sqrt{3}R_{DA}^2) \cdot \log(\sqrt{3}R_0 \cdot \bar{R}_{DA} + R'^2) \end{aligned} \quad (41)$$

$$R'^2 = \bar{R}_{DA}^2 - 2\omega_{DA}^2$$

$$R''^2 = R_{DA}^2 + R_0^2$$

The symbol  $\left| \frac{\bar{R}_{DA} + \sqrt{2}\omega_{DA}}{\bar{R}_{DA} - \sqrt{2}\omega_{DA}} \right.$  means that the  $H(R_{DA})$  function is evaluated at these limits.

### B.2 Fluorescence averaged lifetime

$$\langle \tau_{D(A)} \rangle_F = - \frac{1}{\langle \tau_{D(A)} \rangle_x} \cdot \frac{\tau_{D(0)}^2}{192 \cdot \sqrt{2} \cdot \omega_{DA}^3} \cdot H'(R_{DA}) \Bigg|_{\bar{R}_{DA} - \sqrt{2}\omega_{DA}}^{\bar{R}_{DA} + \sqrt{2}\omega_{DA}} \quad (42)$$

$$H'(R_{DA}) = A' + B' + C' + D' + E' + F' + G'$$

where  $A', B', C', D', E', F'$  and  $G'$  are defined as

$$A' = 24R_{DA}^3 - 72R_{DA}^2 \cdot \bar{R}_{DA} + 72R_{DA} \cdot R'^2 + \frac{12R_0^6 \cdot R_{DA} \cdot (R_{DA}^2 - 2R_{DA} \cdot \bar{R}_{DA} + R'^2)}{R_0^6 + R_{DA}^6}$$

$$B' = 4R_0 \cdot (9R_0^2 - 7\bar{R}_{DA}^2 + 14\omega_{DA}^2) \cdot \tan^{-1}\left(\frac{R_{DA}}{R_0}\right)$$

$$C' = 2R_0 \cdot (2R_0 \cdot (9R_0 - 8\sqrt{3} \cdot \bar{R}_{DA}) + 7R'^2) \cdot \tan^{-1}\left(\sqrt{3} - \frac{2R_{DA}}{R_0}\right)$$

$$D' = -2R_0 \cdot (2R_0 \cdot (9R_0 + 8\sqrt{3} \cdot \bar{R}_{DA}) + 7R'^2) \cdot \tan^{-1}\left(\sqrt{3} + \frac{2R_{DA}}{R_0}\right)$$

$$E' = 32R_0^2 \cdot \bar{R}_{DA} \cdot \log(R'^2)$$

$$F' = R_0 \cdot (-16R_0 \cdot \bar{R}_{DA} + 7\sqrt{3}R'^2) \cdot \log(-\sqrt{3}R_0 \cdot R_{DA} + R'^2)$$

$$G' = R_0 \cdot (16R_0 \cdot \bar{R}_{DA} + 7\sqrt{3}R'^2) \cdot \log(\sqrt{3}R_0 \cdot R_{DA} + R'^2)$$

$$R'^2 = \bar{R}_{DA}^2 - 2\omega_{DA}^2$$

$$R''^2 = R_{DA}^2 + R_0^2$$

The symbol  $\left| \frac{\bar{R}_{DA} + \sqrt{2}\omega_{DA}}{\bar{R}_{DA} - \sqrt{2}\omega_{DA}} \right.$  means that the  $H'(R_{DA})$  function is evaluated at these limits.

## References

- ANDRECKA, J., TREUTLEIN, B., ARCUSA, M. A. I., MUSCHIELOK, A., LEWIS, R., CHEUNG, A. C. M., CRAMER, P. & MICHAELIS, J. 2009. Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex. *Nucleic Acids Research*, 37, 5803-5809.
- ANTONIK, M., FELEKYAN, S., GAIDUK, A. & SEIDEL, C. A. 2006. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *Journal of Physical Chemistry B*, 110, 6970-8.
- BECKER, N. B., ROSA, A. & EVERAERS, R. 2010. The radial distribution function of worm-like chains. *European Physical Journal E*, 32, 53-69.
- BIOU, V., SHU, F. & RAMAKRISHNAN, V. 1995. X-ray crystallography shows that translational initiation factor IF3 consists of two compact alpha/beta domains linked by an alpha-helix. *EMBO J*, 14, 4056-64.
- CHEN, Y., MÜLLER, J. D., SO, P. T. C. & GRATTON, E. 1999. The photon counting histogram in fluorescence fluctuation spectroscopy. *Biophysical Journal*, 77, 553-567.
- DALLAS, A. & NOLLER, H. F. 2001. Interaction of translation initiation factor 3 with the 30S ribosomal subunit. *Mol Cell*, 8, 855-64.
- EGGELING, C., BERGER, S., BRAND, L., FRIES, J. R., SCHAFFER, J., VOLKMER, A. & SEIDEL, C. A. M. 2001. Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *Journal of Biotechnology*, 86, 163-180.
- EGGELING, C., FRIES, J. R., BRAND, L., GÜNTHER, R. & SEIDEL, C. A. M. 1998. Monitoring conformational dynamics of a single molecule by selective fluorescence spectroscopy. *Proc.Natl.Acad.Sci.USA.*, 95, 1556-1561.
- EGGELING, C., WIDENGREN, J., BRAND, L., SCHAFFER, J., FELEKYAN, S. & SEIDEL, C. A. M. 2006. Analysis of Photobleaching in Single-Molecule Multicolor Excitation and Förster Resonance Energy Transfer Measurements. *Journal of Physical Chemistry A*, 110, 2979-2995.
- FABBRETTI, A., PON, C. L., HENNELLY, S. P., HILL, W. E., LODMELL, J. S. & GUALERZI, C. O. 2007. The real-time path of translation factor IF3 onto and off the ribosome. *Mol Cell*, 25, 285-96.
- FELEKYAN, S., KALININ, S., SANABRIA, H., VALERI, A. & SEIDEL, C. A. M. 2012. Filtered FCS: species auto- and cross-correlation functions highlight binding and dynamics in biomolecules. *ChemPhysChem*, 13, 1036-1053.
- FELEKYAN, S., SANABRIA, H., KALININ, S., KUHNEMUTH, R. & SEIDEL, C. A. 2013. Analyzing Förster resonance energy transfer with fluctuation algorithms. *Methods in Enzymology*, 519, 39-85.
- FÖRSTER, T. 1948. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik*, 437, 55-75.
- FRIES, J. R., BRAND, L., EGGELING, C., KÖLLNER, M. & SEIDEL, C. A. M. 1998. Quantitative identification of different single-molecules by selective time-resolved confocal fluorescence spectroscopy. *Journal of Physical Chemistry A*, 102, 6601-6613.
- GARCIA, C., FORTIER, P. L., BLANQUET, S., LALLEMAND, J. Y. & DARDEL, F. 1995a. <sup>1</sup>H and <sup>15</sup>N resonance assignments and structure of the N-terminal domain of Escherichia coli initiation factor 3. *Eur J Biochem*, 228, 395-402.
- GARCIA, C., FORTIER, P. L., BLANQUET, S., LALLEMAND, J. Y. & DARDEL, F. 1995b. Solution structure of the ribosome-binding domain of E. coli translation initiation factor IF3. Homology with the U1A protein of the eukaryotic spliceosome. *J Mol Biol*, 254, 247-59.
- GOPICH, I. & SZABO, A. 2005. Theory of photon statistics in single-molecule Förster resonance energy transfer. *Journal of Chemical Physics*, 122, 14707.

- GOPICH, I. V. & SZABO, A. 2003. Single-macromolecule fluorescence resonance energy transfer and free-energy profiles. *Journal of Physical Chemistry B*, 107, 5058-5063.
- GOPICH, I. V. & SZABO, A. 2007. Single-molecule FRET with diffusion and conformational dynamics. *Journal of Physical Chemistry B*, 111, 12925-12932.
- GOPICH, I. V. & 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, 7747-52.
- HA, T., LAURENCE, T. A., CHEMLA, D. S. & WEISS, S. 1999. Polarization Spectroscopy of Single Fluorescent Molecules. *Journal of Physical Chemistry B*, 103, 6839-6850.
- HAAS, E., WILCHEK, M., KATCHALSKI-KATZIR, E. & STEINBERG, I. Z. 1975. Distribution of end-to-end distances of oligopeptides in solution as estimated by energy transfer. *Proceedings of the National Academy of Sciences of the United States of America*, 72, 1807-1811.
- JULIAN, P., MILON, P., AGIRREZABALA, X., LASSO, G., GIL, D., RODNINA, M. V. & VALLE, M. 2011. The cryo-EM Structure of a complete 30S translation initiation complex from *Escherichia coli*. *PLoS Biol.*, 9, e1001095.
- KALININ, S., FELEKYAN, S., ANTONIK, M. & SEIDEL, C. A. M. 2007. Probability distribution analysis of single-molecule fluorescence anisotropy and resonance energy transfer. *Journal of Physical Chemistry B*, 111, 10253-10262.
- KALININ, S., FULLE, S., KÜHNEMUTH, R., SINDBERT, S., FELEKYAN, S., GOHLKE, H. & SEIDEL, C. A. M. 2015. Diffusion with traps: experiment, simulation, and theory to describe fluorophore dynamics in biomolecular FRET.
- KALININ, S., SISAMAKIS, E., MAGENNIS, S. W., FELEKYAN, S. & SEIDEL, C. A. M. 2010a. On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits. *Journal of Physical Chemistry B*, 114, 6197-6206.
- KALININ, S., VALERI, A., ANTONIK, M., FELEKYAN, S. & SEIDEL, C. A. M. 2010b. Detection of structural dynamics by FRET: A photon distribution and fluorescence lifetime analysis of systems with multiple states. *J. Phys. Chem. B*, 114, 7983-7995.
- KALININ, S. P., T.; SINDBERT, S.; ROTHWELL, P. J.; BERGER, S.; RESTLE, T.; GOODY, R. S.; GOHLKE, H.; SEIDEL, C. A. M. 2012. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nature Methods*.
- KASK, P., PALO, K., ULLMANN, D. & GALL, K. 1999. Fluorescence-intensity distribution analysis and its application in biomolecular detection technology. *Proc.Natl.Acad.Sci.USA.*, 96, 13756-13761.
- KUBO, R. 1969. A Stochastic Theory of Line Shape. *Advances in Chemical Physics*. John Wiley & Sons, Inc.
- KÜHNEMUTH, R. & SEIDEL, C. A. M. 2001. Principles of single molecule multiparameter fluorescence spectroscopy. *Single Molecules*, 2, 251-254.
- KYCIA, J. H., BIOU, V., SHU, F., GERCHMAN, S. E., GRAZIANO, V. & RAMAKRISHNAN, V. 1995. Prokaryotic translation initiation factor IF3 is an elongated protein consisting of two crystallizable domains. *Biochemistry*, 34, 6183-7.
- MAUS, M., COTLET, M., HOFKENS, J., GENSCHE, T., DE SCHRYVER, F. C., SCHAFFER, J. & SEIDEL, C. A. M. 2001. An experimental comparison of the maximum likelihood estimation and nonlinear least squares fluorescence lifetime analysis of single molecules. *Analytical Chemistry*, 73, 2078-2086.
- MCCUTCHEON, J. P., AGRAWAL, R. K., PHILIPS, S. M., GRASSUCCI, R. A., GERCHMAN, S. E., CLEMONS, W. M., JR., RAMAKRISHNAN, V. & FRANK, J. 1999. Location of translational initiation factor IF3 on the small ribosomal subunit. *Proc Natl Acad Sci U S A*, 96, 4301-6.
- MOREAU, M., DE COCK, E., FORTIER, P. L., GARCIA, C., ALBARET, C., BLANQUET, S., LALLEMAND, J. Y. & DARDEL, F. 1997. Heteronuclear NMR studies of *E. coli*

- translation initiation factor IF3. Evidence that the inter-domain region is disordered in solution. *J Mol Biol*, 266, 15-22.
- MUSCHIELOK, 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.
- NIR, E., MICHALET, X., HAMADANI, K. M., LAURENCE, T. A., NEUHAUSER, D., KOVCHEGOV, Y. & WEISS, S. 2006. Shot-noise limited single-molecule FRET histograms: Comparison between theory and experiments. *Journal of Physical Chemistry B*, 110, 22103-22124.
- PATLAK, J. B. 1993. Measuring Kinetics of Complex Single-Ion Channel Data Using Mean-Variance Histograms. *Biophysical Journal*, 65, 29-42.
- PETRELLI, D., LATEANA, A., GAROFALO, C., SPURIO, R., PON, C. L. & GUALERZI, C. O. 2001. Translation initiation factor IF3: two domains, five functions, one mechanism? *EMBO J*, 20, 4560-9.
- ROTHWELL, P. J., ALLEN, W. J., SISAMAKIS, E., KALININ, S., FELEKYAN, S., WIDENGREN, J., WAKSMAN, G. & SEIDEL, C. A. 2013. dNTP-dependent Conformational Transitions in the Fingers Subdomain of KlenTaq1 DNA Polymerase: INSIGHTS INTO THE ROLE OF THE "NUCLEOTIDE-BINDING" STATE. *Journal of Biological Chemistry*, 288, 13575-91.
- ROTHWELL, P. J., BERGER, S., KENSCH, O., FELEKYAN, S., ANTONIK, M., WÖHRL, B. M., RESTLE, T., GOODY, R. S. & SEIDEL, C. A. M. 2003. Multi-parameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 Reverse Transcriptase:primer/template complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 1655-1660.
- SCHAFFER, J., VOLKMER, A., EGGELING, C., SUBRAMANIAM, V., STRIKER, G. & SEIDEL, C. A. M. 1999. Identification of single molecules in aqueous solution by time-resolved fluorescence anisotropy. *Journal of Physical Chemistry A*, 103, 331-336.
- SINDBERT, S., KALININ, S., HIEN, N., KIENZLER, A., CLIMA, L., BANNWARTH, W., APPEL, B., MULLER, S. & SEIDEL, C. A. M. 2011. Accurate Distance Determination of Nucleic Acids via Forster Resonance Energy Transfer: Implications of Dye Linker Length and Rigidity. *Journal of the American Chemical Society*, 133, 2463-2480.
- SISAMAKIS, E., VALERI, A., KALININ, S., ROTHWELL, P. J. & SEIDEL, C. A. M. 2010. Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods in Enzymology*, 475, 456-514.
- TAMARAT, P., MAALI, A., LOUNIS, B. & ORRIT, M. 2000. Ten Years of Single-Molecule Spectroscopy. *Journal of Physical Chemistry*, 104, 1-16.
- TELLINGHUISEN, J., GOODWIN, P. M., AMBROSE, W. P., MARTIN, J. C. & KELLER, R. A. 1994. Analysis of Fluorescence Lifetime Data for Single Rhodamine Molecules in Flowing Sample Streams. *Analytical Chemistry*, 66, 64-72.
- WEISS, S. 1999. Fluorescence spectroscopy of single biomolecules. *Science*, 283, 1676-83.
- WEISS, S. 2000. Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nature Structural Biology*, 7, 724-9.
- WIDENGREN, J., KUDRYAVTSEV, V., ANTONIK, M., BERGER, S., GERKEN, M. & SEIDEL, C. A. 2006. Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. *Analytical Chemistry*, 78, 2039-50.
- ZANDER, C. & DREXHAGE, K. H. 1996. Sequential Counting of Single Molecules in a Capillary. *Journal of Fluorescence*, 7, 37S-39S.

## Chapter D - Quantitative FRET-analysis of measurements in live cells

---

### 1 Introduction

Multi-parameter fluorescence detection (MFD) records the spectral detection window and the polarization of all registered fluorescent photons with picosecond time resolution for a selected spectral excitation window. Hence, MFD is capable of monitoring at least 8 dimensions of fluorescence simultaneously: fluorescence intensities, lifetimes, anisotropy, quantum yields, donor-acceptor distances, stoichiometries, the excitation and the emission window (Förster, 1951, Kuhnemuth and Seidel, 2001, Widengren et al., 2006). This wealth of information recorded in a single experiment provides unsurpassed flexibility in data-analysis. Additional information can be utilized by combining MFD with microscopy as in multi-parameter fluorescence image spectroscopy (MFIS) for qualitative and quantitative improvements fluorescence microscopy. Nowadays, technical improvements allow for highly reliable instrumentation and the determination of average quantifiers of FRET and anisotropy in living cells (Bleckmann et al., 2010, Somssich et al., 2015). However, average FRET quantifiers do not reflect the resolution provided by MFIS-measurements, and tremendous methodological improvements in fluorescence spectroscopy driven by single-molecule MFD are currently overlooked in image spectroscopy. In single-molecule fluorescence measurements, all dimensions (anisotropy, fluorescence lifetimes and FRET) are recorded, explored and connected to structural models on a routine basis (Kalinin et al., 2012). This is currently not the case in MFIS. Hence, by applying next generation of analysis methods developed for single-molecule experiments in combination with molecular simulations to MFIS measurements, MFIS has a remarkable potential to approach the final frontier of molecular biophysics: understanding the movement and mechanism of molecular machines in living cells. This and given the current progress in super-resolution microscopy (Hell and Wichmann, 1994, Dickson et al., 1997, Betzig et al., 2006, Rust et al., 2006), as a future perspective, MFIS in living cells may break the current resolution down to the molecular level.

Current MFIS-analysis cannot achieve this goal because: i) not all recorded information is fully processed, and ii) methodological synergies with molecular simulations are not utilized. In that respect, MFIS cannot hold up to the best practices set by current single-molecule analysis. Thus, even though MFIS oversees cellular, biochemical, structural and conformational aspects of physiologically relevant proteins in living cells, standard MFIS-analysis limits the conclusions that can be drawn.

To gain a deeper understanding of protein function and to unleash the full potential of MFIS, next generation of analysis as outlined below, is vital. To understand the biological function of protein assembly, simulations of different complexity, ranging from atomistic molecular models over coarse grained simulations to simple chemical equilibrium chemistry, should be combined to derive fluorescence observables that are compared to the experimental data. This may allow for a full understanding of biological processes in living cells. The combination of molecular simulations and spectroscopy entitles MFIS to oversee the architecture of biomolecules from Angström to micrometers while single-molecule sensitivity allows one to track gated reaction chemistry linked to biological responses. Thus, when properly analyzed, MFIS may answer exciting biochemical and biophysical questions such as the determination of equilibrium constants or conformational changes within the protein upon a ligand binding in living cells.

To obtain answers by MFIS measurements, precisely phrased questions have to be asked in terms of a fluorescence model: structural properties have to be expressed as anisotropies, lifetimes and intensities, chemical equilibria of multiple components have to be considered, and protein function via binding partners have to be identified to finally determine equilibrium constants of the respective binding partners. Most of these problems have been previously addressed in single-molecule studies which solve structural models by multiple FRET-measurements (Kalinin et al., 2012). The key was to consider the conformational space of the dyes which allows relating structural models to fluorescence properties. In single-molecule measurements, this problem was solved by accessible volume (AV) simulations which sample the conformational space of the fluorophores. The AV simulations were established for small organic molecules (see chapter A) and cannot be directly applied to fluorescent proteins (FPs) used in MFIS due to the longer linkers and the different size of the fluorescent proteins. Therefore, alternative coarse grained simulation techniques were developed and applied. Due to the large size of the fluorescent proteins, their rotational correlation time is significantly longer than that of small organic dyes. In MFIS measurements, this apparent disadvantage can be utilized to measure distances by the fluorescence depolarization due to homo-FRET between two identical fluorophores. Due to higher concentrations of FPs, the presence of multiple acceptors has to be described quantitatively. Hence, these differences between fluorescent proteins and organic dyes demand for adapted analysis models and new types of simulations.

These goals were pursued using model systems of increasing complexity. First, a set of GFP-mCherry fusion proteins in living-cells was analyzed and compared to simulations to establish

a simulation scheme for FPs coupled to proteins. Using these simulations, time-resolved fluorescence decays and other FRET-observables such as the FRET-induced donor decay and steady-state transfer-efficiency were predicted. This allowed to go beyond formal descriptions of time-resolved fluorescence decays recorded by MFIS. By described model simulations the complex interplay between different protein species equilibrium constants in living cells were determined. The developed models and analysis methods were applied in the following manuscripts:

- 1) Revealing structural features and affinities of protein complexes in living cells by MFIS-FRET analysis
- 2) Guanylate binding proteins (GBPs) directly attack *T. gondii* via supramolecular complexes
- 3) MFIS-FRET analysis reveals the structural homo-oligomerization pattern of the G-protein coupled receptor TGR5 in live cells

Personal contributions to these manuscripts are outlined below.

Fluorescent proteins (FPs) are mainly used in MFIS, and they differ from small organic fluorophores used in single-molecule measurements: First, due to their size the rotational correlation times of FPs are significantly longer, therefore, the dynamic  $\kappa^2$ -approximation cannot be applied. Second, FPs are considerably larger, and thus more pronounced steric restrictions are anticipated. Third, the linkers coupling FPs to the host molecules are at least 5 times longer than the ones used in single-molecule studies. Thus, a correct averaging over the sterically allowed space is crucial. This is not the case in small organic fluorophore where a uniform weighting of the sterically allowed space produced satisfactory results (see supplement A). The three aspects above were considered in Monte-Carlo (MC) simulations of the sterically allowed conformational space of the FPs. The MC simulations explicitly consider orientation effects and the entropic effects of the dye-linker. In FPs and small organic fluorophores, specific interactions between the fluorophore and the host molecule may introduce errors in the order of the linker-length. In contrast to fast rotating organic fluorophores, the slow fluorescence depolarization of FPs by rotational motion prohibits monitoring specific interactions. On the other side this allows one to relate the depolarization degree to the FRET rate and thus to measure the distances.

## 2 GFPs and Structure

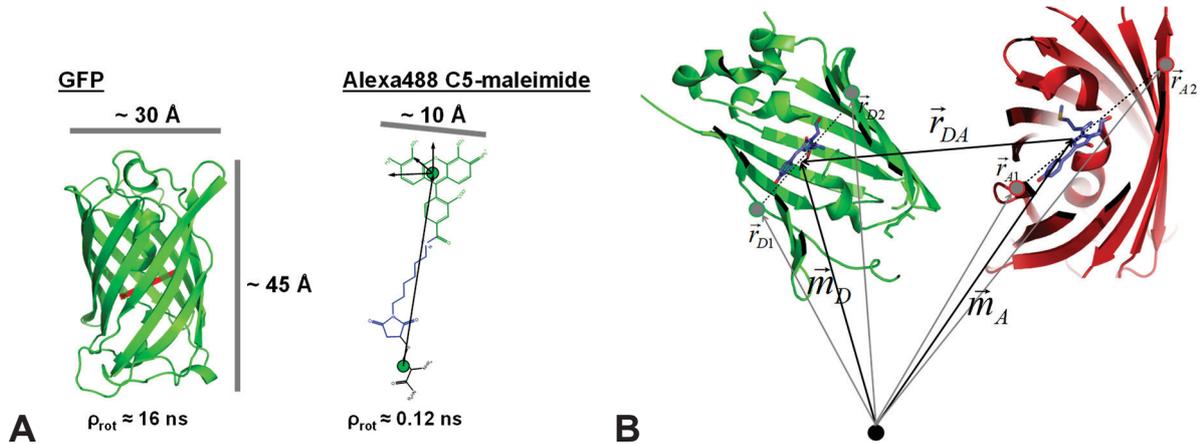
### 2.1 Dye-dimensions and orientation effects

The rate constant of energy-transfer from a donor fluorophore (D) to an acceptor fluorophore (A) via dipolar coupling is given by the FRET-rate constant. It depends on the distance  $R_{DA}$

between the fluorophores and the mutual orientation of their transition dipole moments captured by the orientation factor  $\kappa^2$ :

$$k_{RET} = \frac{1}{\tau_0} \frac{3}{2} \kappa^2 \left( \frac{R_0}{R_{DA}} \right)^6 \quad (1)$$

Herein  $R_0$  is the Förster-radius, a constant for a given fluorophore pair which depends on the spectral overlap of the donor emission spectrum and the acceptor absorption spectrum, and  $\tau_0$  is the lifetime of the donor fluorophore ( $\sim 2.6$  ns in case of GFP). The Förster-radius  $R_0$  of the dye pairs is usually in the range of 10-80 Angströms ( $\text{\AA}$ ). For the most commonly used FRET-pair GFP-mCherry the Förster-radius  $R_0$  is 52  $\text{\AA}$ .



**Figure 1.** Structure of fluorescent proteins (FPs) and nomenclature used for the calculation of the orientation factor and the FRET-rate constant **(A)** Dimensions and structure of GFP in comparison to Alexa488 C5-maleimide. The chromophore of GFP is shown in red. The rotational correlation time of the GFP and free Alexa488 are given below **(B)** Nomenclature of vectors and coordinates to describe FRET between two FPs. The vector  $\vec{r}_{DA}$  connects the chromophores of both FPs. The vectors  $\vec{m}_D$  and  $\vec{m}_A$  point to the center of the donor- and acceptor-chromophore, respectively.  $\hat{\mu}_D$  and  $\hat{\mu}_A$  are the unit-vectors of the transition dipole moments of the donor and the acceptor, calculated using the vectors,  $\vec{r}_{D1}$  and  $\vec{r}_{D2}$  and  $\vec{r}_{A1}$ ,  $\vec{r}_{A2}$  which point to the  $C\alpha$ -atoms on the beta-barrel.

If the orientation factor  $\kappa^2$  is approximated by the isotropic average  $2/3$ , an apparent distance  $R_{DA,app}$  is obtained by

$$R_{DA,app} = R_0 \cdot (k_{RET} \cdot \tau_0)^{-1/6} \quad (2)$$

The FRET-rate constant can be determined if the fluorescence intensity decay of the donor in absence of FRET is known. Assumptions regarding the orientation factor have to be made to determine distances by the FRET-rate constant. In the case of small organic dyes (Kalinin et al., 2015),  $R_{DA,app}$  can be interpreted as real donor-acceptor distance (chapter B). The rotational correlation time of free FPs ( $\sim 16$  ns) (Striker et al., 1999) is approximately 100 times longer

than that of rhodamines or cyanines. Hence, the true donor-acceptor distance  $R_{DA}$  cannot be directly approximated by the apparent donor-acceptor distance  $R_{DA,app}$ , and the orientation factor of each donor-acceptor pair has to be considered to obtain the correct distance  $R_{DA}$  from the FRET-rate constant.

The orientation factor can be calculated if the spatial arrangement of the FPs along the distance is known. This is shown in Fig. 1B where the relevant vectors are depicted. Given those vectors the orientation factor  $\kappa^2$  and the distance between the fluorophores are calculated as it is outlined in Table 1. Now, if the orientation factor  $\kappa^2$ , the lifetime of the fluorophore and the Förster-radius are known, then the distance between the dipoles  $R_{DA}$  can be calculated.

**Table 1.** Calculation of D-A distance and orientation factors.

	Donor (GFP)	Acceptor (mCherry)
<b>Coordinates of the two chosen C<math>\alpha</math>-atoms</b>	$\vec{r}_{D1} \vec{r}_{D2}$	$\vec{r}_{A1} \vec{r}_{A2}$
<b>Distance between the two C<math>\alpha</math>-atoms</b>	$R_{D,21} = \ \vec{r}_{D2} - \vec{r}_{D1}\ _2$	$R_{A,21} = \ \vec{r}_{A2} - \vec{r}_{A1}\ _2$
<b>Unit vector connecting the two C<math>\alpha</math>-atoms</b>	$\hat{\mu}_D = \frac{\vec{r}_{D2} - \vec{r}_{D1}}{R_{D,21}}$	$\hat{\mu}_A = \frac{\vec{r}_{A2} - \vec{r}_{A1}}{R_{A,21}}$
<b>Coordinates of the middle point of the connecting vector</b>	$\vec{m}_D = \vec{r}_{D1} + \frac{R_{D,21}}{2} \hat{\mu}_D$	$\vec{m}_A = \vec{r}_{A1} + \frac{R_{A,21}}{2} \hat{\mu}_A$
<b>Calculation of D-A distance and orientation factor</b>	$\vec{r}_{DA} = \vec{m}_D - \vec{m}_A$ $R_{DA} = \ \vec{m}_D - \vec{m}_A\ _2$ $\hat{\mu}_{DA} = \frac{\vec{r}_{DA}}{R_{DA}}$ $\kappa^2 = (\langle \hat{\mu}_A, \hat{\mu}_D \rangle - 3 \cdot \langle \hat{\mu}_D, \hat{\mu}_{DA} \rangle \cdot \langle \hat{\mu}_A, \hat{\mu}_{DA} \rangle)^2$	

The definition of the parameters and variables are given in Fig. 1

This interplay between the distance and the orientation factor is often used as a thought terminating cliché against accurate distance measurements by FRET. However, if it is considered that for a given distance  $R_{DA}$  not only a single  $\kappa^2$  applies but a whole distribution  $p(\kappa^2)$  has to be considered, then a distribution of FRET-rate constants and thus a distribution of apparent distances will be obtained. In the simplest case,  $p(\kappa^2)$  distribution is given by isotropic oriented dipoles which results in a the probability distribution given by (van der Meer, 2002):

$$p(\kappa^2) = \begin{cases} \frac{1}{2\sqrt{3\kappa^2}} \ln(2 + \sqrt{3}) & 0 \leq \kappa^2 \leq 1 \\ \frac{1}{2\sqrt{3\kappa^2}} \ln\left(\frac{2 + \sqrt{3}}{\sqrt{\kappa^2} + \sqrt{\kappa^2 - 1}}\right) & 1 \leq \kappa^2 \leq 4 \end{cases} \quad (3)$$

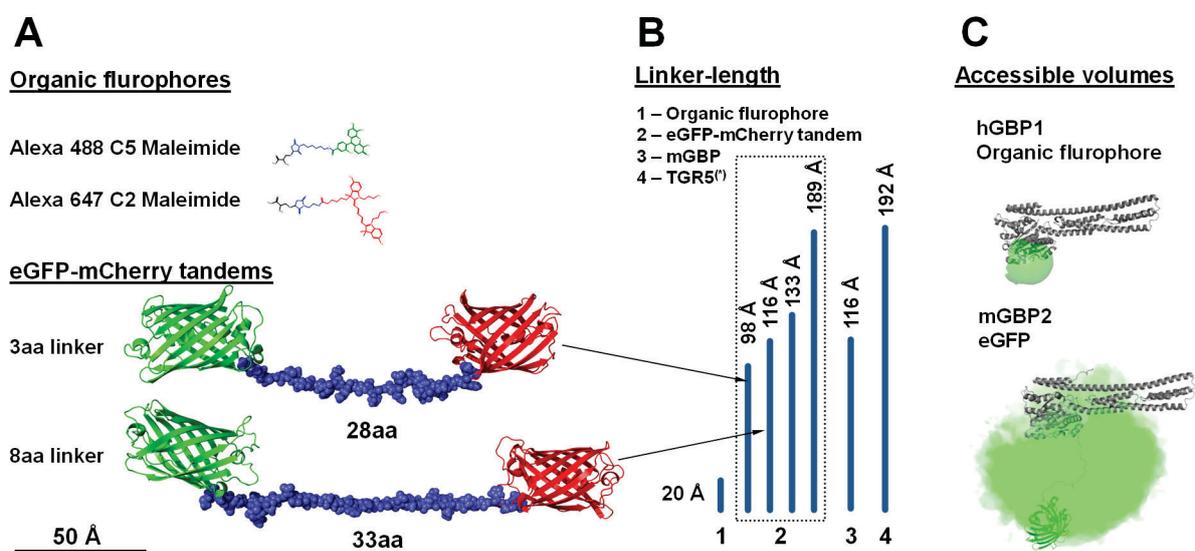
For every distance  $R_{DA}$ , there will be a distribution of orientation factors which results in a distribution of FRET-rate constants. Here, the distance distribution  $R_{DA,app}$  contains the distribution of orientation factors  $p(\kappa^2)$  and information on the distance distribution  $p(R_{DA})$ . In turn, this implies that if the assumptions regarding the distribution of orientation factors can be made, then the distance distribution can be approximated. Surely, correlations between the separation distance and the orientation factor cannot be captured this way. Therefore, orientation and distance distributions are best obtained by molecular simulations and compared to experimental results (van der Meer et al., 2014).

## 2.2 Weighting of distribution

In addition to the orientation effects, the spatial distribution of the fluorophores has to be considered to compare MFIS experiments to structural models. For small organic fluorophores, we previously used accessible volume calculations to describe the conformational space of the dyes (Sindbert et al., 2011). Here, even though it was suggested to use an empirical weighting functions to improve agreement with experiments (Hofig et al., 2014), a uniform weighting of the sterically allowed conformational space provided sufficient accuracy for structural modelling (Kalinin et al., 2012). In small organic dyes transient effects due to diffusion are only of minor importance (Kalinin et al., 2015). As FPs are even larger than the organic dyes, certainly diffusion does not have to be considered, and it can be assumed that the inter fluorophore distance distribution  $p(R_{DA})$  is static during the fluorescence lifetime.

Linkers used to connect the FPs to the host proteins are 5 to 10 times longer compared to the linkers used to tether organic fluorophores to the proteins of interest. This is illustrated in Fig. 2 where the structures of the Alexa488 C5 maleimide and Alexa647N C2 maleimide (typical single molecule fluorophores) are compared to the structure of FPs. Clearly, the end-to-end distance distribution of FP-tandems is not uniform: fully-extended conformations are far less likely than folded conformations. This has to be considered in the analysis of the MFIS experiments. Previously, simple worm-like chain models have been applied to describe the average transfer-efficiency in FP-tandems connected by flexible peptide linkers (Evers et al., 2006), and computational tools have been used to model FPs distance distribution (Pham et al., 2007). But so

far time-resolved fluorescence measurements as MFIS have not been combined with molecular models in living cells.



**Figure 2.** Comparison of linker-length used in different studies – the small organic fluorophores used in the studies targeting the elucidation of structures have a linker length of  $\sim 20$  Å. GFP-mCherry tandems have flexible connection linkers with 28, 33, 38, 54 amino acids, in the mGBP studies the fluorescent proteins (FPs) were coupled by flexible linkers with 33 amino acids, in the studies addressing the G-protein-coupled receptor (GPCR) TGR5 the FPs were coupled by a linker with 55 amino acids. (A) Structural comparison of organic fluorophores typically used in single-molecule studies and fluorescence proteins connected by a 3 amino acid linker and an eight amino acid linker. Due to the flexible N-terminal and C-terminal loops of GFP and mCherry a 3 amino acid spacer between GFP and mCherry results in 28 flexible amino acids. The organic fluorophores and the FPs are to scale relatively to each other. (B) The length of the lines corresponds the distance between the chromophore and the end of the linker (for FPs the peptide chain) at the maximum extension. (C) The differences in the linker length result in differences of the allowed conformational space of the fluorophores. (top) An accessible volume (AV) calculation of Alexa488 C5 maleimide attached to the LG-domain of a crystal structure of the human guanylate binding protein 1 is shown. (bottom) The entropic weighted conformational space of eGFP attached the N-terminal domain of a homology model of mGBP is shown for comparison.

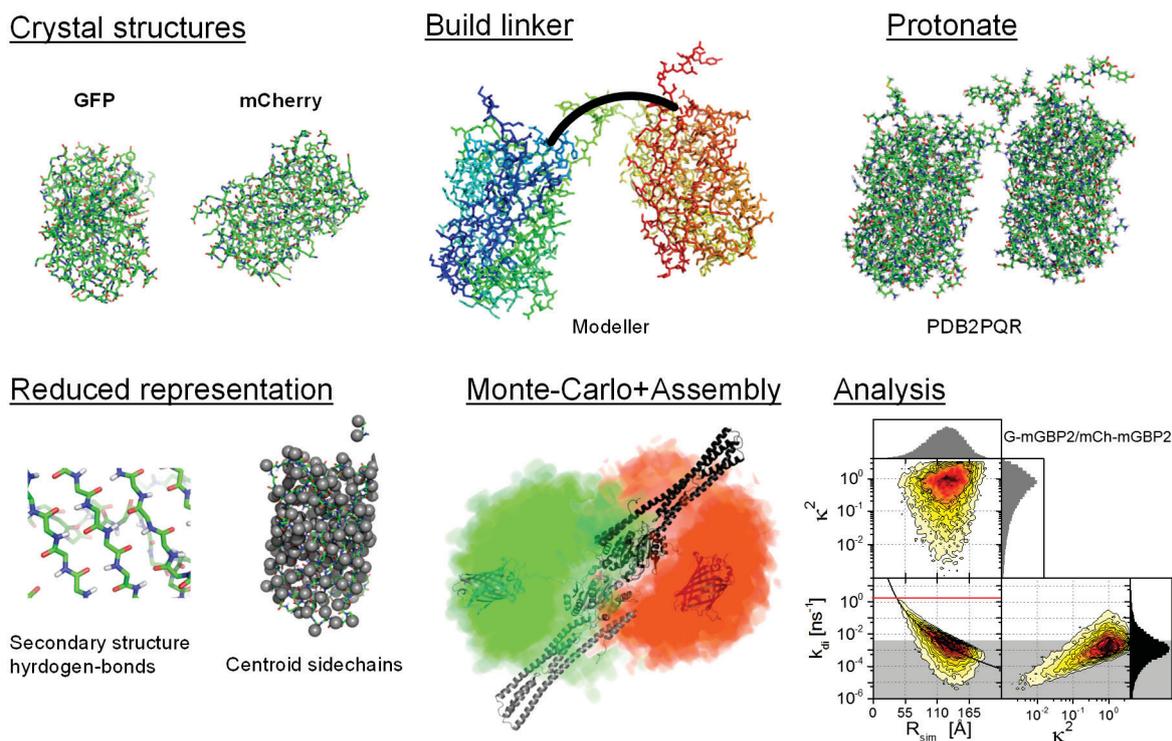
Simulations of the linker connecting the FPs to the protein were performed to compare the calculated fluorescence parameters of a structural model with the experimental data. To validate the method, GFP-mCherry tandems with increasing length were measured and simulated (see eq. 1). Using these simulations the FRET-rate constants distribution were calculated and compared to the experimental data. By directly using the FRET-rate constants distribution we bypassed the orientation factor problem, as it is implicitly contained in the FRET-rate constant. Using these simulations, it was tested if an average orientation factor distribution  $p(\kappa^2)$  is independent of the distance, as this would imply that they can be treated separately.

### 2.3 Monte-Carlo simulations

To compare a structural model to experimental data, the spatial distribution and the mutual orientation of transition dipole moments have to be considered. The fluorescent proteins (FPs)

are coupled by long linkers to the protein of interest. Thus, the conformation space of the protein-linker has to be considered to compare of protein structures to the experimental data. For a given structure, a whole set of linker conformations will be present each having its own interchromophore distance and a mutual orientation factor. Once a conformational ensemble for a proposed structural model of the protein(s) is generated, the experiment can be used as litmus test. To generate such conformational ensembles, we performed Monte-Carlo (MC) simulations of the linker which connects the FPs to its host. Given such an ensemble the FRET-rate constants distribution  $p(k_{RET})$  along with the orientation factor distribution can be directly calculated as outlined in Table 1 and Fig. 1. This distribution of rate constants  $p(k_{RET})$  can be directly compared to homo- or hetero-FRET experiments with fluorescent proteins (FPs).

To obtain sterically allowed conformational ensembles, the crystal structures of FPs have been fused in structural models (see Fig. 3). To sample the sterically allowed conformational space of the connecting linker between the fluorescent protein and the host-molecule, we convert the structure to a reduced representation, as outlined for a GFP-mCherry fusion protein in Fig. 3. The reduced representations in Fig. 3 were constructed based on the GFP and mCherry crystal structures (PDB-ID: 4EUL and 2H5Q, respectively). Homology models of the fusion proteins were constructed using the software MODELLER. These homology models were protonated by the software PDB2PQR. Then the protonated full-length protein models were mapped to a reduced representation solely consisting of the C-, C $\alpha$ -, N-, O- and the hydrogen atoms forming the NH-O bonds. The repulsion between the atom pairs (O, N), (C, O) and (C, N) were modeled as repulsive quadratic potential and the existing hydrogen bonds as simple scaled attractive potential ( $1/r$ ) preserving secondary structural elements. The MC-simulations on reduced structural models were performed using the  $\psi$ - and  $\phi$ -torsion angles of the linker. In each iteration step the torsion angle of one amino acid was changed by a random value taken from a normal distribution with a width of 0.025 rad. Only the internal coordinates of the connecting linker (Fig. 2, in blue) were altered. The internal coordinates of the host-molecule and the beta-barrels of the FPs were kept constant (Fig. 2, in green and red).



**Figure 3.** Simulation workflow – using the known primary sequence, crystal structures of GFP and mCherry (PDB-ID: 4EUL and 2H5Q respectively) homology models have been generated using Modeller. These models have been protonated with PDB2PQR. For faster calculations these initial models were reduced to a coarse grained representation consisting of C-,  $C\alpha$  -, N-, O- and the hydrogen atoms forming the NH-O bonds. The side-chains were replaced by a single centroid atom. The sampling of the possible conformations was performed on dihedral angle. After sampling dimer models were constructed using the known crystal structure interface. Random frames were taken to calculate the orientation factor, distance and FRET-rate constants. In case of clashes the combination of frames was rejected.

To calculate the donor-acceptor distance,  $R_{DA}$ , of every structure, on each fluorophore, two  $C\alpha$ -atoms on the beta-barrel (Asn122 and Asn147 on GFP, Tyr125 and Glu149 on mCherry) were chosen, so that the connecting vector of the two atoms is a good approximation of the transition dipole. The distance between the middle points of the connecting vectors of the donor and acceptor is taken as the distance between the chromophores,  $R_{DA}$ . Table 1 lists the detailed calculation steps. For every simulated structure, given the DA-distance and the orientation factor the FRET-rate was calculated according to Table 1, in which  $\tau_0$  is 2.6 ns and the Förster radius,  $R_0$ , of GFP and mCherry is  $52\text{\AA}$  including  $\kappa^2=2/3$ .

Simulations on mGBP2 were performed similarly. Here, based on the hGBP1 crystal structure (Prakash et al., 2000) homology models of the G-mGBP2 (PDB-ID: 1F5N, 4EUL) and mCherry-mGBP2 fusion protein (PDB-ID: 1F5N, 2H5Q) were constructed using MODELLER (Fiser and Sali, 2003). In both cases the sampling of the conformational space was restricted to the linkage region. Thus, only the internal coordinates of the connecting linker were altered

while the internal coordinates of the beta-barrels as well as the internal coordinates of the mGBP2 model were kept constant. Given the sampled conformation of the mCh-mGBP2 and the G-mGBP2 constructs a putative head-to-head dimer structures was created by superimposing the LG-domains onto the LG-domains in the dimer structure of hGBP1 in presence of GppNHp (PDB-ID: 2BC9) and discarding conformations with clashes (Vöpel et al., 2014).

In both cases the simulation and data analysis was performed in custom software programmed in Cython and Python. To process the trajectory the coordinates were saved in HDF5 file format compatible to MDTraj (McGibbon et al., 2015). A typical outcome of such a simulation is displayed in Fig. 3 for the mGBP2-GFP:mGBP2-mCherry dimer. The simulations allowed to study the correlated effects between  $\kappa^2$  and distance distributions (see Fig. 3).

## 2.4 Effect of orientation factor

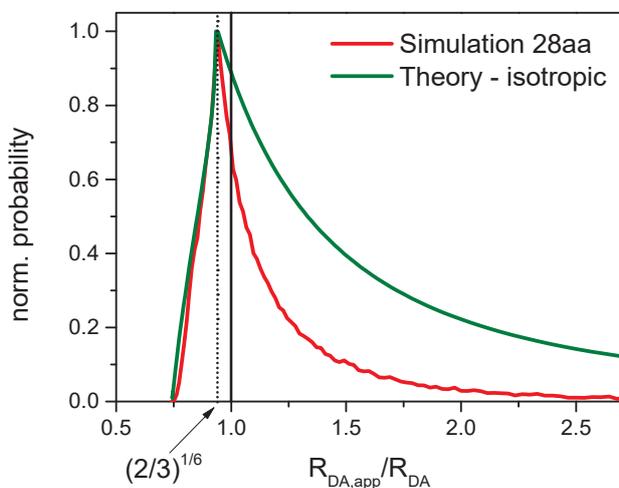
As the actual distributions of the orientation factors,  $p(\kappa^2)$ , are usually unknown, the fluorescence intensity decays or the FRET-induced donor decay are typically analyzed assuming an average orientation factor of  $2/3$ . By this analysis not the actual DA-distance distribution  $p(R_{DA})$  but an apparent distance distribution  $p(R_{DA,app})$  is obtained. The simulations and the theoretical  $\kappa^2$  distribution given by equation (3) are used to highlight differences between  $p(R_{DA})$  and  $p(R_{DA,app})$  in Fig. 4 and Fig. 5.

The theoretical isotropic orientation factor distribution has a sharp initial drop followed by a slower drop for orientation factors bigger than 1. The effect of this distribution on the apparent distance distribution is hard to imagine. However, the orientation factor distribution  $p(\kappa^2)$  can be transformed into an apparent distance distribution using the eqs. (1) and (2) ( $\kappa^2 = 2/3 \cdot (R_{DA}/R_{DA,app})^6 = 2/3 \cdot \xi^6$ ). The partial derivative  $\partial(\kappa^2)/\partial\xi = 4 \cdot \xi^5$  is now used to transform the random variable  $p(\kappa^2)$  in eq. (3) into a distribution of the ratio of the real-distance and the apparent distance  $p(R_{DA}/R_{DA,app})$ :

$$p(\xi) = \begin{cases} \sqrt{2} \cdot \ln(2 + \sqrt{3}) \cdot \xi^2 & 0 \leq \xi \leq \sqrt[6]{\frac{3}{2}} \\ \sqrt{2} \cdot \ln \left( \frac{2 + \sqrt{3}}{\sqrt{\frac{2}{3}} \xi^6 + \sqrt{\frac{2}{3}} \xi^6 - 1}} \right) \cdot \xi^2 & \sqrt[6]{\frac{3}{2}} \leq \xi \leq \sqrt[6]{6} \end{cases} \quad \text{with} \quad \xi = \frac{R_{DA}}{R_{DA,app}} \quad (4)$$

This apparent distance distribution first rises and then decays. This is shown as green line in Fig. 4. This slow decay is due to the fast initial drop of the  $\kappa^2$  distribution for small  $\kappa^2$ . Similarly,

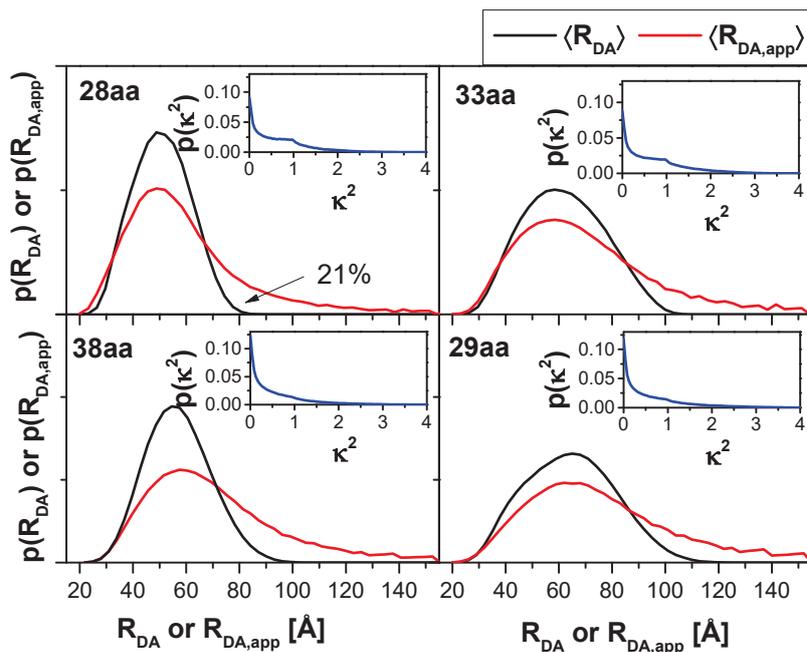
the orientation factor distribution of a simulation can be converted to an apparent distance distribution (see Fig. 4&5).



**Figure 4.** Effect of an orientation factor distribution on an apparent distance distribution. Apparent distance distribution for a single distance  $R_{DA}=55$  Å for a fluorophore with a single donor-lifetime in absence of FRET ( $\tau_{D(0)}=2.6$  ns) given an isotropic orientation factor distribution  $p(\kappa^2)$  (green) or a simulated orientation factor distribution (red) of a GFP-mCherry tandems with 28 flexible amino-acids.

The transformed simulated orientation factor distribution shows the same features as the theoretical orientation factor distribution, a step rise and a slow decay. As the theoretical distribution is broader and decays slower, the effect of the orientation factor on the simulated DA-distance distributions is smaller than theoretically anticipated by equation (4). In both cases the peak of the distribution  $p(\zeta)$  is nearly centered around one. The long tails towards large  $R_{DA,app}/R_{DA}$  ratios results in very low FRET species that may be misinterpreted as molecules missing an acceptor (donor-only). The theoretical orientation factor distribution predicts a much higher fraction of FRET-inactive species compared to the simulated orientation factor distribution.

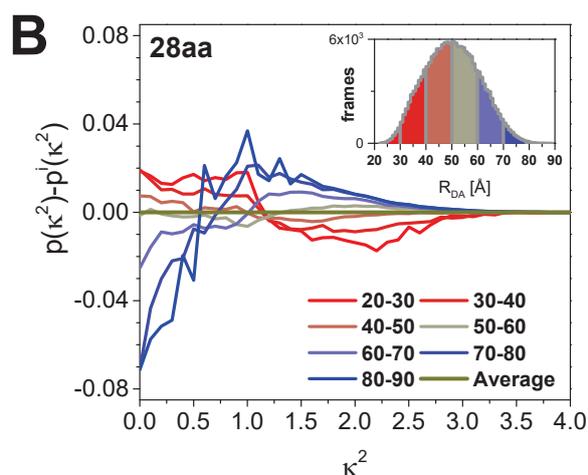
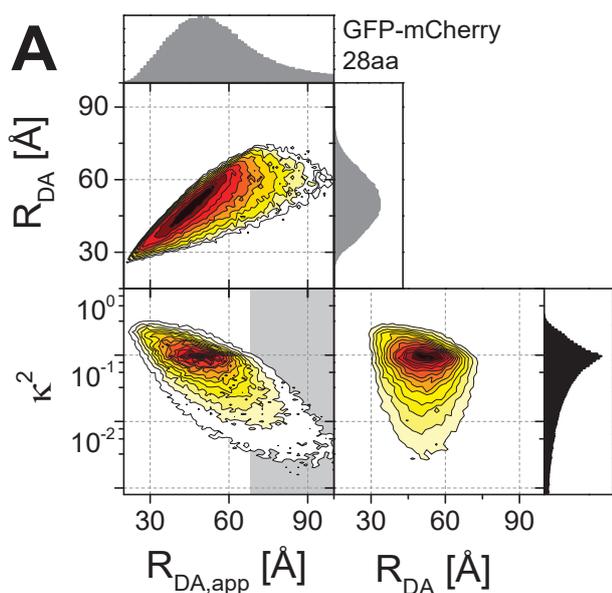
Using the simulations, the effect of the orientation factor can be summarized by overlaying the apparent distance distribution with the actual distance distribution (see Fig. 5). This clarifies that the orientation factor “smears out” the distance distribution and results in tailing towards low-FRET. In a typical FRET-experiment, apparent distances above approx. 70 Å (Förster-radius 52 Å) are indistinguishable from donor-only molecules. Hence, the expected fraction of FRET-inactive species due to orientation effects in case of the GFP-mCherry tandem with the shortest linker is approximately 20 % (see Fig. 5). To sum up, orientation effects result in a broadening of the apparent distance distribution. The fraction of FRET-inactive species increases since the broadening is not symmetric. Molecular configurations with small orientation factor will result in apparent long distances which are implicitly accounted in the analysis by the donor fraction.



**Figure 5** Comparison of the DA-distance distribution  $p(R_{DA})$  (black) and the apparent distance distribution  $p(R_{DA,app})$  (red) of the Monte-Carlo simulations for the GFP-mCherry tandems with 28, 33, 38, 54 flexible elements (amino acids) in the connecting linker. The average orientation factor distributions  $p(\kappa^2)$  are shown as inset.

If the orientation factor distribution is known  $p(\kappa^2)$  and independent of the DA distances, it can be directly accounted for in the analysis of the fluorescence intensity decays to recover the actual DA-distance distributions. As illustrated in Fig. 6A, correlations between simulated DA distances distribution of the orientation factor  $\kappa^2$  and the separation distances  $R_{DA}$  are weak, and a plot of the apparent distance  $R_{DA,app}$  versus the orientation factor  $\kappa^2$  shows as expected a clear correlation. Essentially, no distances above 70 Å are found in the  $R_{DA}$  distribution. Hence, all apparent distances above 70 Å in the  $R_{DA,app}$  distribution correlate with a small orientation factor.

To assess correlations between the separation distance  $R_{DA}$  and the orientation factor distribution in more detail, the difference of the orientation factor distributions for a given distance range from the average orientation factor distribution over the whole distance range are plotted in Fig. 6B. For large separation distances ( $R_{DA} > 70$  Å), this difference is skewed towards lower orientation factors compared to the average distribution over the whole ensemble. In conformations where the chromophores are much separated, the linker between GFP and mCherry is presumably nearly fully extended which results in preferred orientations of the chromophore dipoles. Strong deviations are only observed in the wings of the distance distribution: for very short distances and very large distances. These results indicate that an average orientation factor distribution  $p(\kappa^2)$  over the whole ensemble can be used, once reliably simulated, to recover not only the apparent distance distribution but the actual distance distribution from experiment from experiments.



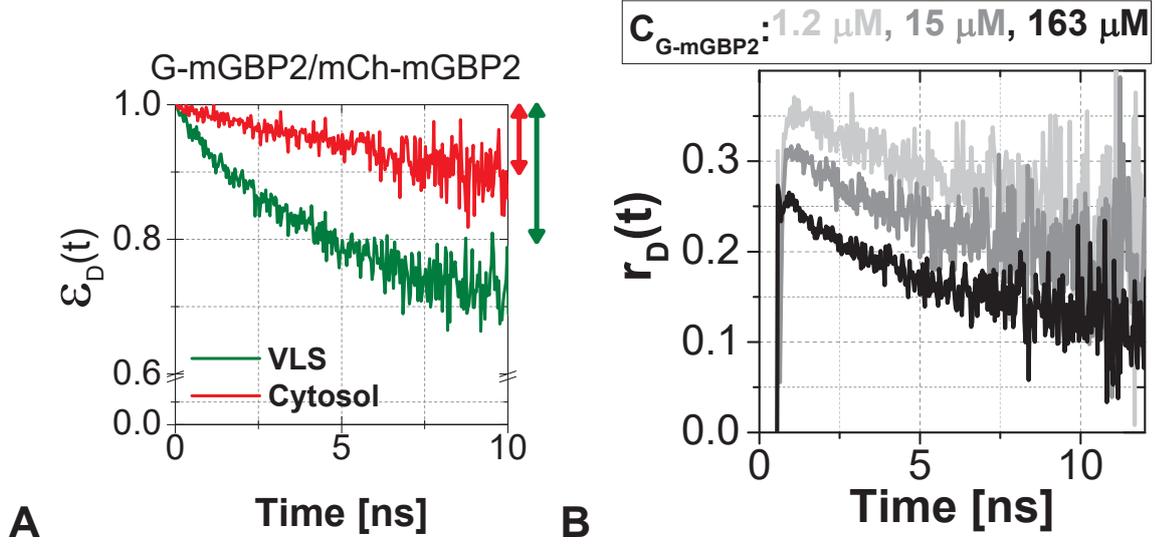
**Figure 6.** Correlation between distance and orientation factor distribution of Monte-Carlo simulations on a GFP-mCherry fusion protein connected by a flexible peptide linker consisting of 28 amino-acids. (A) Multi-dimensional frequency histograms of the apparent donor-acceptor distance  $R_{DA,app}$ , the orientation factor  $\kappa^2$  and the actual donor-acceptor distance  $R_{DA}$ . (B) Deviation of the orientation factor distribution for given donor-acceptor distances (shown as bars in the inset) from the average orientation factor distribution over the whole distance distribution

dimensional frequency histograms of the apparent donor-acceptor distance  $R_{DA,app}$ , the orientation factor  $\kappa^2$  and the actual donor-acceptor distance  $R_{DA}$ . (B) Deviation of the orientation factor distribution for given donor-acceptor distances (shown as bars in the inset) from the average orientation factor distribution over the whole distance distribution

## 2.5 Quantifiers in MFIS

Compared to *in vitro* FRET measurements quantitative MFIS-measurements in living cells are challenging as the precise fluorophore concentration cannot be controlled. In turn some assumptions made in chapter A are not valid any more, i.e. a donor might be surrounded by more than one acceptor and hence be strongly quenched. This is problematic as the fluorescence decay in presence  $f_{DD}^{DA}(t)$  and in absence of FRET  $f_{DD}^{D0}(t)$  are measured individually on different samples and to obtain the FRET-induced donor decay  $\varepsilon_D(t)$  both fluorescence decays have to be scaled to the same initial number of excited molecules at time  $t = 0$ . Consequently, strong quenching of the donor fluorescence by FRET may be overlooked if the fluorescence lifetime of the donor in presence of FRET is much shorter than the instrument response function.

Contrary to the FRET-induced donor decay, in time-resolved anisotropy measurements the initial amplitude is known and defined by the fundamental anisotropy  $r_0$  of the fluorophore. However, as FPs have a long rotational correlation time compared to the time-scale of FRET, the depolarization by homo-FRET can be used to measure distances and strong quenching by FRET is not overlooked like in a hetero-FRET experiment since the initial amplitude is known. This comparison is shown in Fig. 7 where a scaled FRET-induced donor-decay  $\varepsilon_D(t)$  of a hetero-FRET sample is shown next to the time-resolved anisotropy of the donor  $r_D(t)$  of a homo-FRET sample.



**Figure 7.** Time-resolved FRET-observables: (A) FRET-induced donor decay  $\varepsilon_D(t)$  of a representative GFP-mGBP2/mCherry-mGBP2 dimer. The drop in  $\varepsilon_D(t)$  curves, as marked by the arrows, represents the species fractions of FRET-active complexes ( $x_{RET}$ ) in vesicle-like structures (green) and in the cytosol (red). (B) Homo-FRET of mGBP2 at higher concentration exhibited larger quasi instantaneous drop of  $r_D(t)$  from its initial value of  $\sim 0.35$ , which proves the appearance of a very fast depolarization process due to homo-FRET in mGBP2 oligomers.

The presence of multiple acceptors demands for an extension of the formalism presented in chapter A which is only valid for single fluorophore pairs. In ensemble measurements, the donor is surrounded by multiple acceptors and its fluorescence is quenched by all acceptors, each having its own FRET-rate constant. Usually, the contribution by inter-molecular FRET is neglected as the concentration of acceptors is sufficiently low. However, due to the long linkers of the FPs and their high concentration, in the MFIS measurements inter-molecular contributions and multiple acceptor fluorophores have to be considered. In presence of multiple acceptors the normalized population of the excited state of donor in presence of  $N$  acceptors  $n_{D|D}^{DA}(t)$  for a donor fluorophore with a single lifetime  $\tau_0$  is given by:

$$\begin{aligned}
 n_{D|D}^{DA}(t) &= e^{-\frac{t}{\tau_0}} \cdot \prod_{k=1}^N \left( \int p(R_k) e^{-\left(\frac{R_0}{R_k}\right)^6 \frac{t}{\tau_0}} dR_k \right) \\
 &= n_{D|D}^{D0}(t) \cdot \varepsilon_D(t)
 \end{aligned} \tag{5}$$

Analogous to chapter A this can be extended to multi-exponential donor fluorescence decays if the cause of multi-exponential donor fluorescence decays is uncorrelated to FRET. In FPs diffusion can be neglected as already in organic dyes the mutual diffusion of the donor and acceptor dyes is of minor importance. Hence, the FRET-induced donor decay  $\varepsilon_D(t)$  yields the distance distribution in case if only one acceptor is in proximity to the donor ( $N=1$ ) and if orientation is

not explicitly accounted for. Orientation effects are discussed in more detail above. In chapter A inter-molecular quenching of donor fluorophore ( $N > 1$ ) were not considered.

As the protein concentrations in the MFIS measurements are significantly higher and long linkers are used, the acceptors which are further apart from the donor have to be considered and concentration dependent FRET-induced donor decays as result obtained. This concentration dependence was exploited to determine equilibrium constants of dimerization to monitor the formation of higher order oligomers and determine equilibrium constants of complex formation. The long linkers cover a larger space to monitor inter-molecular interactions at the cost of much lower intra-molecular resolution. In case of mGBPs the experimental fluorescence intensity decays were fitted by simulated FRET-rate distributions of the dimer:

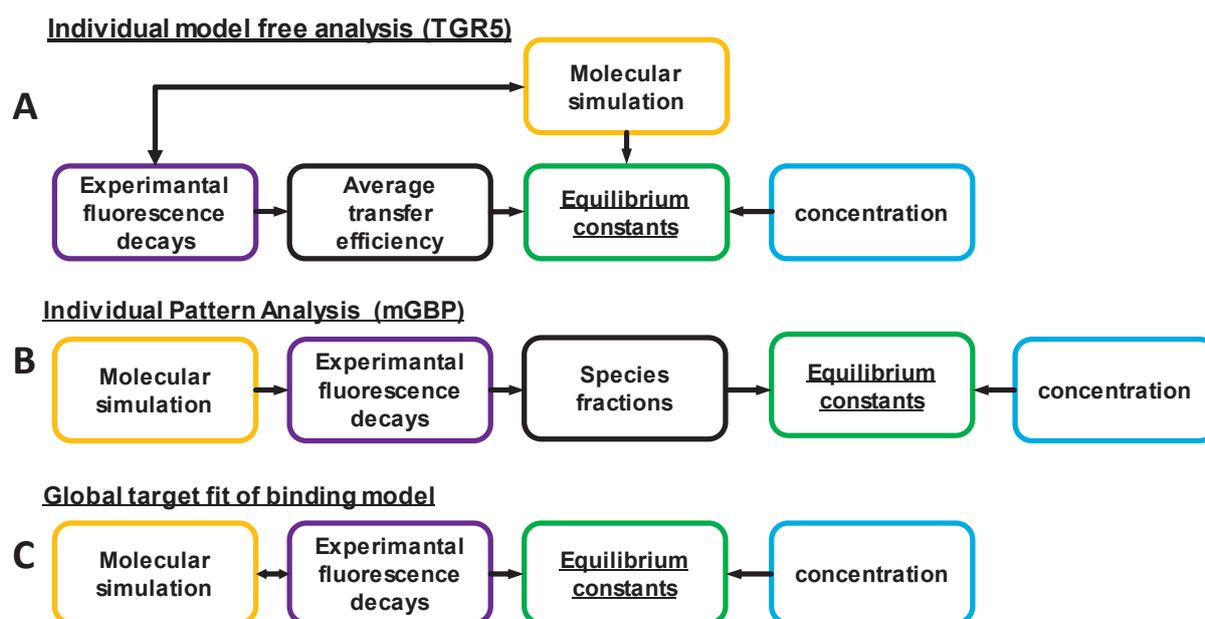
$$\begin{aligned}
f_{D|D}^{DA}(t) &= e^{-\frac{t}{\tau_0}} \cdot \prod_{k=1}^N \left( \int p(R_k) e^{-\left(\frac{R_0}{R_k}\right)^6 \frac{t}{\tau_0}} dR_k \right) \\
&= e^{-\frac{t}{\tau_0}} \cdot \left( \int p(R_1) e^{-\left(\frac{R_0}{R_1}\right)^6 \frac{t}{\tau_0}} dR_1 \right) \cdot \prod_{k=2}^N \left( \int p(R_k) e^{-\left(\frac{R_0}{R_k}\right)^6 \frac{t}{\tau_0}} dR_k \right) \\
&= f_{D|D}^{D0}(t) \cdot \varepsilon_1(t) \cdot \varepsilon_{2-N}(t)
\end{aligned} \tag{6}$$

In mGBP we determine the shape of  $\varepsilon_{di}(t)$  by simulations and measurements at low acceptor concentrations. Hence, we were able to determine the fraction of monomer, dimer and oligomer. In case of TGR5 the FRET-induced donor-decay  $\varepsilon_D(t)$  was fitted formally by two FRET-rate constants.

### 3 Determination of association constants

#### 3.1 Approaches to determine equilibrium constants

Given experimental fluorescence intensity decays and concentrations, equilibrium constants of oligomerization can be determined in at least three different ways (see Fig. 8). (i) The experimental fluorescence intensity decays can be described by a formal model to obtain the average transfer efficiency. Using these average transfer efficiencies and the protein concentrations the equilibrium constants are determined by making assumptions regarding the geometry of the complex. (ii) Alternatively, the molecular models of the complexes are constructed. Using these models, the fluorescence intensity decays are fitted to recover the species fractions. Later, binding isotherms are fitted to the species fractions. (ii) Finally, the equilibrium constants could be determined directly using the fluorescence intensity decays without an intermediate analysis step to determine species fractions using a global fit over all fluorescence decay curves.



**Figure 8.** Workflows to determine equilibrium constants by (A) individual model free analysis (TGR5), (B) individual pattern analysis (mGBP) and (C) global target fit of a binding model. The arrows indicate the direction in which information is processed. (A) The TGR5 experimental data was analyzed by a formal model and summarized in the form of average transfer efficiencies. The molecular simulations were later on cross validated by the experimental data. Using results from the molecular simulation and concentrations determined by the fluorescence intensities, equilibrium constants were determined. (B) The mGBP equilibrium constants were determined following to linear scheme starting with molecular simulations. Individual fluorescence intensity decays of differing concentrations were fitted to determine species fractions. Using the concentrations and the species fraction the equilibrium constants were determined. (C) In the global target fit of a binding model the intermediate step to determine species fractions is skipped and the equilibrium constants are directly obtained by a global analysis.

The first approach shown in Fig. 8A was used in the TGR5-project. It has the advantage that all essential steps (molecular simulations, data-acquisition and analysis) are independent. As solely a formal model and no structural model is required to describe the experimental data it can be processed in early project stages. In the TGR5-project the experimental data was formally analyzed by two FRET-rate constants, which were used to calculate average FRET-efficiencies. At a later project stages these experimental results were compared to a molecular model to unified structural models and experiments and to yield equilibrium constants. A reduced structural model and the concentrations are used together with the given equilibria to calculate the FRET-efficiency (results are shown in Fig. 9). Even though most relevant steps can be performed in parallel, this method has several disadvantages: the experimental data is pre-processed and not directly compared to the binding equilibrium model, the concentration dependent analysis of the FRET-efficiencies is complicated and the molecular simulations are not streamlined and directly incorporated in all analysis steps.

The second approach (Fig. 8B) was used to process the data recorded in mGBPs. Here, the molecular simulations were directly incorporated into the analysis of the experimental data at

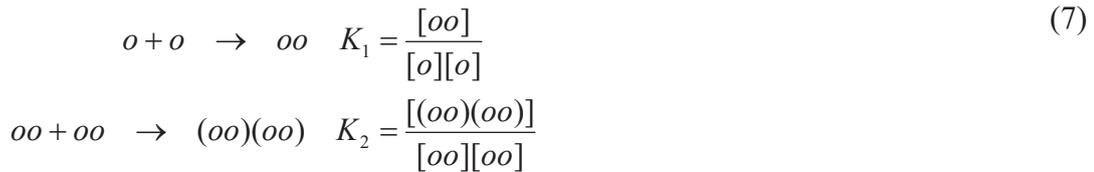
an early state. By that, the determination of the equilibrium constants is simpler as the species fractions of the complexes were determined directly. Using these species fractions, the equilibrium constants of complex formation were determined in a subsequent step (results shown in Fig. 10). The simpler analysis comes at the cost that a final data-analysis is only possible after a molecular model is presented.

Finally, a global set of the fluorescence intensity decay measurements performed at different protein concentrations could be analyzed globally using a target fit (Fig. 8C). This fit has to incorporate the binding model, the molecular models of the different complexes and the concentrations. This approach has the advantage that error-estimates could be determined precisely using the counting statistics of the fluorescence intensity decay.

### 3.2 Forward modelling of oligomer species

To calculate the fluorescence intensity decays the concentrations of all species have to be known. Calculating the species fractions for multiple chemical equilibria is a standard problem in analytical aquatic and geochemistry termed chemical speciation (Parkhurst and Appelo, 2013). For given equilibrium constants and total protein concentration, the species fraction of all oligomer species can be calculated as outlined below on the example of a tetramerization.

For the mGBPs and TGR5, a simple monomer-dimer-tetramer model was used to describe the oligomerization. It was assumed that a tetramer is constituted by a dimer of dimers. Hence, starting from a monomer two equilibriums were treated:



Here  $o$  stands for a monomer while  $oo$  and  $(oo)(oo)$  are dimers and tetramers, respectively.

Then the total protein concentration is given by:

$$c_T = [o] + 2 \cdot [oo] + 4 \cdot [(oo)(oo)] \quad (8)$$

The concentrations of the three species  $o$ ,  $oo$  and  $(oo)(oo)$  for any given total protein concentration was obtained by solving the three equations above. For a given set of equilibrium constants and a total protein concentration  $c_T$  the species concentrations  $[o]$ ,  $[oo]$ ,  $[(oo)(oo)]$  were determined numerically. Compared to geochemical problems, we consider only a small number of equilibria. Hence, instead of using specialized software for chemical speciation calculations (PHREEQC, MINTEQA2, etc.), we solve the fourth order polynomial equation in  $c_T$  obtained by

substitution of eq. (7) into eq. (8) by Ridders method, a simple root-finding algorithm (Ridders, 1979).

For mGBP this model was extended to stepwise oligomer formation in a non-cooperative fashion (i.e. all equilibrium constants are equal) up to a dodecamer. If the total concentration of all oligomers (4-12) is used to display the binding isotherm, one obtains an only slightly broadened binding isotherm compared to the tetramer system. If this binding isotherm is fitted with the simpler tetramer model, a binding constant for dimer binding is obtained, which is slightly (factor 1.6) larger than the simulated value. As we have no information on the cooperativity of binding and the spatially resolved GTP concentration, we approximate the formation of higher order oligomers by the minimal tetramer model for the following reasons: (1) FRET only senses its local environment (i.e. a limited oligomer size) thus the contribution of each monomer unit to the measured signal decreases with increasing oligomer size. (2) This simple model reduces the number of fitting parameters to an adequate level given the spread of the data-points. To conclude, a simple model with a Langmuir binding isotherm (i.e. non-cooperative binding) describes all experiments very well.

### 3.4 Individual model free analysis

#### *Formal description of experimental data*

Alternatively, to the model which assigns species fractions of given structural models to experimental fluorescence decays, the fluorescence decays are formally described for instance by two FRET-rates:

$$\varepsilon(t) = (1 - x_{\text{Donly}}) (x_1 e^{-k_{\text{RET},1}t} + (1 - x_1) e^{-k_{\text{RET},2}t}) + x_{\text{Donly}} \quad (9)$$

Using these fitting results, a mean transfer-efficiency is determined. For each of mean transfer efficiency the concentrations of the donor and the acceptor are determined by the fluorescence intensities. Equilibrium constants are now estimated once a structural model is suggested. In case of the G-protein coupled receptor (GPCR) TGR5 a planar geometry as depicted in Fig 9B was assumed.

#### *Concentration dependent FRET*

As previously described, to estimate/determine the equilibrium constants, the total protein concentration and the equilibrium constants, which describe protein association, are needed to determine the oligomerization state (the chemical speciation). Additionally, the transfer-efficiency for a given oligomerization state, the spatial organization of the molecules within the oligomers

and the concentration of donor, acceptor and non-fluorescent molecules have to be known for average transfer efficiency calculation.

The total protein concentration, which controls the chemical speciation, is given by the sum of the acceptor, the donor and the unlabeled protein concentrations:

$$c_T = c_A + c_D + c_U \quad (4)$$

Here the unlabeled protein concentration  $c_U$  equals to the concentration of inactive acceptor molecules. To calculate the transfer efficiency, we assume that donor, acceptor and unlabeled molecules behave biochemically identical. Hence, the probability of an oligomer composition is given by the probability of finding a donor, acceptor or unlabeled molecule and the counting statistics. The probabilities of finding a donor, acceptor or unlabeled molecule depend on their respective concentrations. For instance, the probability of an acceptor molecule is given by the respective species and total protein concentration:

$$p_A = \frac{c_A}{c_T} \quad (7)$$

In a tetramer the sum of donor, acceptor and unlabeled molecules is constant. Hence, the probability of a certain tetramer composition is obtained by the multinomial distribution:

$$p(n_D, n_A, n_U) = N \cdot p_D^{n_D} p_A^{n_A} p_U^{n_U} \quad (8)$$

$$= \frac{(n_D + n_A + n_U)!}{n_D! n_A! n_U!} \cdot p_D^{n_D} p_A^{n_A} p_U^{n_U}$$

$N$  is the number of combinations for a given composition. Each combination might have a different FRET-rate constant distribution. Hence, in case of two donors and two acceptors six combinations as shown in Fig. 8A contribute to the signal. If only FRET-species with at least one donor and one acceptor are analyzed the FRET-rate constants of overall 38 distinct species and their respective probabilities and FRET-rate constant distributions have to be calculated. The species probabilities summarized by their donor and acceptor composition in dependence of the acceptor to donor ratio  $c_A/c_D$  are illustrated in Fig. 9C.

To calculate the resulting transfer-efficiency the species fractions and the transfer-efficiency of all possible oligomer species has to be considered. The FRET-rate constants are additive. Therefore, in case of multiple acceptors, the total FRET-rate constant experienced by a donor ( $i$ ) is given by the sum of all FRET-rate constants of all acceptors ( $j$ ):

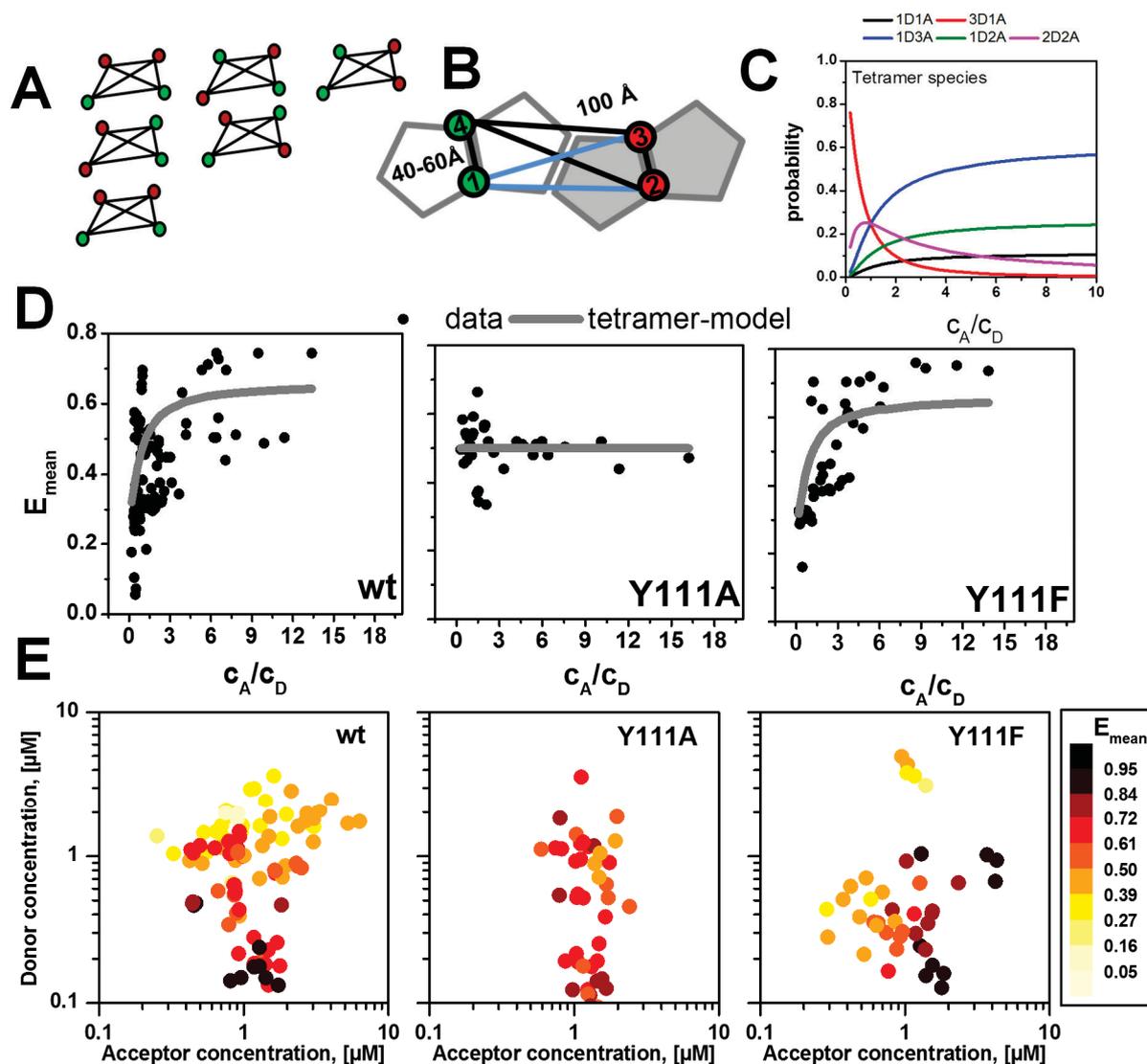
$$k_{RET}^{(i)} = \frac{1}{\tau_0} \cdot \sum_j \left( \frac{R_{DA}^{(ij)}}{R_0} \right)^6 \quad (9)$$

Here  $R_{DA}^{(ij)}$  is the donor acceptor distance between the donor ( $i$ ) and the acceptor ( $j$ ) which is determined by the spatial arrangement of the oligomer. For instance, in the case as illustrated in Fig. 9B the two FRET-rates experienced by the donor at position 1 and the donor at position 4 are given by:

$$k_{RET}^{(1)} = \frac{1}{\tau_0} \cdot \left( \left( \frac{R_{DA}^{(13)}}{R_0} \right)^6 + \left( \frac{R_{DA}^{(12)}}{R_0} \right)^6 \right) \quad (10)$$

$$k_{RET}^{(4)} = \frac{1}{\tau_0} \cdot \left( \left( \frac{R_{DA}^{(42)}}{R_0} \right)^6 + \left( \frac{R_{DA}^{(43)}}{R_0} \right)^6 \right)$$

For a given structural arrangement all FRET-rate constants for all possible compositions (one donor - one acceptor, two donors - one acceptor, etc.) were calculated. Later the average transfer-efficiencies of the tetramer compositions containing at least one donor and one acceptor were calculated. In this analysis the contribution to the fluorescence signal depends on the number of donor molecules. For instance, a tetramer constituted of three donors and one acceptor molecule contributes three times more to the total signal as compared to a tetramer only constituted out of one donor, one acceptor and two unlabeled molecules. Using these FRET-rate constants the mean-transfer efficiency is calculated as described in chapter A. The transfer efficiency for each data point depends now only on the equilibrium constants and the spatial arrangement of the fluorophores.



**Figure 9.** Description of TGR5 by a minimal dimer/tetramer model and effect of the acceptor and donor concentrations on the mean transfer-efficiency. (A) In a tetramer the sum of donor, acceptor and unlabeled molecules is constant. Six tetramer configurations for a case of two acceptor (red) and two donor molecules (green) are possible. (B) Composition of a simplified rectangular tetramer molecule with random arrangement of two donors and two acceptors according to a linear organization of the GPCR. In this model we assume that a tetramer is constituted of a dimer of dimers. (C) Probability distribution of different tetrameric species in dependence of the acceptor/donor-ratio,  $c_A/c_D$  (D) Fitted mean transfer-efficiency in dependence of the  $c_A/c_D$ . (E) Two dimensional plot of the mean transfer-efficiency in dependence of the donor and acceptor concentration.

To illustrate the dependency of the FRET-efficiency on the concentrations we choose two distinct representations of the experimental data. In Fig. 9D the experimental FRET-efficiency is displayed in dependence of the acceptor to donor concentration ratio  $c_A/c_D$ . The gray line in Fig. 9D corresponds to the FRET-efficiency as calculated by the fitted oligomerization model. To a certain degree, this representation is misleading as the FRET-efficiency depends on  $c_U$ ,  $c_D$  and  $c_A$  and all three concentrations were considered in the analysis (compare eq. (8)). Thus, the experimental FRET-efficiencies are additionally displayed Fig. 8E in a two dimensional manner to highlight the dependence of the acceptor and donor concentration,  $c_A$  and  $c_D$ , respectively.

To calculate the FRET-rate constants for TGR5 we assumed that the tetramer can be described by a rectangular geometry where one edge is approximately 100 Å long while the second edge is between 40-50 Å long (Fig. 9B). This assumption is in line with homology models provided by Christoph G.W. Gertzen. Additionally, we assumed that a tetramer is formed in two consecutive steps similar to mGBP2. In the data-analysis only FRET molecules have been selected. Therefore, the first equilibrium from monomer to dimer is not monitored and only the equilibrium constant of the tetramer formation is probed. Thus, only the equilibrium constant of tetramerization and the dimer distance in the range of 40-60 Å is reflected by the data. From the measurements we find that a short distance of approximately 45 Å describes the data best. For the TGR5 wt and Y111F variant we find predominately a tetrameric or higher-order oligomer configuration while in case of the Y111A mutant the molecules are predominately in a dimeric configuration. In principle this approach to analyze FRET-data can be easily extended to time-resolved fluorescence intensities, as in an intermediate step all FRET-rate constants are calculated.

The main advantage of individual model free analysis is that simple fluorescence models can be used to quantify the experimental fluorescence intensity decay. Using the analysis results of these models and assumptions regarding the geometry, the equilibrium constants of the complexes can be deduced by the theory described above. As the analysis, interpretation of the fluorescence intensity decays and the interpretation of concentration dependent FRET-efficiencies are independent of complex molecular simulations, this approach could be implemented straight forward in standard software for MFIS-analysis.

### 3.3 Individual pattern analysis

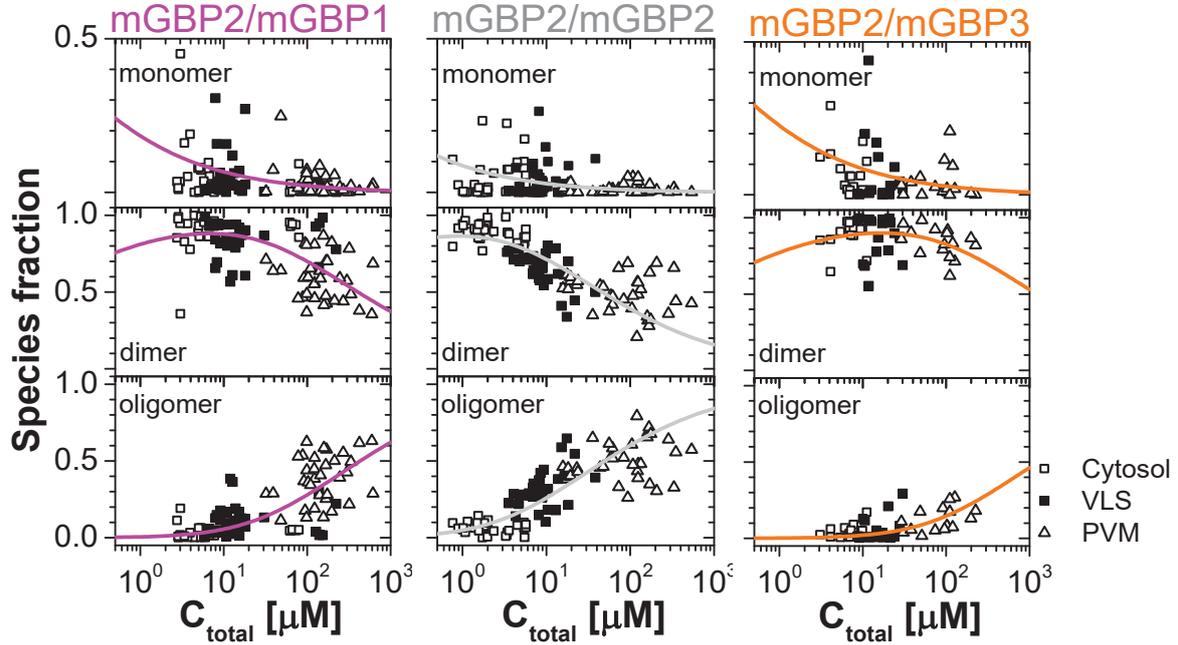
#### *Homo-FRET*

To quantitatively determine the fraction of individual protein species in mGBP, including monomer ( $x_{mono}$ ), dimer ( $x_{di}$ ) and oligomer ( $x_{oligo}$ ) fractions, the time-resolved anisotropy decay ( $r_D(t)$ ) of homo-FRET data was fitted by:

$$r_D(t) = r_0 \cdot \left( x_{mono} + x_{di} \cdot \left( \int p(k_{di}) e^{-2 \cdot k_{di} \cdot t} dk_{di} \right) + x_{oligo} \cdot \left( x_s e^{-2 \cdot k_{olig,s} \cdot t} + x_l e^{-2 \cdot k_{olig,l} \cdot t} \right) \right) \cdot e^{-t/\rho_{global}} \quad (10)$$

Here  $p(k_{di})$  is the FRET-rate distribution of mGBP2 dimer complex as determined by the conformational sampling of the sterically allowed space (see description above).  $k_{olig,s}$  and  $k_{olig,l}$  are formally assigned as the FRET rate constants of mGBPs oligomers of small and large sizes respectively, and  $x_s$  and  $x_l$  are their normalized fractions. It has to be considered that energy can

be transferred in forward and backward direction which doubles the rate constants. The fundamental anisotropy  $r_0$  for GFP molecules is known as 0.38.



**Figure 10.** Concentration dependence of the three mGBP species (monomer, dimer and oligomer) obtained by applying a global pattern fit (eqs. (11) and (10)) of  $r_D(t)$  and  $\varepsilon_D(t)$  for two localizations vesicle like structure (VLS) and the parasitophorous vacuole membrane (PVM). The line depicts the fit of the corresponding binding equilibrium with the dimerization constant  $K_1$ , and the apparent oligomerization constant  $K_2$

The global rotational correlation time  $\rho_{global}$  was approximated by 16 ns, the rotational correlation time of GFP. Oligomer species which produced ultrafast decay components in  $r_D(t)$ , result into a drop of the initial anisotropy (Fig. 7B). With the  $r_0$  knowledge they can be determined together with other species in homo-FRET data. Homo-FRET analysis of the anisotropy decay has the advantage compared to hetero-FRET that strong quenching by FRET is not overlooked, as the initial amplitude  $r_0$  is known.

### Hetero-FRET

Analogous to the homo-FRET, the hetero-FRET data was analyzed by the FRET-induced donor decay  $\varepsilon_D(t)$ :

$$\varepsilon_D(t) = x_{mono} + x_{di} \left( \int p(k_{di}) e^{-k_{di}t} dk_{di} \right) + x_{oligo,s} e^{-t \cdot k_{oli,s}} \quad (11)$$

where  $x_{oligo,s}$  denotes the species fraction of small oligomers. In the case of hetero-FRET, donor molecules in large oligomers (with species fraction  $x_{oligo,l}$ ) could be strongly quenched by nearby acceptors up to nearly 100% and thus became irresolvable owing to the finite width of the instrument response function. Therefore, the information of large oligomers in hetero-FRET

data needed to be recovered according to the homo-FRET data. In the latter, the species fractions of small and large oligomers were found equal in various compartments. Based on the relation  $x_{oligo,s} = x_{oligo,l}$  the large oligomer fractions in hetero-FRET data were extrapolated. Moreover, such a coherent behavior between small and large oligomers indicated that they have a common origin and the broad distribution of their rate constants showed that oligomers may consist of a variety number of units. Hence, it is more meaningful to combine both oligomer species and generally sort protein species as monomer, dimer and oligomer as displayed in Fig. 9. Given the observed species fractions the equilibrium constants were optimized by Broyden–Fletcher–Goldfarb–Shanno (BFGS) method, an iterative method for solving unconstrained nonlinear optimization problems, by minimizing the residuals of the species fraction compared to the calculated species fraction (fitting result see Fig. 10).

Note that the observed reduction in steady-state anisotropy ( $r_D$ ) for cells of high mGBP2 concentration as displayed in Fig. 4C, was mainly due to the large drop in the initial anisotropy of their time-resolved anisotropy ( $r_D(t)$ ) as plotted in Fig. 7D. Therefore the  $K_{D,app}$  value (9  $\mu$ M) derived from  $r_D$  in fact reports the mGBP2 oligomerization processes that could produce such ultrafast depolarizing effect, and is very close to the 8  $\mu$ M obtained by fitting  $r_D(t)$  with the species-resolved model. Hence, the two independent approaches interrogating the same oligomerization process delivered very consistent results, verifying the reliability of the analyses.

## 4 Conclusion

As illustrated above MFIS-measurements in living cells have the enormous potential to provide additional information. As the molecules of interest are measured there is less debate about the biological importance of the findings. As in the mGBP manuscript the MFIS can be utilized to monitor cell over long periods of times. Hence, biologically relevant processes as the invasion of a mammalian cell by a parasite can be directly monitored. Given a single measurement this process can be correlated quantitatively with biochemically relevant quantifiers such as equilibrium constants or concentrations. Additionally, MFIS allows studying structural features. However, given the size of fluorescent proteins and the linkers which are used to attach them to the host-molecule the potential for structural studied is limited.

## References

BETZIG, E., PATTERSON, G. H., SOUGRAT, R., LINDWASSER, O. W., OLENYCH, S., BONIFACINO, J. S., DAVIDSON, M. W., LIPPINCOTT-SCHWARTZ, J. & HESS, H. F. 2006. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313, 1642-1645.

- BLECKMANN, A., WEIDTKAMP-PETERS, S., SEIDEL, C. A. M. & SIMON, R. 2010. Stem Cell Signaling in Arabidopsis Requires CRN to Localize CLV2 to the Plasma Membrane. *Plant Physiology*, 152, 166-176.
- DICKSON, R. M., CUBITT, A. B., TSIEN, R. Y. & MOERNER, W. E. 1997. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature*, 388, 355-358.
- EVERS, T. H., VAN DONGEN, E. M. W. M., FAESEN, A. C., MEIJER, E. W. & MERKX, M. 2006. Quantitative understanding of the energy transfer between fluorescent proteins connected via flexible peptide linkers. *Biochemistry*, 45, 13183-13192.
- FISER, A. & SALI, A. 2003. MODELLER: Generation and refinement of homology-based protein structure models. *Macromolecular Crystallography, Pt D*, 374, 461-491.
- FÖRSTER, T. 1951. *Fluoreszenz organischer Verbindungen*, Göttingen, Germany, Vandenhoeck & Ruprecht.
- HELL, S. W. & WICHMANN, J. 1994. Breaking the Diffraction Resolution Limit by Stimulated-Emission - Stimulated-Emission-Depletion Fluorescence Microscopy. *Optics Letters*, 19, 780-782.
- HOFIG, H., GABBA, M., POBLETE, S., KEMPE, D. & FITTER, J. 2014. Inter-Dye Distance Distributions Studied by a Combination of Single-Molecule FRET-Filtered Lifetime Measurements and a Weighted Accessible Volume (wAV) Algorithm. *Molecules*, 19, 19269-19291.
- KALININ, S., FULLE, S., KÜHNEMUTH, R., SINDBERT, S., FELEKYAN, S., GOHLKE, H. & SEIDEL, C. A. M. 2015. Diffusion with traps: experiment, simulation, and theory to describe fluorophore dynamics in biomolecular FRET.
- KALININ, S., PEULEN, T., SINDBERT, S., ROTHWELL, P. J., BERGER, S., RESTLE, T., GOODY, R. S., GOHLKE, H. & SEIDEL, C. A. M. 2012. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nature Methods*, 9, 1218-U129.
- KUHNEMUTH, R. & SEIDEL, C. A. M. 2001. Principles of single molecule multiparameter fluorescence spectroscopy. *Single Molecules*, 2, 251-254.
- MCGIBBON, R. T., BEAUCHAMP, K. A., HARRIGAN, M. P., KLEIN, C., SWAILS, J. M., HERNANDEZ, C. X., SCHWANTES, C. R., WANG, L. P., LANE, T. J. & PANDE, V. S. 2015. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal*, 109, 1528-32.
- PARKHURST, D. L. & APPELO, C. A. J. 2013. Description of input and examples for PHREEQC version 3—A computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations. *U.S. Geological Survey Techniques and Methods*.
- PHAM, E., CHIANG, J., LI, I., SHUM, W. & TRUONG, K. 2007. A computational tool for designing FRET protein biosensors by rigid-body sampling of their conformational space. *Structure*, 15, 515-523.
- PRAKASH, B., PRAEFCKE, G. J., RENAULT, L., WITTINGHOFER, A. & HERRMANN, C. 2000. Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins. *Nature*, 403, 567-71.
- RIDDERS, C. 1979. A new algorithm for computing a single root of a real continuous function. *IEEE Transactions on Circuits and Systems*, 26, 979-980.

- RUST, M. J., BATES, M. & ZHUANG, X. W. 2006. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, 3, 793-795.
- SINDBERT, S., KALININ, S., NGUYEN, H., KIENZLER, A., CLIMA, L., BANNWARTH, W., APPEL, B., MÜLLER, S. & SEIDEL, C. A. M. 2011. Accurate distance determination of nucleic acids via Förster resonance energy transfer: implications of dye linker length and rigidity. *J Am Chem Soc*, 133, 2463-80.
- SOMSSICH, M., MA, Q. J., WEIDTKAMP-PETERS, S., STAHL, Y., FELEKYAN, S., BLECKMANN, A., SEIDEL, C. A. M. & SIMON, R. 2015. Real-time dynamics of peptide ligand-dependent receptor complex formation in planta. *Science Signaling*, 8.
- STRIKER, G., SUBRAMANIAM, V., SEIDEL, C. A. M. & VOLKMER, A. 1999. Photochromicity and fluorescence lifetimes of green fluorescent protein. *Journal of Physical Chemistry B*, 103, 8612-8617.
- VAN DER MEER, B. W. 2002. Kappa-squared: from nuisance to new sense. *Reviews in Molecular Biotechnology*, 82, 181-196.
- VAN DER MEER, W. B., VAN DER MEER, D. M. & VOGEL, S. S. 2014. Optimizing the Orientation Factor Kappa-Squared for More Accurate FRET Measurements. In: MEDINTZ, I. & HILDEBRANDT, N. (eds.) *FRET - Förster Resonance Energy Transfer: From Theory to Applications*. Weinheim, Germany: Wiley-VCH Verlag.
- VÖPEL, T., HENGSTENBERG, C. S., PEULEN, T.-O., AJAJ, Y., SEIDEL, C. A. M., HERRMANN, C. & KLARE, J. P. 2014. Triphosphate induced dimerization of human guanylate binding protein 1 involves association of the C-terminal helices: A joint double electron-electron resonance and FRET study. *Biochemistry*, 53, 4590-4600.
- WIDENGREN, J., KUDRYAVTSEV, V., ANTONIK, M., BERGER, S., GERKEN, M. & SEIDEL, C. A. M. 2006. Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. *Analytical Chemistry*, 78, 2039-2050.

## Chapter E - Refining fluorescence tools to resolve hidden conformational states of T4L by FRET

---

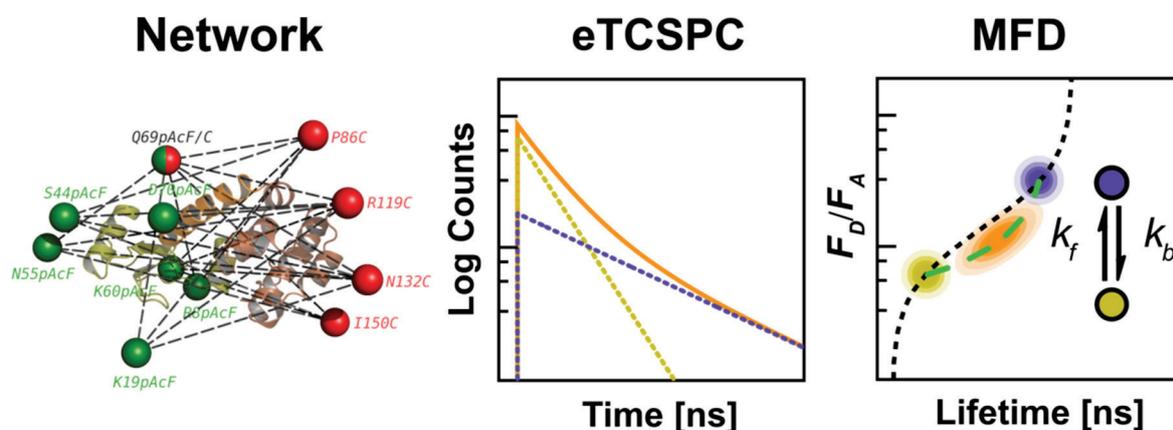
### 1 Introduction

Rigorous data analysis of fluorescence image spectroscopy data, recorded by multi-parameter fluorescence detection (MFIS) using established fluorescent proteins (FPs) as reporters, provides a wealth of information which can be exploited to determine equilibrium constants and validate structural models and thus opens new domains for molecular biophysics (see chapter D). However, the potential of FPs to determine structural models is limited. This is due to a number of reasons. FRET-guided structural modelling, for instance by the FRET-positioning system (FPS) (Kalinin et al., 2012), demands for a large set of high-precision FRET (hpFRET) measurements. FPs are usually either attached to the C- or the N-terminus of the protein of interest by very long peptide linkers (100–200 Å). Hence, the resolution and the number of obtained distance constraints is limited. Additionally, FPs lack brightness and photostability compared to cyanines or rhodamines, their counterparts used in single-molecule spectroscopy which prohibits single-molecule studies.

In nucleic acids similar problems are addressed by highly fluorescent base analogues and specialized modelling approaches (Wilhelmsson et al., 2001; Preus et al., 2013). This combination is extremely successful in nucleic acids but it is not generally applicable. Hence, orthogonal coupling of photostable fluorophores by linkers of sufficient length to guarantee rotational freedom is the current gold standard in FRET-measurements. The linkers (~20 Å) assure the free rotation of the dyes (Sindbert et al., 2011). Therefore, the dynamic orientation regime applies which significantly facilitates the analysis of fluorescence data (Kalinin et al., 2015). Such photostable reporters with all their advantages are on the verge of being regularly used in living cells (Sustarsic and Kapanidis, 2015) and recently made single-molecule measurements in living cells possible (Konig et al., 2015).

Even though the FRET theory is highly developed, FRET is often considered a low-resolution technique. Thus FRET is mainly used for qualitative statements and rarely used for structural modelling. However, it was recently shown that multiple FRET-measurements can be culminated into quantitative structural models by which explicitly consider the spatial probability distribution of the fluorophores (Kalinin et al., 2012). Practically, the fact that FRET is seldom used as a technique for structure determination is on one side caused by the absence of clear procedures and defined standards for streamlined, automated protocols for data-analysis and on the other side by the need for multiple FRET-pairs. This makes the determination of structures

of proteins and enzymes by FRET a tedious endeavor. Therefore, even though FRET has the marvelous potential to identify hidden states (Grossman et al., 2015), this potential is rarely explored for structural modeling of such transiently populated (or hidden) states.



**Figure 1.** T4 Lysozyme (T4L) distance network and schematics of the high precision FRET by fluorescence decay analysis (eTCSPC) and multi-parameter fluorescence detection (MFD). The multidimensional FRET-histograms are used to identify states and serve as fluctuation analysis toolkit. Ensemble time-correlated single photon counting (eTCSPC) resolve stable populations on the ns timescale. Single-molecule MFD experiments of freely diffusing are analyzed by the FRET indicator  $F_D/F_A$  (ratio of donor over acceptor fluorescence) accumulated during single burst duration (millisecond timescale) and detect conformational dynamics. Unimodal distributions occur when dynamic conformational mixing is faster than the burst duration and follow the dashed green line in the MFD plot.

Enzymes are biological catalysts that accelerate chemical reactions. Therefore, the determination of structures of enzymes is interesting. During catalysis, enzymes sample defined structural conformations (Hammes, 1964), including low populated states over a wide temporal domain (Henzler-Wildman and Kern, 2007). In enzymes structure and function are linked by conformational transitions (Smock and Gierasch, 2009). Transient states are also related with disease (Selkoe, 2003). During the last decade, various experimental and theoretical tools were introduced to study the enzymes' dynamic behavior (Henzler-Wildman and Kern, 2007, Chung et al., 2009) and our view of rigid enzymes has changed. However, gathering a full overview of enzyme dynamics and characterizing excited conformational states is one of the most difficult experimental problems in molecular biophysics. Transient states are frequently termed “hidden states” because they are often short lived and hidden in classical methods used in structural biology. All spectroscopic techniques such as nuclear magnetic resonance (NMR) (Kleckner and Foster, 2011), electron paramagnetic resonance (EPR) (McHaourab et al., 1996), FRET (Woźniak et al., 2008), and lower resolution methods (Henzler-Wildman and Kern, 2007), face different method-specific restrictions in sensitivity, size, molecular flexibility as well as spatial and temporal resolution.

In the enzyme T4 lysozyme (T4L), which is a frequently studied model system (Baase et al., 2010), we recently found a new transiently populated conformational state (hidden state) within

its catalytic cycle. Using overall 24 FRET pairs and the collected structural knowledge of the protein database, with more than 500 crystal structures, we demonstrated that this newly found state is indeed unknown. So far, this state was overlooked as it is lowly populated under equilibrium conditions. Experimentally we established this previously hidden state by multiparameter fluorescence detection (MFD) single-molecule measurements and ensemble time correlated single photon counting (eTCSPC) (see Figure 1). In MFD the detected fluorescence photons are characterized by the excitation wavelength, a spectral emission window and their polarization and the time of their detection which is recorded with picosecond resolution. The information provided by the fluorescence photons is recorded without data reduction. Therefore, at least 8 dimensions of fluorescence can be explored with unsurpassed flexibility: fluorescence intensities, fluorescence lifetimes, anisotropies, distances via FRET, stoichiometries, the excitation windows, the emission window and the fluorescence quantum yields (Förster, 1951, Kühnemuth and Seidel, 2001, Widengren et al., 2006). To process this wealth of information we used various analysis techniques such as lifetime filtered correlation spectroscopy, time-resolved fluorescence intensity histograms and single-molecule fluorescence intensity histograms. As under single-molecule conditions no sophisticated strategies to synchronize kinetics are needed, we were able to track the dynamics of T4L over seven orders of magnitude by filtered FCS (fFCS) and by inspecting multi-dimension FRET-efficiency histograms (as described in chapter D).

Later, we made use of the rich set of high quality eTCSPC measurements to determine structural models. We achieve this by combining the fluorescence measurements with molecular simulations and structural pre-knowledge as provided by the protein database. Clearly, even in comparably simple systems such as T4L the degrees of freedom, determined by the number of atoms, tremendously exceeds the number of measurements. Thus, structures determined by FRET are structural models and unify knowledge provided by the simulations and experimental facts. We consider a suggested structural model as good, if it combines the experimental facts with state of the art molecular simulations. We obtain such models by welding experimental results, molecular simulations and fluorescence models tightly together. Ideally, experimental results and molecular models are combined by Bayesian approaches using direct forward modelling of the experimental observables for given structures (Rieping et al., 2005). Unfortunately, the behavior of dyes attached to a biomolecule cannot be predicted with sufficient accuracy. Hence, direct forward modelling of fluorescence observables of dyes attached to proteins is currently impossible. Therefore, we bind the experimental data, fluorescence models, and molecular simulations tightly together by a bi-directional approach: on one side we predict structural models

for a given set of experiments which are interpreted by a fluorescence model; on the other side we predict the fluorescence observables for a given structural model. As the direction from structural models to fluorescence observables is less accurate we solely validate or falsify the used fluorescence model. This approach allows us to cross-validate our fluorescence models which were used in first place to interpret the experimental data in terms of distances. Using this approach, we are able to determine quantitative molecular models for a given set of experiments.

In T4L we solve this multi-factorial problem to determine structural models by FRET-constraints using unnatural amino-acids (UAAs), global analysis of all 24 fluorescence decays, FRET guided molecular dynamics simulations, the estimation of fluorescence properties of the dyes for the found conformational states and cross validation of the structural model against the fluorescence observables. The use of UAAs allows for different labeling chemistry of the donor- and the acceptor-fluorophore. This guarantees for ideal reference samples of high quality. This strategy additionally allowed us to heed specific fluorophore effects in FRET measurements. The global analysis of all fluorescence intensity decays improves the discrimination of the conformational states and allows to assign distances states by their amplitudes. The FRET guided molecular dynamics simulations drive a known conformer into the new conformational states within reasonable simulation times, even though the lifetime of the conformational states is in the microsecond to millisecond regime. The estimation of the position and conformer specific fluorescence properties of the dyes by simulating dynamic quenching guarantees that the model to interpret the fluorescence intensity decays was correct. The cross-validation of the predicted FRET-distances against the experimental distances by analysis detects systematic deviations for the proposed structural model. The already known crystal structures serve as a benchmark for the accuracy of the structural models. Additionally, the size of T4L (~20 kDa) allows for fast molecular simulations, testing for various simulation settings. The small number of relevant principle components in its major states (de Groot et al., 1998) allows for experimental oversampling. In that sense T4L is an ideal benchmark system to establish methodologies for the determination of structural models of protein and hidden-states.

This chapter is based on the manuscript “Dynamics and function of transient states in single enzyme molecules” and the manuscript “FRET-restrained structural modelling resolves an excited state in the catalytic cycle of T4L”. In the first manuscript “Dynamics and function of transient states in single enzyme molecules” we found a new conformer of T4L in addition to the two known major conformers:  $C_I$ , a conformation open at various degrees in the absence of

substrate (Goto et al., 2001, Zhang et al., 1995), and  $C_2$  a conformation adopted when the substrate is bound (Kuroki et al., 1993). Additionally, we found a previously hidden state ( $C_3$ ), which might be related to the release of the substrate. The existence and function of  $C_3$  was shown by interpreting 2-dimensional FRET-efficiency histograms by FRET-lines (see Figure 1 and Figure 2A/B), mutational analysis of the catalytic residues, filtered FCS (fFCS), kinetic simulations of FRET-efficiency histograms and ensemble TCSPC measurements. Additionally, the  $C_3$  state was corroborated by the use of other biophysical tools, chromatography, and EPR spectroscopy (McHaourab et al., 1996). In the second manuscript FRET based structural models of  $C_1$ ,  $C_2$  and  $C_3$  were generated.

In the first manuscript “Dynamics and function of transient states in single enzyme molecules” a multitude of analysis techniques has been used. The fluorescence decay analysis in combination with the mutational analysis clearly demonstrates the existence and function of the new “hidden” state  $C_3$ . Therefore, for simplicity other experimental findings are omitted. The fluorescence decay models of both manuscripts are presented to stress the differences.

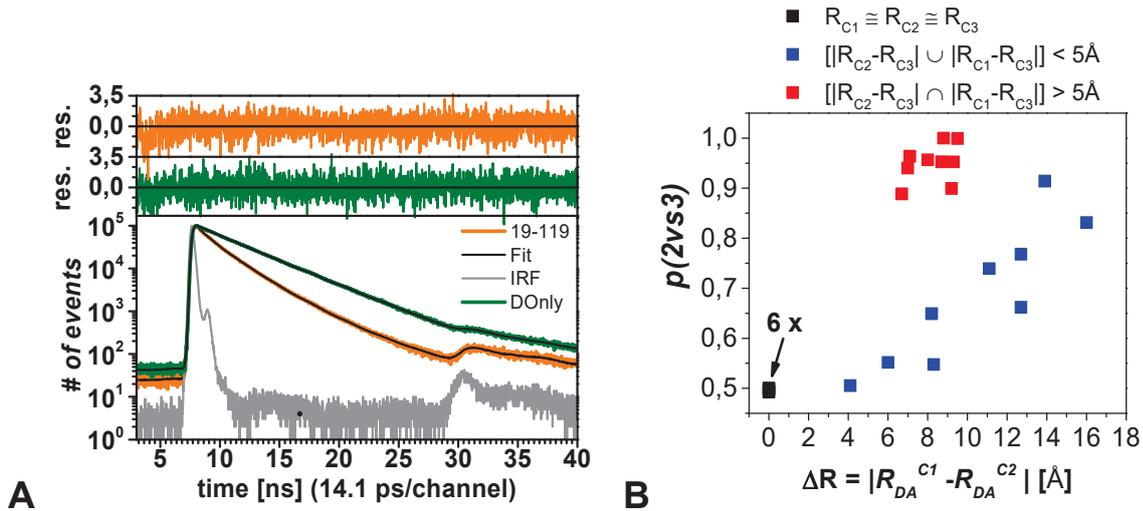
## 2 Identification of a hidden state

### 2.1 Formal fluorescence decay FRET-analysis

To minimize the bias imposed by the fluorescence decay model we first used a formal description of the experimental the fluorescence decays (see Fig. 2A). We deliberately did not include structural knowledge on T4L or linker-distributions and analyzed the data using discrete FRET-rate constants by the method of FRET-induced donor decays (see chapter A). We described the FRET-induced donor decay by a multi-exponential function given by:

$$\varepsilon_D(t) = \sum_{j=1}^N x_{RET}^{(j)} \exp(-t \cdot k_{RET}^{(j)}) + x_{DOnly} \quad (1)$$

Here,  $x_{RET}^{(j)}$  is the species fraction of the FRET-rate constant  $k_{RET}^{(j)}$  and  $x_{DOnly}$  is the fraction of FRET-inactive molecules, which presumably lack an acceptor fluorophore. Here,  $N$  gives the number of components. We call this model a two-state model if  $N=2$ , and the three-state model corresponds to  $N=3$ .



**Figure 2.** (A) Time-resolved fluorescence decay for K19pAcF/R119C-(DA) (orange) respective Donor-sample (green) fitted with three states model (black line), respectively, with the following values for  $R_{DAi}$  and (species fraction  $\xi$ ): 55.0 Å (0.46), 46.3 Å (0.39), 35.2 Å (0.15). Residuals are shown on top. (B) All variants were fitted with the two and three states model. Here, the p-value  $p(2vs3)$  calculated from the F-value  $\chi_{r-rel}^2 = \chi_r^2(2\text{ states}) / \chi_r^2(3\text{ states})$  is plotted against the absolute difference in distances  $C_1$  and  $C_2$  as obtained from the three states model. For six variants, the distance of the states  $C_1$  and  $C_2$  cannot be distinguished (black, arrow).

To describe the fluorescence decay of the donor in the absence of FRET  $f_{D|D}^{D0}(t)$  a multi-exponential relaxation model was necessary:

$$f_{D|D}^{D0}(t) = \sum_i x_{D(0)}^{(i)} \exp(-t / \tau_{D(0)}^{(i)}) \quad (2)$$

Here  $\tau_{D(0)}^{(i)}$  is the donor fluorescence lifetime and  $x_{D(0)}^{(i)}$  are the pre-exponential factors. We assumed that all donor species are quenched by the same distribution of FRET-rate constants. This is true if quenching does not change the donor radiative lifetime and the spectral overlap and donor quenching by the local environment is independent of FRET.

Then the fluorescence intensity decay  $f_{D|D}^{DA}(t)$  of the donor in presence of an acceptor are given as the direct product of equation (1) and (2):

$$\begin{aligned} f_{D|D}^{DA}(t) &= \varepsilon(t) \cdot f_{D|D}^{D0}(t) \\ &= x_{DOnly} f_{D|D}^{D0}(t) + \sum_i \sum_j x_{D(0)}^{(i)} x_{RET}^{(j)} \exp\left(-t \cdot \left(\frac{1}{\tau_{D(0)}^{(i)}} + k_{RET}^{(j)}\right)\right) \end{aligned} \quad (3)$$

This formal description allows us to capture feature of decays and monitor differences in FRET among different decays. For convenience the FRET-rate constants  $k_{RET}$  were expressed as apparent distances:

$$k_{RET}^{(j)} = \frac{1}{\tau_0} \cdot \left( \frac{R_0}{R_{DA,app}^{(j)}} \right)^6 \quad (4)$$

Here,  $R_0$  is the Förster radius and for the Alexa488 and Alexa647 FRET pair is 52 Å by assuming  $\kappa^2$  is 2/3, which is justified by the anisotropy studies,  $\tau_0 = 1/(k_{ISC} + k_{IC} + k_F)$  is the lifetime of the donor that corresponds to the quantum yield of the donor employed in the calculation of the Förster-radius.

## 2.2 Detection of a hidden state

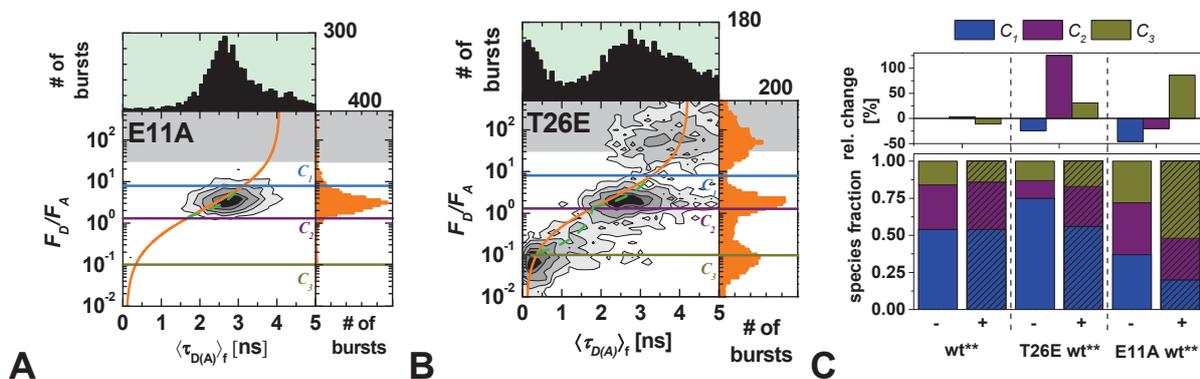
First, we unsuccessfully tried to describe the FRET-induced donor decay by single FRET-rate constants. Next, we used 2- and 3 FRET-rate constant models to consistently describe the FRET-induced donor decays. Clearly, with increasing number of free parameters the quality of the fit increases. To test whether the improvement in  $\chi_r^2$  of the three states model is statistically significant, we postulated the null hypothesis that a two states model is sufficient to describe our data and calculate the F-value given by the ratio  $\chi_r^2(2 \text{ states}) / \chi_r^2(3 \text{ states})$  and calculate the corresponding p-values  $p(2 \text{ vs } 3)$ . This analysis provided the dazzling result that sometimes 2-FRET rate constants are sufficient to describe the data and sometimes a third FRET-rate constant improves the quality of the fit significantly. Next, using the obtained donor-acceptor distances we determine the apparent distance differences  $\Delta R = |R_{DA}^{(C1)} - R_{DA}^{(C2)}|$  and plot this differences against the p-values (see Fig. 2B). In this plot we recognize that the measurements can be grouped in three distinct classes: (1) a class where an additional FRET-rate constant does not improve the quality of the fit and the distance difference  $\Delta R$  between  $C_1$  and  $C_2$  is very small; (2) a class where the fit quality improves with the distance difference  $\Delta R$ ; (3) a class where the improvement in the fit quality is higher than anticipated in comparison to class (2).

To rationalize this unexpected grouping, we assign the FRET-rate constants and the associated apparent distances to the protein in its open conformation  $C_1$ , the supposable catalytic active closed conformation  $C_2$ , or an unknown conformation  $C_3$ . Using this assignment, the grouping can be understood if the differences of the fitted apparent distances among  $C_1$ ,  $C_2$  and  $C_3$  are considered. If two out of the three apparent distances are very similar they are indistinguishable and an additional third FRET-rate constant does not improve the quality of the fit (Fig. 2B, black points). If  $C_1$  and  $C_2$  are distinguishable but the distance difference of  $C_1$  or  $C_2$  to  $C_3$  is small the p-values increase with increasing distance between  $C_1$  and  $C_2$  ( $\Delta R < 5$  Å, blue points). If all three conformers are well resolved by FRET we obtain a group where a third FRET-rate constant improves the quality of the fit more than anticipated ( $\Delta R > 5$  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 is clearly smaller than the F-value of our data (1.046), thus we need at least a three FRET-rate constants to describe our data.

### 2.3 Biochemical trapping

Next, we decided to look deeper into the functional implication of the existence of  $C_3$ . Thus we separately mutated residues E11 and T26 at the active site using the backbone of the S44pAcF/I150C variant. Both additional mutations are known to alter the catalytic activity of T4L (Shoichet et al., 1995, Kuroki et al., 1993) and should help to identify the role of the  $C_3$  state during enzyme catalysis. The E11A mutation is known to inactivate T4L (Shoichet et al., 1995). Moreover, recently it was shown that the variant E11A could bind the substrate but it is unable to catalyze it (Hu and Lu, 2004).



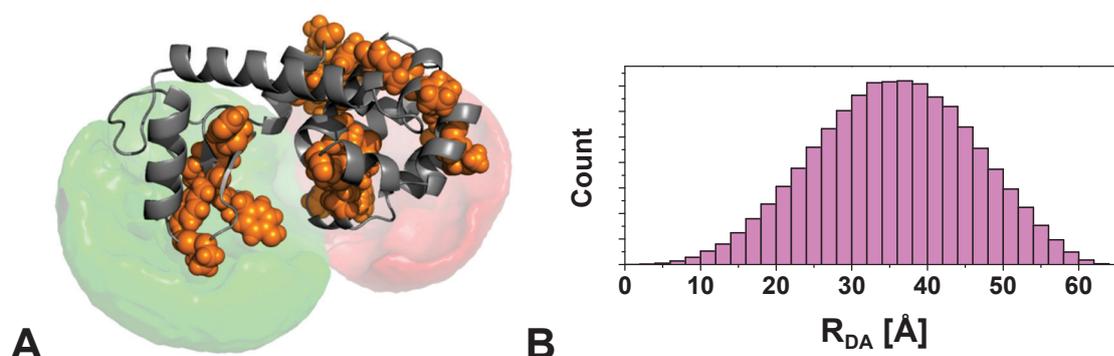
**Figure 3.** (A,B) Two dimensional  $F_D/F_A$  vs.  $\langle \tau_{D(A)} \rangle_f$  MFD histograms of T4L incubated with peptidoglycan with the donor and acceptor attached to the amino acids 44 and 150, respectively (A) E11A mutant (B) T26E mutant (C) Species-fractions of the variants S44pAcF/I150C-(DA), E11A/S44C/I150C-(DA) and T26E/S44pAcF/I150C-(DA) used to mimic free enzyme as determined by ensemble TCSPC measurements. Upon addition of peptidoglycan (+), the fractions of states of the functional variants are shifted towards  $C_2$  and  $C_3$ , respectively (shaded bars). On top, the relative change in fractions upon addition of peptidoglycan is shown.

In the E11A variant the high-FRET state is not populated upon addition of the substrate (see Fig. 3A). The mutant T26E is the mutant used in the crystallization of the adduct form of T4L (Kuroki et al., 1993). In T26E we observe a significant increase in the population of the high-FRET state in the single-molecule measurements as well as in the ensemble TCSPC measurements (see Fig. 3B/C). This is the first hint that the high-FRET state is indeed a functional state that is involved in the catalytic cycle of T4L. As in the E11A mutant we find an increase of the  $C_2$ -state upon substrate addition, we interpreted the anticipated  $C_2$ -state as catalytic active state, while we interpret the high-FRET as  $C_3$ -state or product release state. We find that  $C_3$  is preferentially populated after substrate hydrolysis. Nonetheless, one cannot compare the changes of the species fractions directly because each of these variants in the absence of substrate shows slightly different equilibrium conditions. Therefore, instead of looking at absolute changes of

the species fractions we consider their mutant specific relative changes upon substrate addition (Fig. 3C).

## 2.4 Limitations of formal FRET-analysis

The simple formal model of the fluorescence intensity decays in presence of FRET by a number of discrete FRET-rate constants, as described above, was successfully applied to identify the previously hidden  $C_3$ -state. However, the potential of this formal description for structural modelling is limited due to a number of reasons: (1) The fluorophores are tethered to the protein by long flexible linkers. Hence, distance distributions are anticipated but not explicitly accounted for. (2) The assignment of distance restraints to conformational states is not automated and thus maybe biased. (3) Quenching of the donor dye by the local environment was assumed to be independent of FRET. This assumption was not validated.



**Figure 4.** Accessible volume (AV) of the donor (Alexa488) and the acceptor (Alexa647) at position K19pAcF and I150C, respectively (**A**) the AV of the donor and the acceptor are shown as green and red surface, respectively. The amino-acids methionine, tryptophan, tyrosine, histidine and proline, which are known to quench Alexa488 dynamically, are shown as orange spheres (**B**) Calculated distance distribution position between the two labels given their accessible volumes with a mean distance  $\langle R_{DA} \rangle = 35.6$  Å and a distribution width of  $\sigma_{DA} = 10.3$  Å.

For accurate structural modelling the spatial distribution of the fluorophores has to be considered (Kalinin et al., 2012). As shown in Fig. 4A the conformational space, approximated by accessible volume (AV) calculations of the donor as well as the acceptor, is considerable. As the dye diffusion of fluorophores is relatively slow compared to the fluorescence lifetime of the donor dye, a distribution of distances (see Fig. 4B) is expected (compare chapter A). If the distance distributions of the fluorophores are not explicitly accounted for, the recovered apparent distances may be biased and not necessary represent the average donor-acceptor distance. Hence, for structural modelling we prefer to use a fluorescence decay model with distributed donor-acceptor distances instead of a formal description of the experimental fluorescence decay.

In the analysis scheme as described above, the assignment of FRET-rate constants to states was based on the pre-knowledge of the already known crystal structures of  $C_1$  and  $C_2$ . In T4L we consider this approach reasonable, as already multiple structural studies have been performed

which confirmed the existence of  $C_1$  and  $C_2$ . However, this approach lacks rigor to be generally applicable. Furthermore, in some cases, for instance if two of the states are very similar, this approach may be susceptible to personal bias.

Finally, in this formal analysis it was assumed that the donor quenching by FRET and the quenching by the local environment are independent. As shown in chapter This approximation is valid for flexible coupled dyes which are dynamically quenched by their local environment *within* their linker-length. In T4L, however, the local environment of the dyes in the  $C_1$ -,  $C_2$ -, and  $C_3$ -state might differ. The major-states  $C_1$  and  $C_2$  are probably also major states in the donor only reference samples. Therefore, the conformer specific effects on the recovered distances of  $C_1$ , and  $C_2$  are probably small.  $C_3$ , however, is a minor-state. Hence, the contribution of  $C_3$  to the fluorescence intensity of the donor in absence of FRET is also small. Therefore, the recovered distance may deviate significantly from the correct donor-acceptor distance (compare Fig. 6, chapter A).

### 3 Refined fluorescence tools for structure validation

#### 3.1 Global analysis of fluorescence decays

In chapter A it is demonstrated that the local quenching of the donor *within* the conformational space of a dye attached to a protein can be well approximated by factorizing the effect of FRET from the effect of donor quenching:

$$f_{D|D}^{DA}(t) = \varepsilon_D(t) \cdot f_{D|D}^{D0}(t) \quad \text{with} \quad \varepsilon_D(t) = \int_0^{\infty} p(R_{DA}) \cdot \exp\left[-\frac{1}{\tau_0} \left(\frac{R_0}{R_{DA}}\right)^6 \cdot t\right] dR_{DA} \quad (5)$$

It was not shown that this factorization is independent of the conformational state of the protein. However, different conformational states of T4L might lead to different exposure of quenching aromatic amino-acids (shown in Fig. 4A as orange spheres) and a different local environment of the donor dye and hence, a different fluorescence decay of the donor in absence of FRET. Thus, a more complicated analysis scheme as outlined in Fig. 5 in chapter B might be applicable, where every conformational state is associated to its FRET- and donor-state by a state specific FRET-rate constant distribution and donor lifetime distribution. For now, we assume that the approximation given by the factorization provides the correct results.

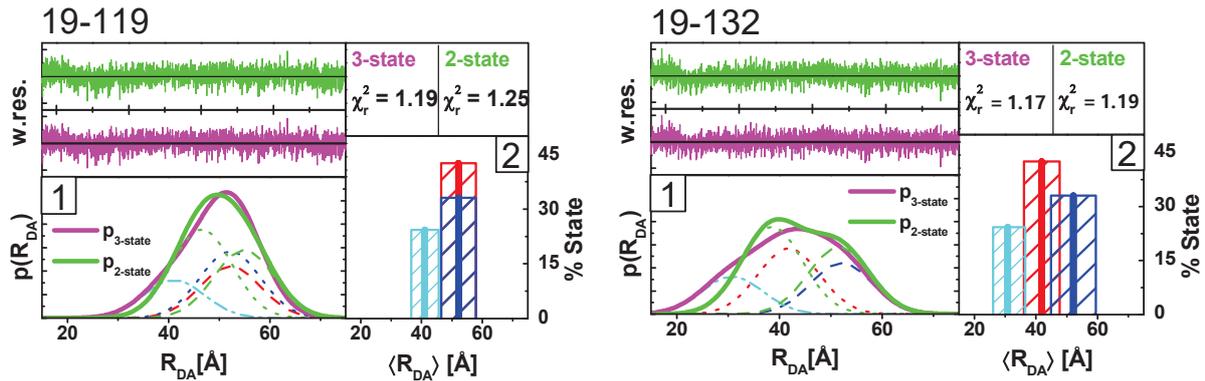
The distance distribution given in equation (5) is then simply given by the superposition of the donor-acceptor distance distributions of the  $N$  states:

$$p(R_{DA}) = \sum_{j=1}^N x^{(j)} \cdot p^{(j)}(R_{DA}) \quad (6)$$

Here,  $x^{(j)}$  are the species fractions, with the associated probability distributions of the distances  $p^{(j)}(R_{DA})$  of the state ( $j$ ). We approximate the distance distributions of the  $N$  states by normal distributions:

$$p^{(j)}(R_{DA}) = \frac{1}{w_{DA} \sqrt{\pi/2}} \exp\left(-2 \left[\frac{R_{DA} - \bar{R}_{DA}^{(j)}}{w_{DA}}\right]^2\right) \quad (7)$$

Herein,  $w_{DA}$  is the width of the distributions and  $\bar{R}_{DA}^{(j)}$  is the center of the distribution of state ( $j$ ). The width of this normal distribution  $w_{DA}$  should not be misinterpreted as the experimental/statistical-error but it describes a real physical property of the donor-acceptor pair. We assumed independent of the separation distances a global width of 12 Å. This is in agreement with typical widths observed by fluorescence decay analysis of such fluorophores (Sindbert et al., 2011, Hofig et al., 2014).



**Figure 5.** Comparison of the fit-result of a 3-state and a 2-state model with normal distributed distances (the width was fixed to 12 Å). Shown in (1) as dashed lines are the components of the 3-, and 2-state fit the solid green line and solid magenta line represent the total distance distribution given by the sum of the individual components. In (2) the mean shown as solid lines and the errors shown as shaded boxes of the mean of the Gaussians of the 3-state fit are shown. In top the reduced sum of the weighted deviations is shown. The weighted residuals of the datasets are shown in the top left. The shown dataset corresponds to two datasets out of 24 datasets which were globally analyzed.

Instead of FRET-rate constants and fractions in this analysis we determine  $x^{(j)}$ , the fractions of the states, and  $\bar{R}_{DA}^{(j)}$ , the center of the donor-acceptor distance distributions. Hence, compared to the previous analysis the number of free parameters is identical. Analogous to the previous analysis we fit the experimental decays using a 2-state and a 3-state model. However, we further restrict the parameter space by using global fractions among all FRET-states. Overall 24 FRET-samples and 7 donor-references were globally analyzed. After optimization of the free parameters using the Levenberg Marquardt algorithm the statistical uncertainties were determined using an effective invariant ensemble sampler for Markov chain Monte Carlo (Goodman and

Weare, 2010, Foreman-Mackey et al., 2013). By this analysis a combination of distances with corresponding errors and fractions was obtained.

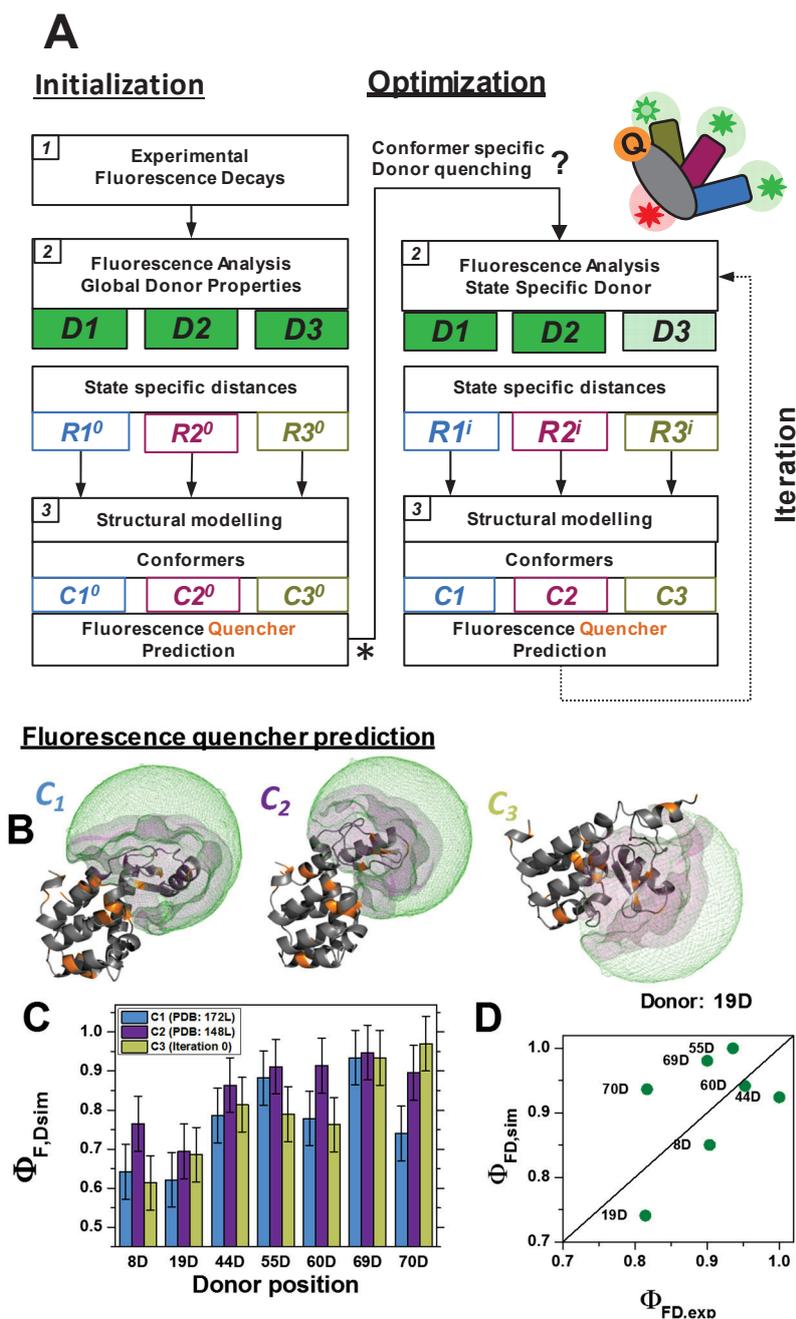
This global analysis approach has numerous advantages. As we assume a reasonable width of the donor-acceptor distance distributions we were able to resolve overlapping distributions of the individual states. The global analysis is able to assign states to samples even in cases where the distances among two states are identical. This is possible as the global analysis implicitly utilizes the information contained in the relative amplitudes. Hence, structural pre-knowledge is not needed to assign distance constraints to states, as the assignment is solely based on the amplitudes. This exemplified in the comparison of the samples 19-119 and 19-132 in Fig. 5. In the FRET-pair 19-119 effectively only two distances are observed. However, in other FRET-pairs, for instance 19-132, three states with distinct distances with corresponding amplitudes are found. Hence, despite in 19-119 apparently only two distances are found, one of the two distances is given by the superposition of two states. This example can be better understood using PDB:172L and PDB:148L as a reference for  $C_1$  and  $C_2$ , respectively. In the FRET-pair 19-119 average distances of 50.8 Å and 47.4 Å are obtained by accessible volume calculations for  $C_1$  and  $C_2$ , respectively. Within the statistical error of the data these two distances are indistinguishable. Hence, two distance distributions, one for  $C_1/C_2$  and one for  $C_3$ , are effectively enough to describe the data. However, in a dataset where all three states are distinguishable the amplitudes are determined. Thus, as all parameters are mutually dependent, the global analysis stabilizes the fit-result and allows assignment of states to distances.

### 3.2 Considering FRET and fluorescence quenching in structural modelling

In the analysis it is initially assumed that the FRET-induced donor decay  $\varepsilon_D(t)$ , which is simply given by the ratios of the normalized fluorescence decays in presence and absence of FRET, provides the exact distance distribution by equation (5). This assumption ignores possible correlations between the FRET-rate constant distribution and the quenching of the donor by the local environment. To obtain accurate distances by analysis of the time-resolved fluorescence it is vital to consider such correlations in particular for minor states (see chapter B, Fig. 6). Without taking structural models of the identified states into account an alternative model, which considers species specific quenching, cannot be constructed. In Fig. 6 an iterative scheme is depicted that can be applied to determine structural models by fluorescence decays. It allows for state specific fluorescence decays models by combining structural models derived by FRET-restraints and simulations of the donor fluorescence. Currently, donor fluorescence is simulated using the quenching and traps accessible volume simulations (qtAV) as introduced in chapter A.

Following this scheme, it initially is assumed that all donor lifetime components are quenched by the same FRET-rate distribution. This is a reasonable assumption as initially little structural knowledge is available. Next, by global analysis of all measurements a set of distance constraints is determined (see Fig. 6). The global analysis groups distances by amplitudes. Hence, a manual assignment of the distances to states is not necessary. Using these distance constraints, a set of structural models of all conformers is determined. This set of structural models is used to test the initial assumption whether the donor fluorescence properties are indeed independent of the conformational states. This is accomplished by simulating the fluorescence properties of the donor of each conformer (see Fig. 6B). Simulations of donor fluorescence properties and quenching by the local environment were briefly outlined in chapter B. In case differences between are detected, the initial model that all fluorescence states are equally quenched by FRET is adjusted. For instance, if the fluorescence of the donor in one conformational state is stronger quenched by the local environment short lifetime components are attributed to that conformational state. Using an adjusted fluorescence model, a new set of distances is obtained which is again used for structural modelling.

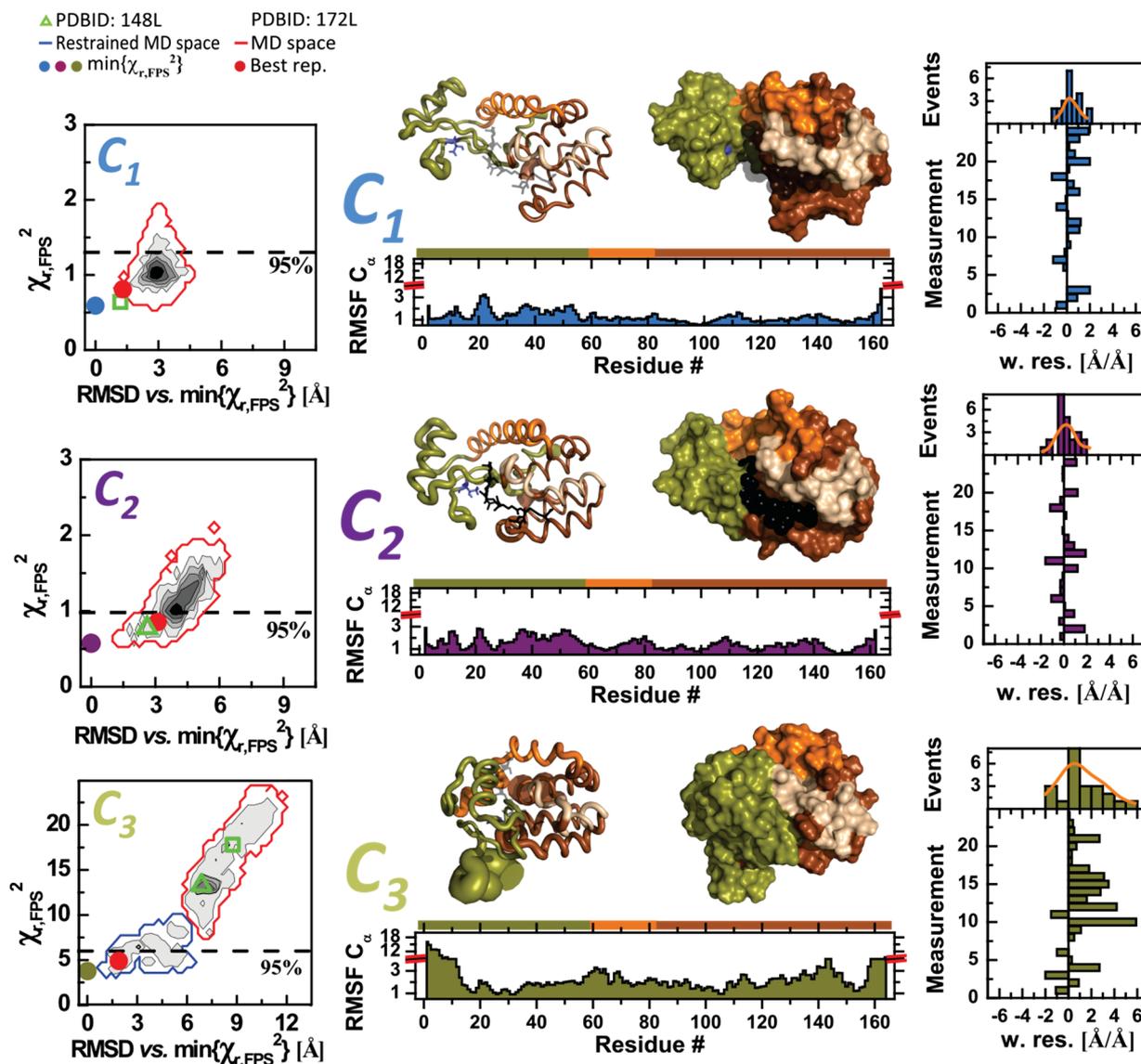
In T4L no final structural model was determined so far. However, the crystal structures of T4L in the  $C_1$ - and  $C_2$ -state (PDB-ID: 172L, 148L) and a preliminary structural model of the  $C_3$ -state are used to test the assumption that all three states are equally quenched (Fig. 6B). After simulating the donor fluorescence properties little differences in the donor fluorescence among these structural models of conformational states were found (see Fig. 6C). Hence, no additional optimization of the fluorescence model is needed and the initial assumption that all donor states are equally quenched by FRET is valid.



**Figure 6.** Analysis scheme for the determination of T4L-structures by FRET (A) Flow-chart describing the applied analysis scheme of the time-resolved fluorescence data. (B) Found conformers after initialization (highlighted by \*), the quenching amino-acids are shown in orange, in green the accessible volume of the donor at the given position is outlined, magenta are parts of the accessible volume in which the donor diffuses slowly. (C) Comparison of the simulated fluorescence quantum yields of all donor positions for the three conformational states shown, the error bars are estimated errors given the simulations as shown in Chapter B (D) Experimentally determined fluorescence quantum yield of the fluorescent species for different donor positions on T4L in comparison to the simulated fluorescence quantum yields; the black line shows a 1:1 correlation. For better comparison the fluorescence quantum yields of the donor were normalized by the maximum fluorescence quantum yield determined by each method.

### 3.3 Structural model validation

FRET-guided molecular dynamics simulations are followed by a statistical analysis of the conformations given the experimental distance constraints and their uncertainties. We call this procedure “screening” as we select structural models which fulfil the experimental constraints given the generated conformational ensemble. Preliminary screening results of the  $C_1$ ,  $C_2$  and  $C_3$  state are shown in Fig. 7.



**Figure 7.** Preliminary screening results and comparison with experimental distances. The three conformers have the following color codes: blue for  $C_1$ , purple for  $C_2$ , dark yellow for  $C_3$ . We selected our states according to the confidence interval of  $2\sigma$  or 95% (left). Cluster analysis is presented as  $\chi_{r,FPS}^2$  vs. RMSD( $C_\alpha$ ) plots. The structure with minimum  $\chi_{r,FPS}^2$  was taken as the reference structure for RMSD calculations; the best representative for each conformation is shown as solid red circle. Conformational space sampled by the unrestrained and restrained MD simulations is shown as red and blue contours, respectively. The 95% confidence interval for the precision of the proposed FRET derived models is shown as dashed line. The best representatives for each conformation are given in surface representation and as “putty” cartoons, where the thickness of the backbone representation corresponds to RMSFs within the confidence interval. The NTsD and CTsD of T4L are colored in olive and brown, respectively. Models are aligned such that CTsD overlay optimally. Substrate is shown in solid black for  $C_2$ , where it is known from crystal structure, and overlaid as semi-transparent for visual reference in  $C_1$  and  $C_3$ . Helix c is colored in orange, and helices f and g in wheat. Stick representation is used for Thr 26. The agreement of structural models with their specific FRET restraints is judged by weighted residuals.

To illustrate the sampled conformational space, the best structural model is taken as a reference to calculate the RMSD against all other structures. Next the reduced  $\chi^2$  is plotted vs. the RMSD and a histogram is constructed. To reduce the influence of a potential sampling bias in the MD trajectories on the structures, they were clustered according to their RMSD using the UPGMA method implemented in fastcluster (Müllner, 2013). From the dendrogram, 1600, 800, and 400 clusters were selected by choosing appropriate merging levels.

The representatives of these clusters (structure with minimal RMSD from the mean of the cluster) were used to estimate the accuracy and precision of the structures for  $C_1$ ,  $C_2$ , and  $C_3$  reported in Fig. 7. These structures were chosen from the MD-generated ensemble such that they are closest in RMSD to the weighted average atomic coordinates  $\langle \vec{r}_j \rangle$  of the cluster representatives  $i$  below a maximum allowed threshold  $\chi_{r,FPS,max}^2$  corresponding to a confidence level of 95%:

$$\langle \vec{r}_j \rangle = \sum_{i=1}^N w_i \cdot \vec{r}_{i,j} \quad (8)$$

$N$  is the number of cluster representatives below the threshold,  $w_i$  are the weighting factors, and  $\langle \vec{r}_j \rangle$  are the average atomic coordinates of atom  $j$ . The weighting factors  $w_i$  were calculated based on the probability by which the cluster representative is distinguishable from the best structure according to FRET. We calculated the degrees of freedom (#d.o.f.  $\sim$  #measurements - #normal modes = 24 - 5 = 19) by using the normal modes computed by NMSim. Given the determined  $\chi_{r,FPS}^2$ , the minimum  $\chi_{r,FPS,min}^2$ , and the degrees of freedom (#d.o.f = 19), the weighting factors were calculated according to:

$$w_i = \frac{v_i}{\sum_i v_i} \quad \text{with} \quad v_i = F(\chi^2 / \chi_{min}^2, \text{dof}) \quad (9)$$

where  $F(\chi^2 / \chi_{min}^2, \text{dof}) = F(f, m, n)$  is the F-distribution with the f-value  $f = \frac{\chi_m^2 / m}{\chi_n^2 / n}$ , and  $m = n = 19$  are the corresponding degrees of freedom.

To illustrate systematic deviations of the found conformational states from the experimental data, the distances predicted for the structural model, which is in best agreement with the experimental data, is compared with the fitted experimental distances. To consider experimental errors, the weighted residuals of these distances are used. This is shown in Fig. 7 for  $C_1$ ,  $C_2$  and  $C_3$ . For  $C_1$  and  $C_2$  we find a good agreement between our structural model and the experimental data. As the current structural model for  $C_3$  still deviates systematically from the experiment we consider the results as presented in Fig. 7 as preliminary results.

## 4 Conclusion

It was demonstrated how a model free approach that uses a formal description of the FRET-induced donor decay can be utilized to identify a formerly unknown conformational state in T4 lysozyme. This approach uncovered the previously hidden  $C_3$  state by a network of FRET-pairs

and the relative changes of amplitudes upon addition of substrate. The functional states of the protein were assigned based on mutational analysis and the structural pre-knowledge as provided by the protein database.

To determine a structural model with minimum bias by structural pre-knowledge which explicitly considers the donor-acceptor distance distribution of the individual states a refined fluorescence decay model was used. By treating all measurements simultaneously by global-analysis of all recorded fluorescence decays, the assignment of distances to states was based on the recovered fitted species fractions. By unifying all data-sets and effectively oversampling we compensated for variability and uncertainty associated with single pair FRET-experiments. The well-defined statistics of the fluorescence decays allowed us to determine statistical uncertainties and assure the existence of  $C_3$ -state. We used a bi-directional/iterative approach in which first the experiments are interpreted by a fluorescence model to determine constraints for FRET-guided simulations and later the results of the simulations are used to validate the fluorescence model. This gives us confidence in the constraints which were used to guide the simulations. However, at current stage the simulated structure determined by FRET-guided molecular dynamics simulations systematically deviates from the experimental constraints and is still under revision.

In future the developed methodology may be applied to more complex systems. Additionally, our efforts in correlating the dynamic quenching of the donor fluorescence with the local environment might lead to refined models of fluorophores coupled to proteins. Currently, fluorescence decays are analyzed in terms of fluorescence decay models and the result of that analysis is compared to a structural model. Once the behavior of the fluorophores attached to proteins is better understood, direct forward modelling is possible. This brings structural models closer to fluorescence decays and could allow for more rigorously determined structural models of higher quality by FRET.

## References

- BAASE, W. A., LIU, L., TRONRUD, D. E. & MATTHEWS, B. W. 2010. Lessons from the lysozyme of phage T4. *Protein Sci*, 19, 631-41.
- CHUNG, H. S., LOUIS, J. M. & EATON, W. A. 2009. Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 11837-11844.
- DE GROOT, B. L., HAYWARD, S., VAN AALTEN, D. M., AMADEI, A. & BERENDSEN, H. J. 1998. Domain motions in bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data. *Proteins*, 31, 116-27.

- FOREMAN-MACKEY, D., HOGG, D. W., LANG, D. & GOODMAN, J. 2013. emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*, 125, 306-312.
- FÖRSTER, T. 1951. *Fluoreszenz organischer Verbindungen*, Göttingen, Germany, Vandenhoeck & Ruprecht.
- GOODMAN, J. & WEARE, J. 2010. Ensemble Samplers with Affine Invariance. *Communications in Applied Mathematics and Computational Science*, 5, 65-80.
- GOTO, N. K., SKRYNNIKOV, N. R., DAHLQUIST, F. W. & KAY, L. E. 2001. What is the average conformation of bacteriophage T4 lysozyme in solution? A domain orientation study using dipolar couplings measured by solution NMR. *J Mol Biol*, 308, 745-64.
- GROSSMAN, I., YUVAL AVIRAM, H., ARMONY, G., HOROVITZ, A., HOFMANN, H., HARAN, G. & FASS, D. 2015. Single-molecule spectroscopy exposes hidden states in an enzymatic electron relay. *Nat Commun*, 6, 8624.
- HAMMES, G. G. 1964. Mechanism of Enzyme Catalysis. *Nature*, 204, 342-3.
- HENZLER-WILDMAN, K. & KERN, D. 2007. Dynamic personalities of proteins. *Nature*, 450, 964-972.
- HOFIG, H., GABBA, M., POBLETE, S., KEMPE, D. & FITTER, J. 2014. Inter-Dye Distance Distributions Studied by a Combination of Single-Molecule FRET-Filtered Lifetime Measurements and a Weighted Accessible Volume (wAV) Algorithm. *Molecules*, 19, 19269-19291.
- HU, D. & LU, H. P. 2004. Placing single-molecule T4 lysozyme enzymes on a bacterial cell surface: toward probing single-molecule enzymatic reaction in living cells. *Biophys J*, 87, 656-61.
- KALININ, S., FULLE, S., KÜHNEMUTH, R., SINDBERT, S., FELEKYAN, S., GOHLKE, H. & SEIDEL, C. A. M. 2015. Diffusion with traps: experiment, simulation, and theory to describe fluorophore dynamics in biomolecular FRET.
- KALININ, S., PEULEN, T., SINDBERT, S., ROTHWELL, P. J., BERGER, S., RESTLE, T., GOODY, R. S., GOHLKE, H. & SEIDEL, C. A. M. 2012. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nature Methods*, 9, 1218-1225.
- KLECKNER, I. R. & FOSTER, M. P. 2011. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta*, 1814, 942-68.
- KONIG, I., ZARRINE-AFSAR, A., AZNAURYAN, M., SORANNO, A., WUNDERLICH, B., DINGFELDER, F., STUBER, J. C., PLUCKTHUN, A., NETTELS, D. & SCHULER, B. 2015. Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. *Nat Methods*, 12, 773-9.
- KÜHNEMUTH, R. & SEIDEL, C. A. M. 2001. Principles of single molecule multiparameter fluorescence spectroscopy. *Single Molecules*, 2, 251-254.
- KUROKI, R., WEAVER, L. H. & MATTHEWS, B. W. 1993. A covalent enzyme-substrate intermediate with saccharide distortion in a mutant T4 lysozyme. *Science*, 262, 2030-3.
- MCHAOURAB, H. S., LIETZOW, M. A., HIDEG, K. & HUBBELL, W. L. 1996. Motion of spin-labeled side chains in T4 lysozyme. Correlation with protein structure and dynamics. *Biochemistry*, 35, 7692-704.
- MÜLLNER, D. 2013. fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python. *Journal of Statistical Software* 53, 1-18.
- PREUS, S., KILSA, K., MIANNAY, F. A., ALBINSSON, B. & WILHELMSSON, L. M. 2013. FRETmatrix: a general methodology for the simulation and analysis of FRET in nucleic acids. *Nucleic Acids Res*, 41, e18.
- RIEPING, W., HABECK, M. & NILGES, M. 2005. Inferential structure determination. *Science*, 309, 303-306.
- SELKOE, D. J. 2003. Folding proteins in fatal ways. *Nature*, 426, 900-4.

- SHOICHET, B. K., BAASE, W. A., KUROKI, R. & MATTHEWS, B. W. 1995. A relationship between protein stability and protein function. *Proc Natl Acad Sci U S A*, 92, 452-6.
- SINDBERT, S., KALININ, S., NGUYEN, H., KIENZLER, A., CLIMA, L., BANNWARTH, W., APPEL, B., MULLER, S. & SEIDEL, C. A. 2011. Accurate distance determination of nucleic acids via Forster resonance energy transfer: implications of dye linker length and rigidity. *J Am Chem Soc*, 133, 2463-80.
- SMOCK, R. G. & GIERASCH, L. M. 2009. Sending signals dynamically. *Science*, 324, 198-203.
- SUSTARSIC, M. & KAPANIDIS, A. N. 2015. Taking the ruler to the jungle: single-molecule FRET for understanding biomolecular structure and dynamics in live cells. *Curr Opin Struct Biol*, 34, 52-59.
- WIDENGREN, J., KUDRYAVTSEV, V., ANTONIK, M., BERGER, S., GERKEN, M. & SEIDEL, C. A. M. 2006. Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. *Analytical Chemistry*, 78, 2039-50.
- WILHELMSSON, L. M., HOLMEN, A., LINCOLN, P., NIELSEN, P. E. & NORDEN, B. 2001. A highly fluorescent DNA base analogue that forms Watson-Crick base pairs with guanine. *J Am Chem Soc*, 123, 2434-5.
- WOŹNIAK, A. K., SCHRÖDER, G., GRUBMÜLLER, H., SEIDEL, C. A. M. & OESTERHELT, F. 2008. Single molecule FRET measures bends and kinks in DNA. *Proc.Natl.Acad.Sci.USA.*, 105, 18337-18342.
- ZHANG, X. J., WOZNIAK, J. A. & MATTHEWS, B. W. 1995. Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. *J Mol Biol*, 250, 527-52.

---

## Mapping the motions and structure of a state necessary for oligomerization of a large GTPase

**Carola S. Hengstenberg<sup>1†</sup>, Thomas-Otavio Peulen<sup>2†</sup>, Ralf Biehl<sup>3</sup>, Mykola Dimura<sup>2,4</sup>, Alessandro Valeri<sup>2</sup>, Semra Ince<sup>1</sup>, Tobias Vöpel<sup>1</sup>, Bela Faragó<sup>5</sup>, Holger Gohlke<sup>4</sup>, Christian Herrmann<sup>1</sup>, Johann P. Klare<sup>6\*</sup>, Andreas Stadler<sup>3\*</sup>, Claus A. M. Seidel<sup>2\*</sup>**

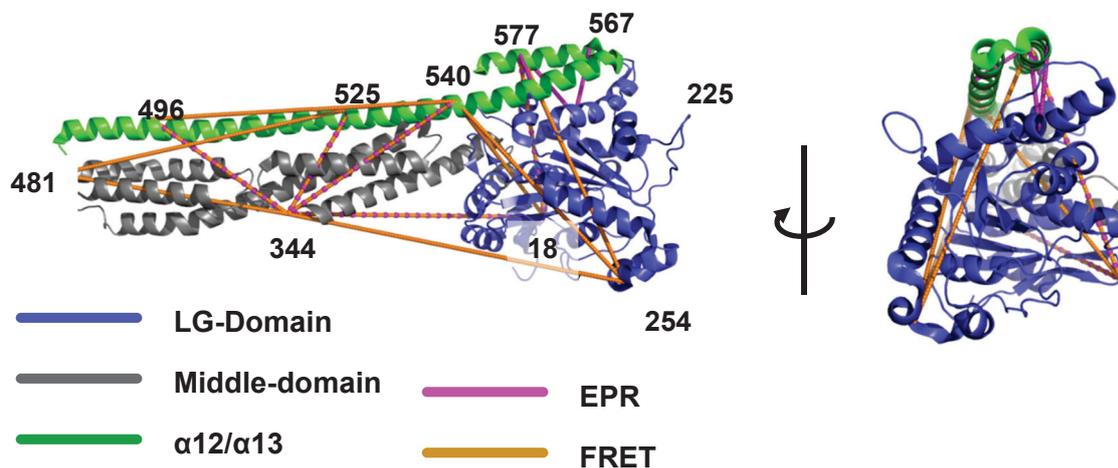
<sup>1</sup>Physical Chemistry I, Faculty of Chemistry and Biochemistry, Ruhr-University Bochum, Universitätsstr. 150, 44780 Bochum, Germany; <sup>2</sup>Chair for Molecular Physical Chemistry, Heinrich-Heine University, Universitätsstr. 1, 40225 Düsseldorf, Germany; <sup>3</sup>Institute of Complex System, Forschungszentrum Jülich, D-52425 Jülich, Germany; <sup>4</sup>Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität, Düsseldorf, Germany; <sup>5</sup>Institut Laue-Langevin, Grenoble, France; <sup>6</sup>Macromolecular Structure Group, Department of Physics, University of Osnabrück, Barbarastr. 7, 49076 Osnabrück, Germany

### 1 Introduction

Deciphering the mechanism that controls conformational changes within proteins is a fundamental issue of molecular biophysics. A deeper understanding of the underlying principles has the potential to open new possibilities for clinical and/or pharmacological applications. A key question is how information within proteins is transduced and propagated via conformational changes over long (and short) distances (Yu and Koshland, 2001). Substrate induced conformational changes are of uttermost importance to control oligomerization and aggregation of proteins in a controlled manner. While unregulated oligomerization, for instance by domain swapping can lead to protein precipitation and even amyloidosis (Bennett et al., 1995). However, the evolutionary and physical mechanisms of oligomerization are very diverse and its general principles have not yet been formulated (Hashimoto et al., 2011).

The members of the dynamin superfamily show nucleotide-dependent oligomerization (Praefcke and McMahon, 2004). Within this superfamily the human guanylate binding protein 1 (hGBP1) is a prime example of the class of large GTPases (GBPs) (Ghosh et al., 2006, Prakash et al., 2000b). In hGBP1 multiple potential interaction sites were identified (Syguda et al., 2012). By FRET- and EPR-experiments we recently found that binding between two spatially separated sites can be induced by addition of the substrate GTP (Vopel et al., 2014) (see Supplement B). This was interpreted as substrate induced conformational changes upon binding to the catalytically active GTPase-domain. Thus homo-dimerization is only possible, if hGBP1 undergoes conformational transitions so that both binding sites come closed to each other (Vopel et al., 2014) (see Supplement B). To draw conclusions on potential oligomerization pathways, a set of questions has to be addressed. A central question is, if conformations, which serve as precursors for dimerization, already exist spontaneously or, if they are induced by the substrate. One potential oligomerization pathway of slow exchanging conformations might differ from the pathway of fast exchanging conformations. Hence, both kinetic and dynamic aspects have to be addressed to be able to understand how distinct

regions of multi-domain protein interact. Recently, encounter complexes of interacting proteins were detected by single-molecule spectroscopy and the formation of an intermediate encounter complex was shown (Gambin et al., 2011).



**Figure 1.** EPR- and FRET-distance network used to probe conformational states of the human guanylate binding protein 1 (hGBP1). hGBP1 can be decomposed into three different domains: the LG-domain (blue), a middle domain (gray) and a helical-domain (green). The amino-acids highlighted by the numbers used for the EPR and FRET measurements are connected by magenta (EPR) and orange (FRET) lines.

Because of its high biological relevance, hGBP1 is a perfect model system to study the potential impact of conformational dynamics for protein function. Early studies reported a contribution of hGBP1 to defense against the vesicular stomatitis virus and the encephalomyocarditis virus (Anderson et al., 1999). Additionally, hGBP1 showed suppressive effects on Hepatitis C virus replication which involved a direct interaction between hGBP1 and the virus (Itsui et al., 2006, Itsui et al., 2009). In an extensive study with mouse GBPs, several GBPs were found to promote oxidative killing and the delivery of antimicrobial peptides to autophagolysosomes (Lee et al., 2011). Recently we found that GBPs are not only promote killing of cellular parasites but directly attack the directly targets the plasma membrane of the cellular parasites via supramolecular complexes (Kravets et al., in revision) (see Supplement C). The image spectroscopic methods which were used to measure equilibrium constants of the complex formation between different GBPs in living cells is presented in more details in chapter D, section 3.3 of this dissertation.

The nucleotide-free and the GppNHp-bound crystal structures of the full length protein show the typical architecture of a dynamin superfamily member (Prakash et al., 2000b, Prakash et al., 2000a). hGBP1 consists of a large GTPase domain (LG domain), a mediating middle

domain and an elongated purely alpha-helical effector domain of the helices  $\alpha_{12}$  and  $\alpha_{13}$  with an overall length of the protein around 120 Ångström (Figure 1). Mutagenesis studies on the guanine cap revealed a coupling of nucleotide binding and guanine cap reorientation (Wehner et al., 2012). So far, hGBP1 is the only member in the dynamin superfamily that converts GTP to GMP via GDP in two subsequent steps (Schwemmle and Staeheli, 1994). It shows similar binding affinities for all three guanine nucleotides. GTP hydrolysis was shown to be fast and cooperative (Praefcke et al., 1999b). Consistently with the other dynamin superfamily members, hGBP1 shows nucleotide-dependent oligomerization (Ghosh et al., 2006, Prakash et al., 2000b, Praefcke and McMahon, 2004). A correlation between homo- and hetero-oligomerization of GBP isoforms and their subcellular localization was described recently (Britzen-Laurent et al., 2010). Upon GTP binding, structural rearrangements enable hGBP1 to self-assemble (Vopel et al., 2009, Vopel et al., 2010).

In a dimer, both binding sites are catalytically active and cleave off phosphate ions from the nucleotide in two successive steps. Inactivation of the catalytic machinery, most probably by dissociation of the hGBP1 dimer, competes with GDP hydrolysis to GMP (Kunzelmann et al., 2006). By mutational analysis, Multiparameter Fluorescence Image Spectroscopy, and live cell microscopy we showed previously that the murine GBPs form multimers (Kravets et al., in revision) and (chapter D).

To enable interaction with various GBP isoforms as well as other binding partners, a high conformational flexibility is a prerequisite for successful homo- and hetero complex formation enabling hGBP1 to fulfill a multitude of biological functions. Based on the available X-ray structures it is not possible to rationalize the required structural rearrangements. By solving possible conformational ensembles of the solution structure, we hope to obtain a better understanding of the function of hGBP1. We combine experimental methods which highlight dynamic features of biomolecules with techniques which resolve spatial features to obtain a detailed picture on hGBP1. We use small-angle X-ray scattering (SAXS, collaboration Andreas Stadler, Ralf Biehl, Forschungszentrum Jülich) as a label free methodology to recover the shape and average distribution of length scales in solution. Additionally, we use electron paramagnetic resonance (EPR, collaboration Johann Klare, University Osnabrück) and Förster resonance energy transfer (FRET) as techniques which provide high resolution information on distance distributions between introduced labels. To relate the recovered distances to structural models, a network of EPR- and FRET-pairs was generated (collaboration Johann Klare) and measured (see Fig. 1). By probing the global structure (SAXS) and distance distributions between labels introduced (FRET, EPR), we increase our

combined resolution and resolve two distinct conformational states. By the use of molecular simulations, we collimate our experimental findings into structural models of two conformers. Furthermore, using the label free neutron spin-echo spectroscopy (NSE, in collaboration with Andreas Stadler and Ralf Biehl) and the lifetime filtered fluorescence correlation spectroscopy (fFCS), we interrogate the conformational dynamics of hGBP1. NSE is highly sensitive for detecting global conformational dynamics up to 300 ns, whereas fFCS quantifies molecular kinetics in the range from  $\mu\text{s}$  to ms, limited by the diffusion time of the molecules (milliseconds). While NSE measurements determine only global dynamics, fFCS allows us to study and quantify intramolecular kinetics between specific positions. This unique combination of experimental methods guarantees to detect all relevant conformational dynamics. By taking advantage of our network of FRET-variants we map the flexibility and determine neuralgic points for the conformational transitions. Using the gathered knowledge on the molecular kinetics and the conformational states we discuss potential implications of protein flexibility in protein aggregation.

## 2 Material and Methods

### 2.1 Protein expression and labeling

The cysteine mutations were introduced into the cysteine free hGBP1 construct in a pQE80L vector (Qiagen, Germany) following the instructions of the QuikChange site-directed mutagenesis kit (Stratagene, USA). The mutagenesis was verified by DNA sequencing with a 3130xl sequencer (Applied Biosystems, USA). New cysteines were introduced at various positions of interest (N18C, Q254C, Q344C, T481C, A496C, Q525C, 540C, Q577C). The hGBP1 was expressed in *E. coli* BL21(DE3) and purified following the protocol described previously (Praefcke et al., 1999a). A Cobalt-NTA-Superflow was used instead of a Nickel-NTA-Superflow for affinity chromatography. No glycerol was added to any buffer as it did not make any detectable differences. To not interfere with the following labeling reactions, the storage buffer did not contain any DTT or DTE. Protein concentrations were determined by absorption at 280 nm according Gill and Hippel using an extinction coefficient of  $45400 \text{ M}^{-1} \text{ cm}^{-1}$  (Gill and Vonhippel, 1989).

FRET-labeling was performed in two subsequent steps. The protein was first incubated with a deficit amount of Alexa647 maleimide C2 (Alexa647) (Invitrogen, Germany). To start the first labeling reaction, a protein solution of 100-300  $\mu\text{M}$  concentration in labeling buffer containing 50 mM Tris-HCl (pH 7.4), 5 mM  $\text{MgCl}_2$ , 250 mM NaCl was gently mixed with a 1.5-fold molar excess of Alexa647. After 1 hour incubation on ice, the unbound dye was

removed using a HiPrep 26/20 S25 Desalting column (GE Healthcare, Germany) with a flow rate of 0.5 ml/min. After this first labeling step, double, single and unlabeled proteins were separated based on the charge difference introduced by the coupled dyes using anion exchange chromatography on a ResourceQ column (GE Healthcare, Germany) and a salt gradient running from 0-500 mM NaCl over 120 ml at a pH of 7.4 and flow rate of 2.0 ml/min. The peaks appearing in the elugram were analyzed for their labeling degree by measuring their absorption by UV/Vis spectroscopy at a wavelength of 280 nm and 651 nm. The fraction with the highest, single-acceptor labeled protein amount was subsequently labeled with a 4-fold molar excess of Alexa488 C5 maleimide (Alexa488) and the unreacted dye was separated as described for the first labeling step. Finally, the degree of labeling for both dyes was determined and was usually 70-100% for each dye. The labeling efficiencies were determined by absorption using  $71.000 \text{ M}^{-1} \text{ cm}^{-1}$  and  $265.000 \text{ M}^{-1} \text{ cm}^{-1}$  as extinction coefficients for Alexa488 and Alexa647, respectively. The labeled proteins were aliquoted into buffer containing 50 mM Tris-HCl (pH 7.9), 5 mM MgCl<sub>2</sub>, 2 mM DTT, shock-frozen in liquid nitrogen and stored at -80 °C.

The spin labeling reactions were conducted at 4 °C for 3 hours using an 8-fold excess of (1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate (MTSSL) as a spin label (Enzo Life Sciences GmbH, Germany). The reaction was performed in 50 mM Tris, 5 mM MgCl<sub>2</sub> solved in D<sub>2</sub>O at pH 7.4. Unbound spin labels were removed with Zeba Spin Desalting Columns (Thermo Fisher Scientific GmbH, Germany) equilibrated with 50 mM Tris, 5 mM MgCl<sub>2</sub> solved in D<sub>2</sub>O at pH 7.4. Concentrations were determined as described above. Labeling efficiencies have been determined by double integration of CW room temperature (RT) EPR spectra in comparison with EPR samples of known concentrations, and were ~90-100% in all cases. In all EPR experiments the protein concentration was 100 µM.

## 2.2 FRET

### 2.2.1 Experimental methods

#### *Ensemble TCSPC*

Ensemble time-correlated single-photon-counting (eTCSPC) measurements of the donor fluorescence were either performed on an IBH-5000U (HORIBA Jobin Yvon IBH Ltd., UK) equipped with a 470 nm diode laser LDH-P-C 470 (Picoquant GmbH, Germany) operated at 8 MHz or on a EasyTau300 (PicoQuant, Germany) equipped with an R3809U-50 MCP-PMT detector (Hamamatsu) and a BDL-SMN 465 nm diode laser (Becker & Hickl, Germany) operated at 20 MHz. To detect the donor fluorescence, the emission wavelength was set to 520 nm using a slit-width that resulted in a spectral resolution of 16 nm in the emission path

of the machines. Additionally, a 500 nm cut-off filter placed in the detection path was used to reduce the contribution of the scattered light. All measurements were conducted at room temperature under magic-angle conditions with a protein concentration of approximately 200 nM. Here typically photons  $14 \cdot 10^6$  to  $20 \cdot 10^6$  were recorded at a channel-width of 14.1 ps (IBH-5000U) or 8 ps (EasyTau300). Differential non-linearities were considered in the fitting procedure by multiplying the model function with the previously recorded normalized/smoothed uncorrelated instrumental response. The fits cover the full instrument response function (IRF) and 99.9% of the total fluorescence to resolve short lifetime components. The experimental IRFs had typically FWHM of 254 ps (IBH-5000U) or 85 ps (PicoQuant EasyTau300).

### *Single-molecule measurements*

The freely diffusing molecules were excited by a linearly polarized pulsed argon-ion laser (Sabre®, Coherent) through a corrected Olympus objective (60X, 1.2 NA collar (0.17)). The laser was operated at 496 nm and 72 MHz at an excitation power of 120  $\mu$ W at the objective. The fluorescence light was collected through the same objective and spatially filtered by a 100  $\mu$ m pinhole which defines an effective confocal detection volume. A polarizing beam-splitter divided the collected light into parallel and perpendicular components. Next, the light passed a dichroic beam splitter that defines a “green” and “red” wavelength range (below and above 620 nm). After passing through band pass filters (AHF, HQ 520/35 and HQ 720/150) the light was detected by two “green” (either  $\tau$ -SPADs, PicoQuant, Germany or MPD-SPADs, Micro Photon Devices, Italy) and two “red” detectors (APD SPCM-AQR-14, Perkin Elmer, Germany). Two SPC 132 counting boards (Becker & Hickel, Berlin) recorded the detected photon stream. Thus for each detected photon the arrival time after the laser pulse, the time since the last photon, the polarization and the wavelength was recorded. All measurements were performed at concentrations of approximately 50 pM to assure that only single-molecules were detected.

### **2.2.2 Analysis methods**

#### *Burst-wise analysis*

Prior to subsequent data-analysis steps fluorescence bursts were discriminated from the background signal of 1–2 kHz of the single-molecule measurements by applying certain threshold intensity criteria (Eggeling et al., 2001). Next the anisotropy (Schaffer et al., 1999) and the fluorescence averaged lifetime were determined for each burst (Maus et al., 2001). Additionally, we consider the background fluorescence, the detection efficiency-ratio of the “green” and “red” detectors, and the spectral cross-talk to determine absolute FRET-

efficiencies for each burst (Sisamakris et al., 2010b). As the freely diffusing molecules were only observed for a brief moment of time (in the order of milliseconds) the number of detected photons and the complexity of the model function are limited (Kollner and Wolfrum, 1992). Hence, the determined lifetime of each burst corresponds to the fluorescence averaged lifetime even in presence of lifetime distributions (Kalinin et al., 2010b). To allow for more complex models the fluorescence of multiple burst was combined into a joint fluorescence decay histogram (sub-ensemble analysis, seTCSPC). Using species averaged lifetime of the donor in absence of an acceptor  $\langle \tau_{D(0)} \rangle_x$ , the experimental fluorescence averaged lifetime of the donor in presence of an acceptor  $\langle \tau_{D(A)} \rangle_x$  and the FRET-efficiency  $E$ , the mean  $\langle \tau_{D(A)} \rangle_x = (1-E_S) \cdot \langle \tau_{D(0)} \rangle_x$ , and the variance  $\langle \tau_{D(A)} \rangle_x = (1-E) \cdot \langle \tau_{D(0)} \rangle_x$  of the lifetime distribution for each burst were calculated and shown in the supplement as histogram to highlight conformational dynamics.

#### *Fluorescence decay analysis*

The probability distribution for emission of a single-photon is modulated by FRET and follows the intensity distribution of all emitted photons. By analysis of the fluorescence intensity decays of the donor in presence ( $f_{D/D}^{DA}(t)$ ) and absence of FRET ( $f_{D/D}^{D0}(t)$ ) FRET can be quantified, thus donor-acceptor distances are estimated. In seTCSPC additionally the FRET-sensitized emission of the acceptor ( $f_{A/D}^{DA}(t)$ ) and the fluorescence decay of the directly excited acceptor ( $f_{A/A}^{A0}(t)$ ) were fitted. Site-specific fluorescence properties of the dyes were accounted for by separate reference samples of the donor and the acceptor using single cysteine variants. Both, donor and acceptor were tethered to the biomolecule by cysteine-maleimide labeling chemistry. Hence, the distribution of the dyes among the two labeling sites was initially unknown but determined using two donor labeled single cysteine reference samples by limited proteolysis, anisotropy-PDA and global analysis of six independent fluorescence intensity decays (see chapter C). Prior to fitting of the polarized seTCSPC histograms, the magic-angle fluorescence intensity decays were calculated using separately measured correction factors which correct for polarization dependent detection efficiencies and anisotropy mixing in the objective (Koshioka et al., 1995).

Differences in the local environment of the dyes within their accessible volume result in complex fluorescence decays of the donor  $f_{D/D}^{D0}(t)$  and the acceptor  $f_{D/D}^{A0}(t)$  even in the absence of FRET. The sample specific decays were analyzed formally by a multi-exponential model functions:

$$f_{D|D}^{D0}(t) = \sum_i x_i^{(i)} \exp(-t/\tau_i^D) \quad (1)$$

$$f_{A|A}^{A0}(t) = \sum_i x_i^{(i)} \exp(-t/\tau_i^A)$$

Here  $D|D$  refers to the donor fluorescence under the condition of donor excitation and  $A|A$  refers to the acceptor fluorescence under acceptor excitation. The individual species fractions  $x_i^D$  and  $x_i^A$  are and lifetimes of the donor  $\tau_i^D$  and the acceptor  $\tau_i^A$  are summarized in Table S2/S3.

We assume that all fluorescence states of the donor are quenched by the same FRET-rate constant distribution and that the radiative lifetime of the donor is not affected by quenching. Therefore the fluorescence decay of the donor in presence of acceptor,  $f_{D|D}^{DA}(t)$ , can be expressed by the following product:

$$\begin{aligned} f_{D|D}^{DA}(t) &= \left( \sum_{i=1} x_i^D \exp(-t/\tau_i^D) \right) \cdot \left( \sum_{j=1}^K x_j^{RET} \exp(-t \cdot k_j^{RET}) \right) \\ &= f_{D|D}^{D0}(t) \cdot \varepsilon_D(t) \end{aligned} \quad (2)$$

Here  $\varepsilon_D(t)$  is the FRET-induced donor decay which is given by  $(x_j^{RET}, k_j^{RET})$  the distribution of FRET-rate constants. Under these assumptions the FRET-rate constant ( $k_{RET}$ ) is only determined by the donor-acceptor distance and their relative orientation. Dyes coupled by flexible long linkers are best described by the “diffusion with traps” model (Kalinin et al., 2015). Our single-molecule anisotropy measurements demonstrate that the major fraction of the dyes is mobile during the fluorescence lifetime. Therefore, we approximate the orientation factor  $\langle \kappa^2 \rangle = 2/3$  and relate the fluorescence decay of the donor in presence of FRET to the donor-acceptor distance distribution by:

$$\varepsilon_D = \frac{f_{D|D}^{DA}(t)}{f_{D|D}^{D0}(t)} = \int_{R_{DA}} p(R_{DA}) \cdot \exp(-t \cdot k_0 \cdot (R_0 / R_{DA})^6) dR_{DA} \quad (3)$$

Therein  $p(R_{DA})$  is a FRET-rate distribution expressed as distance and  $R_0$  is the Förster-radius (in this case  $R_0 = 52$  Å) and  $k_0 = 1/\tau_0$  is the radiative rate of the unquenched dye. The fluorescence decay of the FRET-sensitized acceptor  $f_{A|D}^{DA}(t)$  was fitted by the convolution of

the direct excited fluorescence decay of the acceptor,  $f_{A|A}^{A0}(t)$ , and the donor fluorescence decay in presence of FRET,  $f_{D|D}^{DA}(t)$ :

$$f_{A|D}^{DA}(t) = f_{D|D}^{D0}(t) \cdot \varepsilon_D(t) \otimes f_{A|A}^{A0}(t) \quad (4)$$

Considering the fluorescence decay of all donors (DOnly and DA)  $f_{D|D}(t)$ , the fraction of FRET-inactive species  $x_{DOnly}$  attributed to incompletely labeled or bleached molecules was considered as a constant offset in the FRET-induced donor decay,  $\varepsilon_D(t)$ :

$$f_{D|D}(t) = f_{D|D}^{D0}(t) \cdot (\varepsilon(t) + x_{DOnly}) \quad (5)$$

In seTCSPC the fraction of FRET-inactive species  $x_{DOnly}$  was minimized using the acceptor fluorescence as an additional selection criterion (see Table S1).

The fluorescence intensity decay curves were fitted using the iterative re-convolution approach (Phillips et al., 1985). In addition to the

$$g(t) = N_F \cdot f(t) \otimes IRF(t) + N_{BG} \cdot IRF(t) + bg \quad (6)$$

Here  $g(t)$  is the model function fitted to the experimental data using the modified Levenberg–Marquardt algorithm *lmdif* implemented in MINPACK (Moré et al., 1980),  $N_F$  is the number of fluorescence photons,  $N_{BG}$  is the number of background photons due to Rayleigh or Raman scattering and  $bg$  is a constant offset attributed to detector dark counts and afterpulsing. In seTCSPC the fraction of scattered light and the constant background was calculated given the experimental integration time and the buffer reference measurements. In eTCSPC the fraction of scattered light and the constant offset were free fitting parameters. In all cases the model function  $g(t)$  was scaled to the data by the experimental number of photons.

The experimental fluorescence decays were fitted by custom software written in Python which allows for the global analysis of multiple experimental curves. The statistical errors were estimated by sampling the parameter space of the fitting parameters and applying the F-distribution to 68% confidence given the minimum determined  $\chi^2$  (Goodman and Weare, 2010, Foreman-Mackey et al., 2013).

#### *FRET-lines*

The maximum likelihood estimator determines the fluorescence weighted average lifetime of the donor  $\langle \tau_{D(A)} \rangle_F$  while the FRET-efficiency  $E$  quantifies the species averaged lifetime of the donor. We use a generalized methodology similar as described previously to describe dynamic

FRET-lines (Kalinin et al., 2010a) to relate these two moments of the lifetime distribution (see chapter C). For a given distance distribution  $p(R_{DA})$  the FRET-rate constant distribution  $p(k_{RET})$  is calculated. Next, knowing the lifetime distribution of the donor in absence of FRET  $p(\tau_{D(0)})$  the lifetime distribution of the donor in presence of FRET  $p(\tau_{D(A)})$  is calculated. Using these lifetime distributions, we calculate the transfer-efficiency  $E$  and  $\langle \tau \rangle_F$ , the fluorescence averaged lifetime, using the species average and the fluorescence average, respectively. This procedure is repeated for a set of model parameters which are changed within a given range. The projection of the change of this model parameter to the two observables we call a FRET-line. In this paper we use two types of FRET-lines: dynamic and static FRET-lines. Dynamic FRET-line quantify the relative fraction of typically a two component model (usually two distances) while static FRET-lines are characterized by the change of a parameter which quantifies a mean donor-acceptor distance. Here all FRET-lines were calculated assuming that the donor-acceptor distances are Gaussian distributed. All reported histograms are corrected for instrumental parameters such as spectral cross-talk, background fluorescence and detection efficiency ratios while the FRET-lines were calibrated for sample specific fluorescence properties such as the donor and acceptor fluorescence quantum yields, the fraction of acceptor in the fluorescent trans-state of Alexa647 and multiple lifetimes of the donor.

Alternatively, to a “classical” FRET-line  $\langle \tau_{D(A)} \rangle_F$ , the fluorescence weighted average lifetime and  $E$ , the measured transfer-efficiency, are combined with  $\langle \tau_{D(0)} \rangle_x$ , the species weighted fluorescence lifetime of the donor in absence of FRET, to determine the variance of the lifetime-distribution for a single-molecule:

$$\text{Var}(\tau_{D(A)}) = \langle \tau_{D(A)} \rangle_F \cdot \langle \tau_{D(A)} \rangle_x - \left( \langle \tau_{D(A)} \rangle_x \right)^2 \quad \text{with} \quad \langle \tau_{D(A)} \rangle_x = (1 - E) \langle \tau_{D(0)} \rangle_x \quad (7)$$

Using this transformation, conformational dynamics during the integration time of the experiment is directly determined by a non-zero variance.

Static and dynamic FRET-lines were calculated numerically as described in Chapter C of this dissertation. All FRET-lines are corrected for the lifetime distribution of the donor-reference samples. The static FRET-lines were calculated assuming using normal distributions with a width of 12 Å. The dynamic FRET-lines were calculated numerically by varying the relative species fraction between two anticipated limiting states. These limiting states were determined by analyzing the donor-fluorescence decays and the time-resolved FRET-sensitized acceptor emission of the FRET-molecules by a two component model with two normal distributions with a width of 12 Å each.

*Lifetime-filtered species cross-correlation*

Photon bursts of differing mean transfer-efficiency were used to define limiting states. These states are designated as high- (HF) and low-FRET (LF) state. Next, polarization and spectral (green/red) resolved fluorescence intensity decays of the non-fluorescent background and HF and LF were generated. These decays were used to create filters  $w_j^{(i)}$  as previously described (Felekyan et al., 2012). Correlation functions  $G^{(n,m)}(t_c)$  were calculated using the recorded signal intensities  $S_j(t)$  weighted by the filters  $w_j^{(i)}$ :

$$G^{(n,m)}(t_c) = \frac{\langle F^{(n)}(t) \cdot F^{(m)}(t+t_c) \rangle}{\langle F^{(n)}(t) \rangle \cdot \langle F^{(m)}(t+t_c) \rangle} \text{ with } F^{(n)}(t) = \left( \sum_{j=1}^{d \cdot L} w_j^{(n)} \cdot S_j(t) \right) \quad (8)$$

Herein  $(n)$  and  $(m)$  are two species in a mixture,  $d$  is the number of detectors,  $L$  is the number of TAC-channels and  $S_j(t)$  is the signal recorded in the TAC-channel  $j$ . The choice of  $n$  and  $m$  defines the type of the correlation function. If  $n$  equals  $m$   $G^{(n,m)}(t_c)$  is a species autocorrelation function otherwise  $G^{(n,m)}(t_c)$  is a species cross-correlation function (Felekyan et al., 2013). Given the selected High-FRET (H) and Low-FRET-state (L), overall four correlations were calculated per sample: two species autocorrelation functions  $SACF^{H,H}(t_c)$ ,  $SACF^{L,L}(t_c)$  and two species cross-correlation functions  $SCCF^{H,L}(t_c)$ ,  $SCCF^{L,H}(t_c)$ .

To interpret the correlation curves, a model function which factorizes into a diffusion-term  $G_{\text{Diff}}^{(n,m)}(t_c)$  and a kinetic-term  $G_{\text{kin}}^{(n,m)}(t_c)$  was used:

$$G^{(n,m)}(t_c) = 1 + \frac{1}{N_{\text{eff}}^{(n,m)}} \cdot G_{\text{Diff}}^{(n,m)}(t_c) \cdot G_{\text{kin}}^{(n,m)}(t_c) \quad (9)$$

Here,  $N^{(n,m)}$  is the effective number of molecules. It was set as a free fitting parameter in the  $SACFs$  and global among the  $SCCFs$  of each sample.

As found to be appropriate for modern confocal detection, the spatial detection profile  $W(x,y,z)$  can be described by a three dimensional Gaussian. Consequently, the diffusion term  $G_{\text{Diff}}$  of the correlation is given by (Schaefer, 1973):

$$G_{\text{Diff}} = \left(1 + t_c/t_{\text{Diff}}\right)^{-1} \left(1 + \left(\frac{\omega_0}{z_0}\right)^2 (t_c/t_{\text{Diff}})\right)^{-1/2} \quad (10)$$

The parameters  $\omega_0$  and  $z_0$  are define the shape of the three dimensional Gaussian  $W(x, y, z) = \exp(-2(x^2 + y^2)/\omega_0^2) \cdot \exp(-2z^2/z_0^2)$  in  $x$ ,  $y$ - and  $z$ -direction, respectively. We assume that the correlation curves of each sample are described by the same diffusion term.

The kinetic factor of all four correlation curves was described formally by a set of correlation (*SACF*) and anti-correlation terms (*SCCF*):

$$\begin{aligned} G_{kin}^{L,H}(t_c) &= (1 - A_0^{LH} \cdot (A_1 \cdot e^{-t_c/t_1} + A_2 \cdot e^{-t_c/t_2} + A_3 \cdot e^{-t_c/t_3})) \\ G_{kin}^{H,L}(t_c) &= (1 - A_0^{HL} \cdot (A_1 \cdot e^{-t_c/t_1} + A_2 \cdot e^{-t_c/t_2} + A_3 \cdot e^{-t_c/t_3})) \cdot (1 - A_b^{HL} \cdot e^{-t_c/t_b}) \\ G_{kin}^{L,L}(t_c) &= (1 + A_1^{LL} (e^{-t_c/t_1} - 1) + A_2^{LL} (e^{-t_c/t_2} - 1) + A_3^{LL} (e^{-t_c/t_3} - 1)) \\ G_{kin}^{H,H}(t_c) &= (1 + A_1^{HH} (e^{-t_c/t_1} - 1) + A_2^{HH} (e^{-t_c/t_2} - 1) + A_3^{HH} (e^{-t_c/t_3} - 1)) \cdot (1 + A_b^{HH} (e^{-t_c/t_b} - 1)) \end{aligned} \quad (11)$$

In these equations  $A_0$  determines the amplitude at  $t=0$ ,  $A_b$  is the amplitude attributed to bleaching of the acceptor molecule from the high-FRET state,  $t_b$  is the characteristic bleaching time of the acceptor (under the given conditions typically 5-10 ms). The parameters  $A_1$ ,  $A_2$  and  $A_3$  together with the times  $t_1$ ,  $t_2$  and  $t_3$  define the relaxation time spectrum of the transition between the high-FRET and the low-FRET state. The amplitudes  $A_1$ ,  $A_2$  and  $A_3$  were additionally constrained to positive values and sum up to unity. The relaxation times  $t_1$ ,  $t_2$  and  $t_3$  were fitted globally among all samples while the amplitudes were sample specific. Overall 12 samples were analyzed resulting in 48 correlation curves which were globally fitted. The data analysis software was written in Python and is available upon request.

## 2.3 Simulations

### 2.3.1 Rigid-body docking

To obtain a set of structures which fulfil the experimental FRET and EPR distance constraints, we performed rigid body docking as previously described (Kalinin et al., 2012c). First hGBP1 was decomposed into its individual domains: the LG-domain (aa 1-309), the middle-domain (aa 310-481) and the helices  $\alpha_{12}$  (aa 482-563) and  $\alpha_{13}$  (aa 564-583). To allow for rearrangements within the middle-domain, it was decomposed into two sub-units (aa 310-373, aa374-481). The helices  $\alpha_{12}$  (aa 482-563) and  $\alpha_{13}$  (aa 564-583) were treated as individual units. Following the primary sequence, the N- to the C-terminal parts of the rigid bodies were connected via bonds with a weak quadratic potential. As this reduced model does not allow for bending of the individual domains we deliberately choose a very soft clash-potential. Details on the rigid body docking are given in the supplement.

### 2.3.2 Normal mode based geometric simulations

In addition to rigid body docking with experimental constraints we used the computationally more demanding normal mode based geometric simulations (NMSim, collaboration Holger Gohlke, University Düsseldorf) which generates all-atom representations with stereochemical accurate conformations. Conventional NMSim is a three-step protocol for multiscale modelling of protein conformations and incorporates information about preferred directions of protein motions into a geometric simulation algorithm (Ahmed et al., 2011). First, static properties of the protein are determined by decomposing the molecule into rigid clusters and flexible regions using the graph theoretical approach FIRST (Jacobs et al., 2001). Next, dynamical properties of the molecule are identified by an elastic network model representation of the coarse-grained protein, as implemented in the RCNMA approach (Ahmed et al., 2010, Ahmed and Gohlke, 2006). In this step, only rigid body motions are allowed for rigid clusters, while links between them are treated as fully flexible. In the final step, termed NMSim, the recently introduced idea of constrained geometric simulations of diffusive motions in proteins (Wells et al., 2005) is extended (Ahmed et al., 2011). Details of the implementation are given in (Ahmed et al., 2011). We incorporated experimental FRET-constraints into NMSim by repeating the stochastic diffusive motion at each iteration step three times. For the next iteration the conformation closer to the experimental constraints, judged by the sum of all squared weighted deviations, is taken.

### 2.3.3 Principle component analysis

To identify collective degrees of freedom, essential movements and correlated domain motions of hGBP1 from the known crystal structure (PDB code: 1DG3) we used Principal Component Analysis (PCA) of accelerated MD simulations (aMD) (Hamelberg et al., 2004). hGBP1 was protonated with PROPKA (Bas et al., 2008) at a pH of 7.4, neutralized by adding counter ions and solvated in octahedral box of TIP3P water (Jorgensen et al., 1983) with a water shell of 12 Å around the solute. The obtained system was used to perform unbiased molecular dynamics (MD) simulations and accelerated MD simulations. The Amber14 package of molecular simulation software (D.A. Case, 2015) and the ff14SB (Wang et al., 2000) force field were used to perform five unrestrained all-atom MD simulations of 50 ns length each. “Particle Mesh Ewald” (Darden et al., 1993) method was utilized to treat long-range electrostatic interactions; SHAKE algorithm (Ryckaert et al., 1977) was applied to bonds involving hydrogen atoms. The mass of solute hydrogen atoms was increased to 3.024 Da and the mass of heavy atoms was decreased respectively according to the Hydrogen Mass Repartitioning method (Hopkins et al., 2015). The time step in all MD simulations was 4 fs

with a direct-space, non-bonded cutoff of 8 Å. For initial minimization, 17500 steps of steepest decent and conjugate gradient minimization were performed; harmonic restraints with force constants of 25 kcal·mol<sup>-1</sup> Å<sup>-2</sup>, 5 kcal·mol<sup>-1</sup>·Å<sup>-2</sup>, and zero during 2500, 10000, and 5000 steps, respectively, were applied to the solute atoms. Afterwards, 50 ps of NVT simulations (MD simulations with a constant number of particles, volume, and a thermostat) were conducted to heat up the system to 100 K, followed by 300 ps of NPT simulations (MD simulations with a constant number of particles, barostat and thermostat) to adjust the density of the simulation box to a pressure of 1 atm and to heat the system to 300 K. A harmonic potential with a force constant of 10 kcal·mol<sup>-1</sup> Å<sup>-2</sup> was applied to the solute atoms at this initial stage. During the next 100 ps of NVT simulations the restraint forces on the solute atoms were gradually reduced to zero, followed by 200 ps unrestrained NVT simulations. Additionally, two a MD simulation of 100 ns and 225 ns were performed. Boost parameters were chosen according to the method suggested in (Pierce et al., 2012).

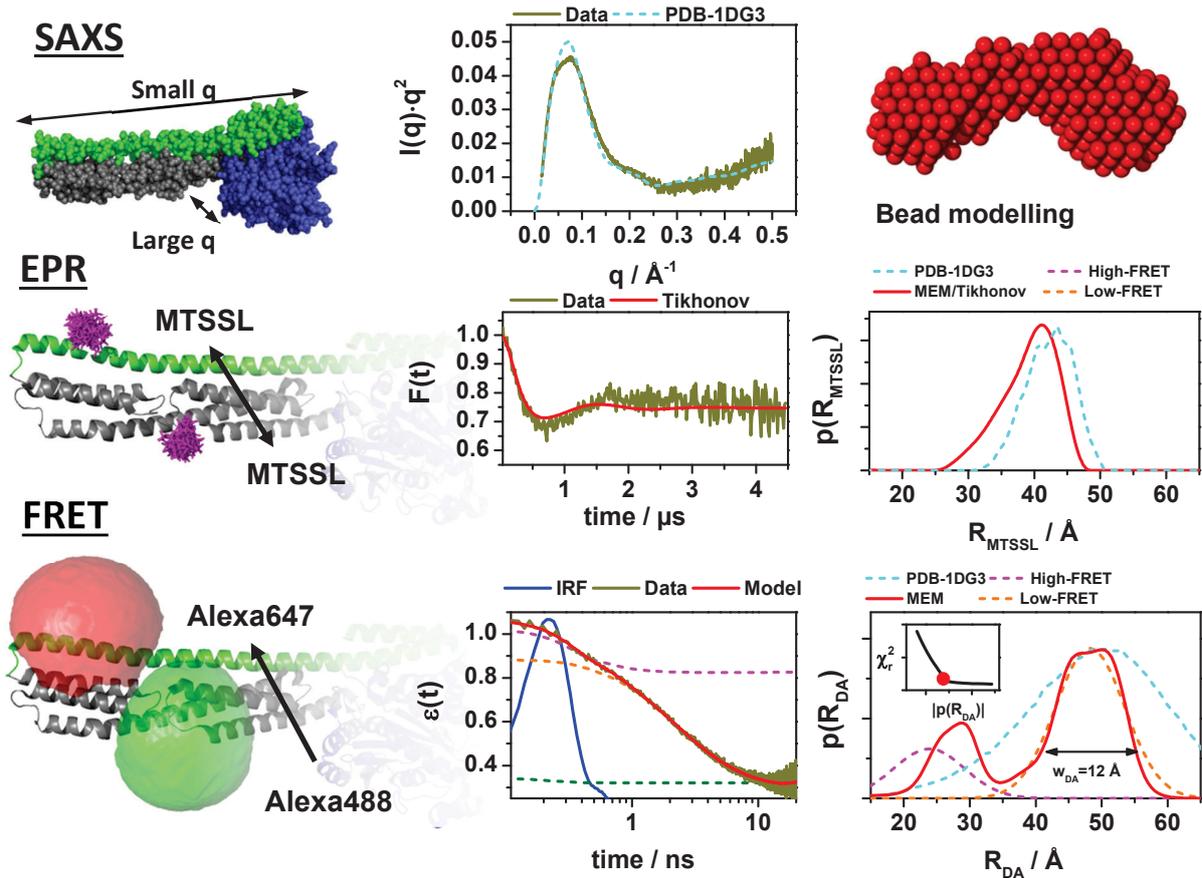
### 2.3.4 Accessible volume simulations and FPS

The spatial distribution of the fluorophores modelled by the accessible-volume (AV) approach according to (Cai, Kusnetzow et al. 2007; Muschielok, Andrecka et al. 2008; Sindbert, Kalinin et al. 2011). A geometric search algorithm determines all dye positions within the linker-length from the attachment point which do not cause steric clashes with the protein surface. The chromophores were approximated by ellipsoids. The centre of each ellipsoid was connected by a linker to the C $\beta$ -atoms of the reactive amino-acid. The longest distance from the attachment point to the centre of the chromophore was used as length of the linker. The linker includes the reactive group, a spacer and the internal linker of the dye. Both donor and acceptor were attached to the protein by cysteines. Thus, both Alexa Fluor 488 C5 maleimide (Alexa488) and Alexa Fluor 647 C2 maleimide (Alexa647) were modelled with a linker width  $L_{width}$  of 4.5 Å; the linker-length  $L_{link}$  of 20.5 Å and 22 Å were used, for Alexa488 and Alexa647, respectively. The radii of the ellipsoids ( $R_{dye1}$ ,  $R_{dye2}$  and  $R_{dye3}$ ) were determined by the spatial dimensions of the dyes. Alexa488 was modelled using radii of 5.0 Å, 4.5 Å and 1.5 Å. Alexa647 was modelled using radii of 11.0 Å, 4.7 Å and 1.5 Å.

## 3 Results

### 3.1 Experimental equilibrium distributions

The FRET-, SAXS- and EPR- measurements probe distinct properties of hGBP1. SAXS measures the global shape while EPR and time-resolved FRET measurements determine distances between introduced labels (Fig. 2).



**Figure 2.** Used experimental techniques exemplified on the hGBP1 cysteine double variant Q344C/A496C. In the left panels the characteristic properties of the respective techniques are exemplified. SAXS measures the scattered intensity in dependence of the scattering vector  $q$ . By this information on the radial distribution function of scattering object is obtained. EPR measures the dipolar coupling between two spin-labels (MTSSL) introduced at specific sites. In FRET measurements the energy transfer from a donor-fluorophore to an acceptor fluorophore is measured. In FRET- and EPR-measurements the spatial distributions of the labels have to be considered. The spatial distribution of the EPR MTSSL-label is shown in magenta as calculated by the MTSSL-wizard (Hagelueken et al., 2012). The length of the linkers for the dyes Alexa488 and Alexa647 is larger. Therefore, the conformational space is determined by accessible volume calculations and presented as semi-transparent “clouds” (donor, green) and acceptor (acceptor, red) (see Supplement A). The middle panel shows representations of the experimentally measured data. Top Kratky-plot of the experimental data (dark yellow) overlaid by the calculated scattering curve of the PDB 1DG3 (cyan), middle experimental DEER-trace (dark yellow) as described by Tikhonov regularization (red line), bottom experimental FRET-induced donor decay (dark yellow) overlaid by a maximum entropy model and a fitted model function composed of two Gaussian distributed distances (magenta and orange dashed lines) with a width of  $w_{\text{DA}}=12 \text{\AA}$  and a constant offset (FRET-inactive molecules, green). Right panels: top low resolution model determined by SAXS shown as scattering beads. The recovered distance distribution (red) is compared to the distance distribution as expected by the crystal structure (PDB: 1DG3) shown as dashed cyan line. Middle the inter-spin label distance distribution  $p(R_{\text{MTSSL}})$  as recovered by Tikhonov analysis of DEER-trace. Bottom the donor-acceptor distance distribution  $p(R_{\text{DA}})$  recovered from the fluorescence intensity decays by the maximum entropy method (red) and by fitting the fluorescence intensity decays (magenta, and orange peaks). The dashed orange and magenta lines correspond to the high-FRET and mid-FRET state, respectively, as determined by Gaussian fitting.

A first analysis of the measured SAXS scattering profile clearly showed that the average radial distribution function of the atoms within hGBP1 is in disagreement with the known crystal structure of the full length protein. This is visualized by the Kratky-plot shown in Fig. 2 (middle top) where the calculated scattering curve of the crystal structure (PDB-ID: 1DG3) is overlaid with the experimental data. The most likely distribution of scattering beads

resembles the shape of a full length protein with an additional kink between the LG- and the middle domain (Fig. 2, right top).

The distance network measured by EPR and FRET offers additional details by probing distances between the LG- and the middle-domain, the middle domain and helix  $\alpha_{12/13}$  and within the  $\alpha_{12/13}$ -domain. For the double variant Q344C/A496C model free analysis of the EPR-data reveals a distance distribution which is significantly shorter than the distance distribution anticipated by the crystal-structure and is in clear disagreement with it (see Fig. 2, middle of middle and right panel). The recovered inter-spin distance distribution is asymmetric which might be due to an overlaying second minor state or due to linker artifacts.

The time-resolved FRET-measurements are visualized by the FRET-induced donor decay. In the semi-logarithmic plot of the FRET-induced donor decay ( $\epsilon_D(t)$ ) shown in Fig. 2 (middle bottom) a single FRET-rate constant should become visible as a well resolved “step”. The location of the “step” is a measure of the FRET-rate constant whereas its step height measures the species fraction. The FRET-rate constant is a measure of the donor-acceptor distance. However, the lack of a single step in  $\epsilon_D(t)$ , indicates a more complex scenario with multiple DA distances. To recover and visualize the underlying distance distribution we use the maximum-entropy method which reveals two separated peaks (Fig. 2, right bottom). To compare these peaks with the crystal structure, we calculate the distance distribution as expected by accessible volume (AV) calculations. The AV calculations determine the sterically allowed conformational space of the donor- and acceptor dye. Using these calculations, the donor-acceptor distance distribution was determined using a uniform weighting of the AVs. This results in distance distributions which are systematically too broad (Hofig et al., 2014, Sindbert et al., 2011b) but reproduce the average inter-fluorophore distances reliably (Kalinin et al., 2012c). The position of one peak is in good agreement with the crystal structure while the second peak clearly disagrees (Fig. 2, right bottom). This is an evidence for at least two distinct conformational states. To determine reliable statistical estimates on the inter-fluorophore distances and to assign conformational states unambiguously by using the amplitudes of the two states, a model composed of two normal distance distributions with a width of 6 Å was constructed. All measured ensemble fluorescence decays were consistently analyzed by this model, keeping the relative amplitudes of the supposed conformational states global.

The result of this analysis is summarized in Table 1. Remarkably, even though all fluorescence decays were consistently analyzed by a two component model, certain FRET-

pairs solely require a single normal distributed distance distribution while other FRET-pairs demand for a second distance distribution.

**Table 1.** Analysis of ensemble fluorescence by a global model constituted of two normal distributed donor-acceptor distance distributions with a width of 12 Å. The distances of higher amplitude are addressed by M1, distances of the lower amplitude are addressed by M2

FRET-pair	$\langle R_{DA} \rangle_{M1}$ [Å]	Estimated error [Å]	$\langle R_{DA} \rangle_{M2}$ [Å]	Estimated error [Å]	$\Delta \langle R_{DA} \rangle$ [Å]	Relative difference
N18C/Q344C	73.6	8.3	67.0	5.4	7	0.13
N18C/V540C	57.8	3.6	36.6	2.7	21	0.65
N18C/Q577C	64.2	4.7	47.4	2.8	17	0.43
Q254C/Q344C	81.3	15.1	72.3	7.6	9	0.16
Q254C/V540C	63.6	4.5	36.8	2.7	27	0.77
Q254C/V577C	70.8	9.0	52.9	7.4	18	0.40
Q344C/T481C	37.9	2.6	54.4	3.3	17	0.52
Q344C/A496C	48.0	2.9	23.5	8.2	25	0.99
Q344C/Q525C	46.7	2.8	20.7	13.3	26	1.11
Q344C/V540C	59.3	3.8	45.5	2.8	14	0.38
T481C/Q525C	69.6	6.3	69.5	6.3	0	0.00
A496C/V540C	63.6	4.6	63.6	4.6	0	0.00
<b>Species fraction</b>	<b>0.61</b>		<b>0.39</b>			

The estimated errors ( $\sigma$ ) includes statistical uncertainties, potential systematic errors of the donor references, estimated uncertainties of the orientation factor determined by time-resolved anisotropy measurements of donor samples and uncertainties of the accessible volumes due to the differences of the donor and acceptor linker-length. Reference measurements of single donor and acceptor labeled variants are summarized in Table S2 and S3, respectively.

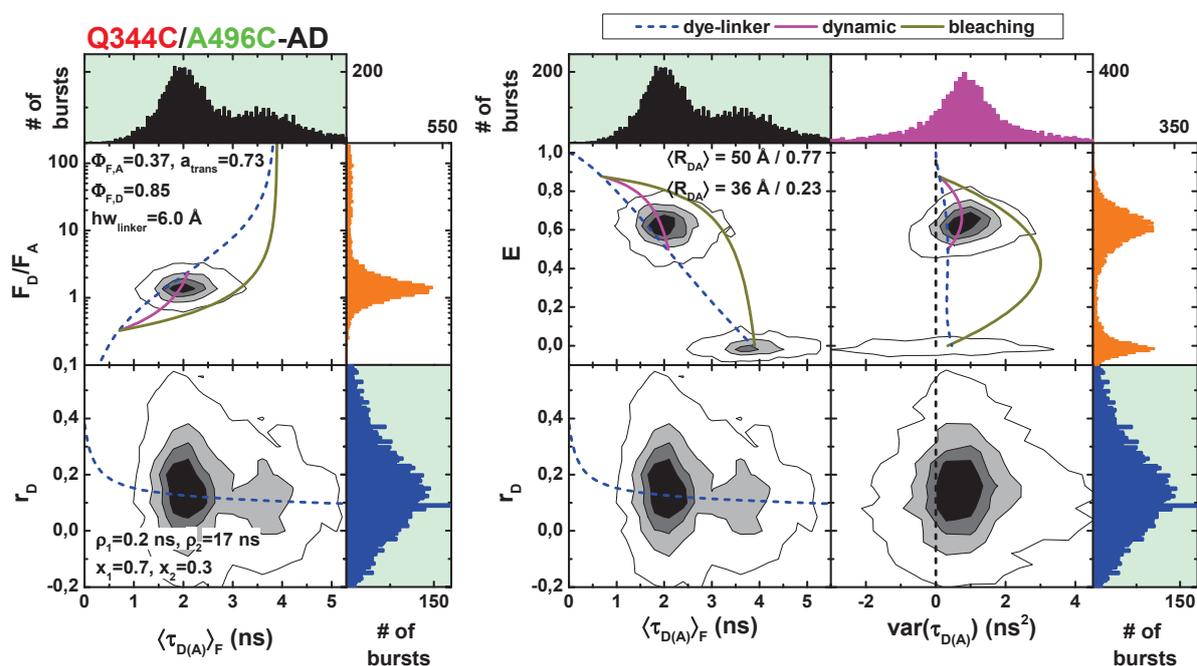
Measurements on the variant A496C/V540C, which was designed to test the stability of helix  $\alpha_{12}$ , showed no significant broadening of the recovered distance distribution. Similarly, no significant differences were detected in the FRET-pair T481C/Q525C which probes for movements from the end of the middle-domain to helix  $\alpha_{12}$ . The variants N18C/Q344C and Q254C/Q344C which measure distances from the middle domain to the LG-domain reveal only relatively small changes. All other variants, which probe for distance changes to helices  $\alpha_{12/13}$ , show significant differences.

The fact that in certain cases only one normal distribution is sufficient while in other cases two normal distributions are necessary gives us confidence that the global analysis of all datasets indeed stabilizes the fitting and does not introduce significant bias.

### 3.2 Identification and quantification of molecular kinetics

To check and to quantify potential dynamics exchange between the detected conformational states, we performed single-molecule MFD-measurements of freely diffusing, labelled hGBP1 molecules without ligands (Apo-state). By determining the fluorescence weighted average lifetime ( $\langle \tau_{D(A)} \rangle_F$ ) and the steady-state transfer-efficiency  $E$  for each detected single molecule we constructed multi-dimensional FRET-efficiency frequency histograms using fully corrected transfer efficiencies (Sisamakris et al., 2010a). In these histograms the determined

transfer-efficiency is displayed in dependence of the fluorescence weighted average lifetime (see Fig. 3).



**Figure 3.** Frequency histograms of a single-molecule measurement recorded with multi-parameter fluorescence detection of the FRET labeled double cysteine variant Q344C/A496C. The dashed blue lines describe the expected behavior of a single protein conformation in absence of protein dynamics. The one dimensional histograms are projections of the multi-dimensional frequency histograms. The top rows,  $F_D/F_A$  (fluorescence intensity ratio of the donor and the acceptor) and  $E$  (the FRET-efficiency), are measures of FRET. In the bottom row the steady-state anisotropy of the donor  $r_D$  is shown. The  $\langle \tau_{D(A)} \rangle_F$  is the fluorescence weighted average lifetime of the donor in presence of FRET. The two-dimensional anisotropy histograms in dependence of  $\langle \tau_{D(A)} \rangle_F$  are described by a Perrin-equation with two rotational correlation times WELCHE. The transfer-efficiency is corrected for the donor- and the acceptor fluorescence quantum yields as determined by separate reference samples. The acceptor fluorescence quantum yield is additionally corrected for the fraction of Alexa647N in the bright trans-state, as estimated by fluorescence correlation spectroscopy. The FRET-lines were calculated as described in the method section assuming a normal distributed donor-acceptor distribution with a width of 12 with a width  $w = 12 \text{ \AA}$ . To highlight the presence of conformational dynamics the variance and the mean of the lifetime distribution is calculated for each detected single-molecule and plotted in dependence of the FRET-efficiency. This is equivalent to a mean-variance histogram. The analysis results of globally fitted donor-and FRET-sensitized acceptor fluorescence decays of the FRET-molecules are shown as inset. The dynamic FRET-lines of interconversion between these two fitted states are shown in magenta. The yellow FRET-line describes the bleaching from the high-FRET state. The displayed FRET-lines are exact solutions numerically calculated given the fit-results as reported as insets. The Perrin-equation is given by eq. 9 in chapter B.

Surprisingly only a single peak is visible even though the ensemble measurements revealed two distinct populations. This apparent discrepancy between the fluorescence decay analysis and the single-molecule burst analysis can be understood by considering conformational dynamics. Suppose a single-molecule changes its conformational state several times during the observation time (the burst duration), then only a averaged transfer efficiency is detected. Fortunately, by comparing the determined fluorescence weighted average lifetime to the transfer-efficiency fast conformational dynamics becomes detectable by a shift from the static-FRET line (see chapter C). Alternatively, the fluorescence weighted average lifetime is

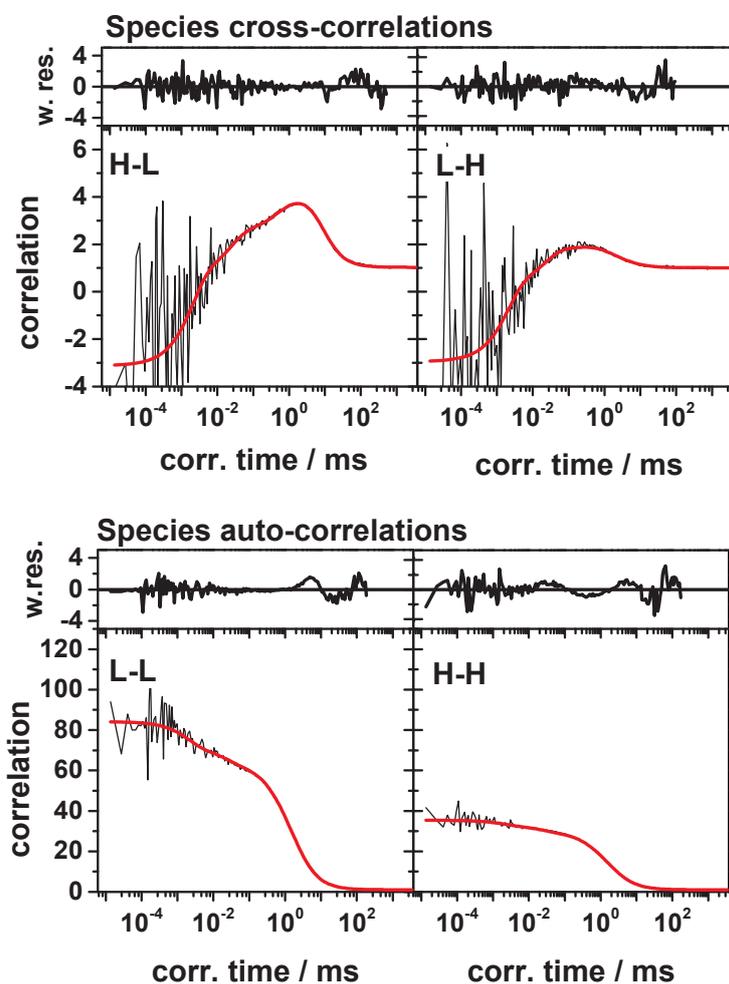
combined with the detected transfer-efficiency to determine the variance of the underlying fluorescence lifetime distribution (see Fig. 3, right). A positive variance is a direct evidence for a distribution of lifetimes, and hence conformational dynamics. Practically, dye-linker dynamics and dynamic quenching of the donor contributes to broadening of the lifetime distributions. Here, these effects were explicitly considered. Nevertheless, significant additional shifts were detected, which are a clear indication for conformational dynamics of the host-molecule.

To understand the origin of these shifts and to cross-validate the ensemble measurements with the single-molecule measurements, fluorescence decays of the sub-ensemble of FRET-molecules were compiled (Fig. S3). The fluorescence intensity decay of the donor and the FRET-sensitized acceptor emission were globally analyzed by a two component model composed of two normal distributed donor-acceptor distance distributions (eqs. XXX) (Table S1). Using this analysis result, a dynamic FRET-line which describes the exchange between the fitted distributions was calculated. These dynamic FRET-lines describe the peak position in the multi-dimensional FRET-histograms in Fig. 3 perfectly. These shifts were consistently found in all variants (Fig. S2), which required a second distance distribution, while the mean distances and species fractions are in good agreement with the ensemble measurements. This confirms our supposed theory that we find two conformational states in solution which are in dynamic exchange. The observation time set by the burst-duration and the peak width allows us to conclude, that the time-scale of dynamics is significantly faster than the burst-duration (~ms).

To determine the exact time-scale(s) of the dynamic exchange between the two conformations, we performed lifetime filtered correlations of all 12 FRET-pairs (see Methods section). Surprisingly, a single relaxation time was insufficient to describe the anti-correlation as determined by the individual samples. This can be already evident by a visual inspection of the experimental data (see Fig. 4).

The addition of a second relaxation time significantly improved the fit; however, as the addition of a third component did not reduce  $\chi^2$  considerably and also the noise level of the curves prohibited the use of a more complex model, we accepted the two relaxation time model as sufficient to describe the individual datasets. Though, both, the found relaxation times *and* the amplitudes differed among the datasets. Therefore, a direct comparison between the datasets was not possible. Therefore we performed global analysis of the all fFCS curves of

all variants for achieving a consistent interpretation. A formal model with three global correlation times was necessary to describe all FRET-samples (Fig. 4, Fig. S4).



**Figure 4.** Experimental lifetime filtered correlation of the FRET-labeled Q344C/A496C variant (black lines) overlaid by a model function (red) fitted using global correlation times among all 12 FRET-pairs and free amplitudes for each variant. Left species cross-correlations between the high-FRET (H) and the low-FRET (L) species. Right species auto-correlations of the L- and the H-species. Curves of the other variants are shown in Figure S4.

This significantly facilitated the interpretation as the sample specific amplitudes can be directly compared. As for each FRET-pair the two possible species auto-correlations and the two cross-correlations were calculated, hence, in total 48 FCS-curves were globally analyzed. An excerpt of this analysis for variant Q344C/A496C is shown in Fig. 4. The amplitudes of the recovered relaxation times are summarized in Table 2.

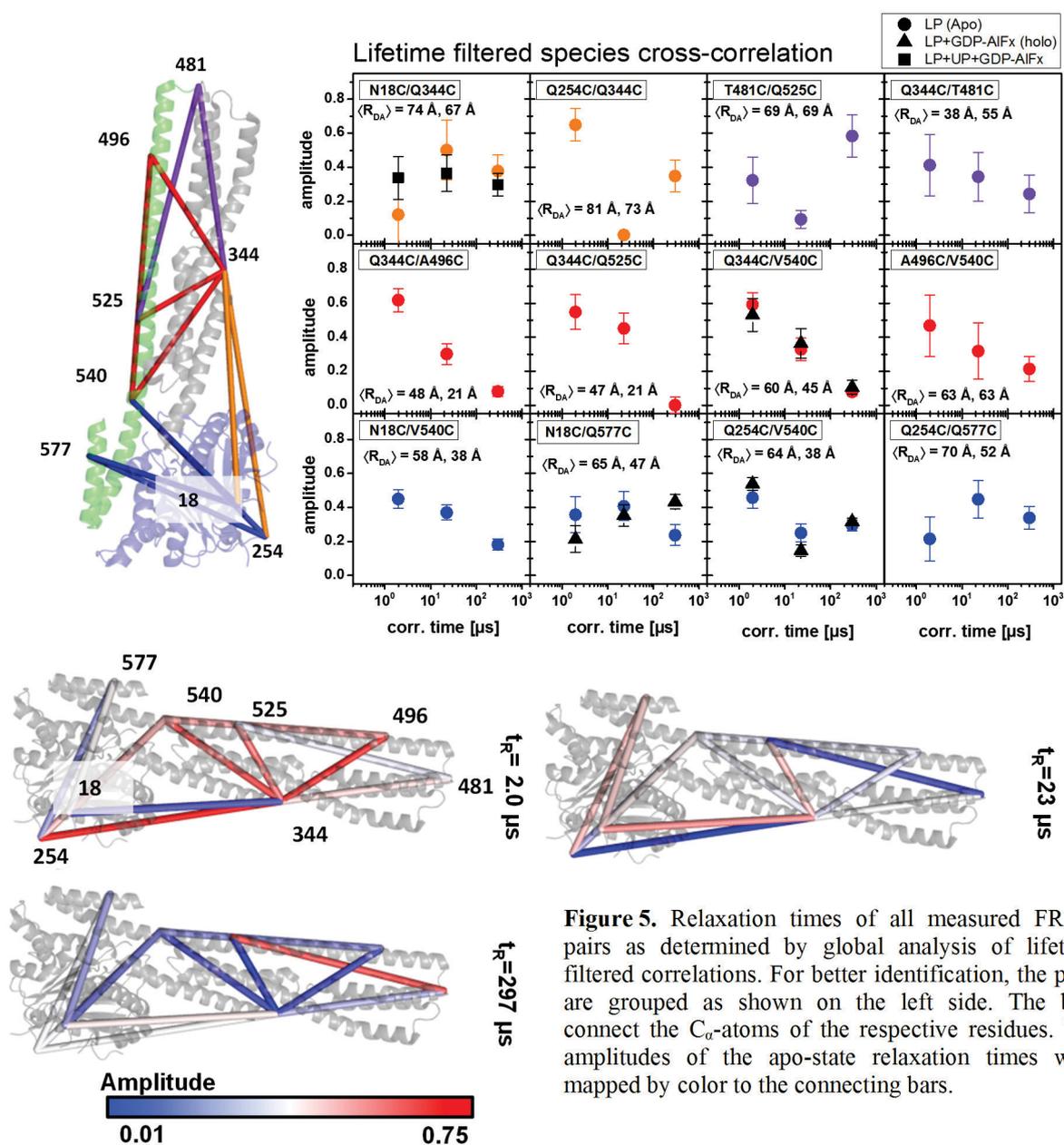
The shown excerpt visualizes the power of lifetime filtered correlations to amplify the contrast between distinct FRET-species. At the same time, it highlights the limits of the analyzed dataset in resolving fast relaxation times. The noise level of the correlation curves increase significantly with decreasing relaxation time. Therefore, for the given lifetime filtered correlations solely kinetics faster than the presented shortest correlation time of 2  $\mu$ s cannot be excluded. However, the neutron-spin echo measurements revealed (personal communication with Andreas Stadler and Ralf Biehl, FZ-Jülich) that the amplitude of conformational dynamics up to a correlation time of 300 ns is negligible. This makes us confident that our relaxation time distribution covers all relevant time-scales.

**Table 2.** Distribution of hGBP1 relaxation times (mean (M) and standard deviation (SD) determined by global analysis of filtered species auto and cross-correlation curves.

Sample	Relaxation time $t_c$ [ $\mu$ s]						$N_1$	$N_2$
	297.6		22.6		2.0			
	M	SD	M	SD	M	SD		
N18C/Q344C	0.38	0.09	0.50	0.18	0.12	0.20	0.001	0.010
N18C/V540C	0.18	0.03	0.37	0.04	0.45	0.06	0.013	0.123
N18C/Q577C	0.24	0.06	0.41	0.09	0.36	0.11	0.053	0.009
Q254C/Q344C	0.35	0.09	0.00	0.00	0.65	0.09	0.013	0.030
Q254C/V540C	0.34	0.03	0.21	0.05	0.45	0.06	0.019	0.036
Q344C/T481C	0.24	0.11	0.34	0.14	0.41	0.18	0.020	0.021
Q344C/A496C	0.08	0.03	0.30	0.06	0.62	0.07	0.012	0.029
Q344C/Q525C	0.00	0.05	0.45	0.09	0.55	0.10	0.058	0.055
Q344C/V540C	0.08	0.02	0.33	0.07	0.59	0.07	0.049	0.038
T481C/Q525C	0.58	0.13	0.09	0.05	0.32	0.14	0.029	0.015
A496C/V540C	0.21	0.07	0.32	0.17	0.47	0.18	0.005	0.010
Q254C/Q577C	0.34	0.07	0.45	0.11	0.21	0.13	0.036	0.008

The donor and acceptor column indicate the anticipated position of the donor and the acceptor fluorophores, M and SD are the mean and the standard deviation as determined by the fitting procedure,  $N_1$  and  $N_2$  are the effective numbers of molecules of species 1 and 2, respectively.

A direct inspection of the amplitude spectrum presented in Table 2 was not very elusive. Overall, significant differences between the FRET pairs were detectable. To relate the recovered relaxation time spectrum with structural properties it was mapped to the crystal structure as a preliminary model (Fig. 5).



**Figure 5.** Relaxation times of all measured FRET-pairs as determined by global analysis of lifetime filtered correlations. For better identification, the pairs are grouped as shown on the left side. The bars connect the C $_{\alpha}$ -atoms of the respective residues. The amplitudes of the apo-state relaxation times were mapped by color to the connecting bars.

By this representation differences among the samples in their apo state became evident. In most cases the shortest relaxation time had the highest amplitude and the relaxation time spectrum decayed with longer correlation time. This result was consistent with the burst-wise analysis of the single-molecule measurements. All single-molecule FRET-efficiency histograms revealed no significant broadening (Fig. S3). Hence, fast correlation times as presented in Fig. 5 and Table 2 were anticipated. Interestingly, a distribution of relaxation time was found. Thus, a kinetic model with two interchanging states would be inappropriate to describe the experimental data. Notably, the amplitude spectra of many variants did not have a shape which would be equivalent to a stretched exponential model, so that a more complex exchange dynamics had to be considered.

To study the effect of substrate on the dynamics, we additionally performed EPR-, SAXS- and FRET-measurements in presence of GDP-AlFx, and GppNHP. At micro-molar (unlabeled) GTPase concentrations dimerization of hGBP1 took place and severe conformational changes were detected by FRET. Reporting on all these measurement results and the conformational changes goes beyond the scope of this work. Thus, we solely report on substrate effects in “absence” of other GTPases. Under our single-molecule measurement conditions the concentration of hGBP1 was approximately 50 pM. Hence, even if the nucleotide was bound no dimerization could occur. We used GDP-AlFx as substrate as it showed the strongest effects at higher hGBP1 concentrations. We anticipate that, under the sm measurement conditions, GDP-AlFx was bound to the LG-domain but due to the low protein concentration no dimerization occurs. We refer to this measurement conditions as holo-form of the protein. We selected set of variants (N18C/Q577C, Q254C/V540C, Q344C/V540C) in which we found the biggest substrate induced effects at higher hGBP1 concentrations. Additionally, we measured the relaxation time spectrum in the variant N18C/Q344C at higher total protein concentrations. Surprisingly, the detected relaxation time spectrum is within errors indistinguishable from the measurements of the nucleotide free apo-forms.

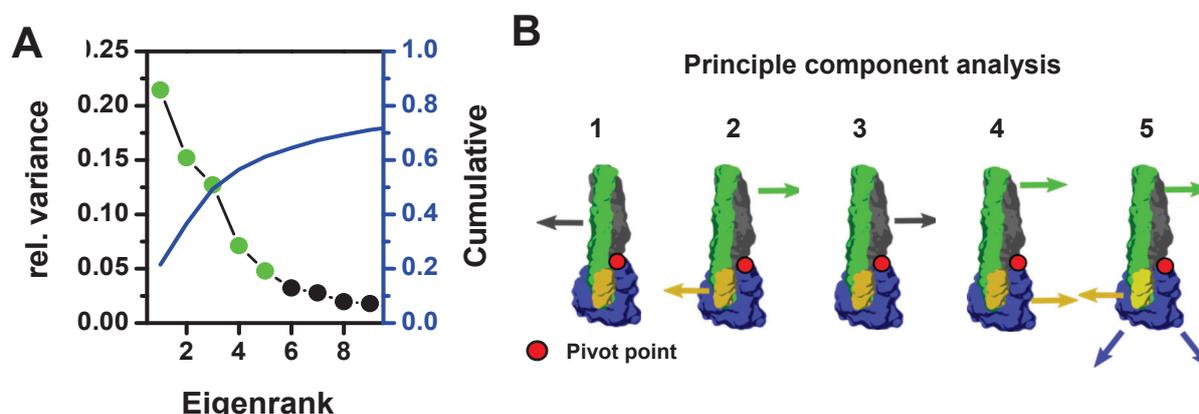
To conclude, we found distributions of relaxation times that are presumably unaffected by binding of a nucleotide to the catalytically active LG-domain. The relaxation time distributions are FRET-pair specific. On average the fast relaxation times were most prominent for FRET pairs located at the middle domain and the helices  $\alpha$ 12/13 (Q344C/A496C, Q344C/Q525C, Q344C/V540C).

### 3.3 Essential motions determined by molecular dynamics simulations

We performed accelerated molecular dynamics (MD) simulations starting from the full-length crystal structure of the apo-form (PDB-ID: 1DG3) to obtain a better understanding on possible motions within hGBP1. A visual inspection of the MD-simulations revealed a motion of the middle-domain and helices  $\alpha$ 12/13 towards the LG-domain. As we started from the crystal structure, such a motion is required to obtain a shape as determined by the bead-modelling of the SAXS-profile. Hence, this finding is consistent with the experimental data and suggests that the elongated straight structure of hGBP1 might be caused by crystal packing effects.

To reveal and identify additional correlated domain motions we performed a Principal Component Analysis (PCA) of the accelerated MD simulations. The principle components were sorted by the magnitude of the eigenvalues. The first 5 principle components contribute to 60% of the total variance of all simulations (see Fig. 6, Fig. S1). As already inferred by

visual inspection of the trajectories the main principle component characterizes a kinking movement of the middle- and  $\alpha$ 12/13-domain towards the LG-domain. The second principle component describes a “rolling” movement of the  $\alpha$ 12/13-domain on the LG-domain. This “rolling” movement is anti-correlated, which means it corresponds to a rotation around a pivot point located in proximity of helix  $\alpha$ 6 on the boundary of the LG- and the middle-domain. The third principle component describes the counter movement of the first component. The fourth principle component shows a correlated motion of the  $\alpha$ 12/13-domain with respect to the LG-domain, which does not involve a rotation around the suggested pivot point. Interestingly, the fifth principle component shows the same directionality of the second component, except that the “rolling” motion of the middle-domain and helices  $\alpha$ 12/13 are correlated with a breathing motion of the catalytic LG-domain.



**Figure 6.** Principle component analysis (PCA) of five MD simulations of 100 ns each, starting from the crystal structure 1DG3. (A) Magnitude of the eigenvalues of the determined principal components. The shown five components (green) correspond to 60% of the total variance in the simulations. (B) First five principle components illustrated by a reduced representation of hGBP1. In this representation the LG-domain is colored in blue the middle-domain is colored gray, the helix  $\alpha$ 12 and  $\alpha$ 13 are colored in green and yellow, respectively. The arrows show the direction of the motion. (1) the middle-domain moves with respect to the LG-domain preferentially to the “left” (in direction of the binding cap), (2) middle- and helical-domain move anti-correlated, helix  $\alpha$ 12 and helix  $\alpha$ 13 move along of the LG-domain while the middle domain moves out of the plane, (3) is a motion orthogonal to the first component with a eigenvalue that is two times smaller, (4) similar to the second principal component except that motion of the middle domain and helix  $\alpha$ 12/13 are correlated and move in the same direction, and (5) same directionality of the second component, except that the motion of the middle-domain and helices  $\alpha$ 12/13 are correlated with the motion of the catalytic LG-domain. All motions detected by the MD-simulations and the PCA can be reduced to rotations around a single pivot point (shown in red).

The molecular dynamics simulations cover only short time-scales compared (several hundred nanoseconds) to the relaxation times found by fFCS ( $\sim 2 \mu$ s). Nevertheless, the results are extremely valuable, as they indicate potential directions of motions for exchange between to limiting conformations of hGBP1, M1 and M2. In agreement with fFCS the  $\alpha$ 12/13-domain is highly mobile with respect to the LG-domain and exhibits a distinct “rolling” motion. The PCA helped to identify a pivot point (depicted in red in Fig. 6B) between the LG- and the middle-domain.

### 3.4 Structural modelling by experimental constraints

We utilized the measured distance distributions to obtain structural models of the two supposed conformational states. To determine correct structural models, the assignment of the recovered distance constraints to conformational states was essential. Additionally, to estimate the uncertainties of the structural model, the errors of the experimental constraints had to be quantified and the conformational states of the possible structural models were exhaustively sampled. The experimental error should ideally include assignment errors of distance constraints.

To determine structural models by FRET and EPR distance constraints, we considered the spatial distribution of the labels around their attachment point and steered the simulations by the deviation of the simulated and experimentally recovered distances. We utilized models of increasing complexity from rigid body docking (Kalinin et al., 2012a) over fast normal mode based atomistic geometric simulations to highly accurate all atom molecular simulations. This multi-scale approach allowed us to sample the experimentally allowed conformational space as vastly as possible to provide a realistic error estimate of the determined structural models.

In the EPR-measurements such a clear assignment of conformational states to the distances distributions recovered by Tikhonov-regularization was not possible. Unfortunately, the shape and the width of the inter-spin distance distributions depend on the conformational states and can only be estimated *a posteriori* once a structure is proposed. Thus, even though DEER-EPR measurements are recognized to provide experimental distance constraints with Ångström accuracy, we use the width of the recovered distance distributions as uncertainty to determine possible structural models.

In the FRET-experiments the assignment of distance constraints to conformational states was possible even though the individual peaks of the conformational states overlap in the donor-acceptor distance distributions. This was possible due to a global analysis of all measured FRET-pairs and is corroborated by the conformational dynamics which were detected by burst-wise analysis and lifetime filtered correlation of the single-molecule multi-parameter fluorescence experiments.

## F - Mapping the motions and structure of a state necessary for oligomerization of a large GTPase

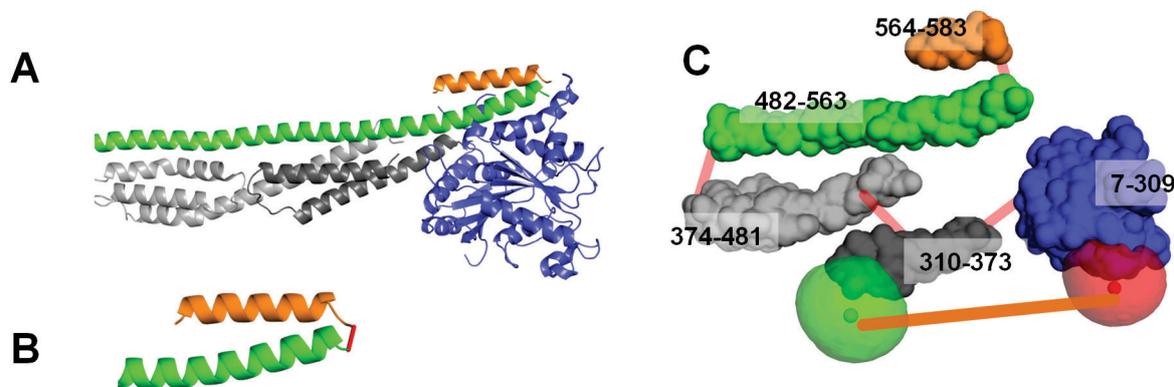
**Table 3.** EPR distance restraints as determined by Tikhonov regularization of the experimental DEER-traces and assignments to the supposed conformational states

Spin pair	$\langle R \rangle_{M1}$ [Å]	Uncertainty [Å]	$\langle R \rangle_{M2}$ [Å]	Uncertainty [Å]
N18C/Q344C	64.6	12.2	64.6	12.2
N18C/Q577C	53.2	7.6	53.2	7.6
C225C/K567C	15.0	8.0		8.0
C225C/Q577C	23.1	6.0		6.0
A551C/Q577C	21.0	8.0	21.0	8.0
Q344C/A496C	40.0	9.6	40.0	9.6
Q344C/Q525C	32.1	2.6	32.1	2.6
Q344C/V540C	46.6	6.3	46.6	6.3

The uncertainties were estimated by the peak width. The variants C225C/K567C and C225C/Q577C were deliberately only assigned to the first conformational state

The EPR-measurements could not resolve the conformational heterogeneities detected by FRET which were highlighted by the dynamic fingerprint of the molecule, global analysis of all FRET-datasets and amplified effects due to longer linkers. However, the EPR-measurements provided highly valuable information by short range distance constraints which are used to position the short helix  $\alpha 13$  relatively to helix  $\alpha 12$ .

To determine a coarse grained structural model which agree with the distance sets provided by the FRET-measurements and the EPR-measurements we performed rigid body docking. We considered the spatial distributions of the introduced labels as previously described.



**Figure 7.** Decomposition of human guanylate binding protein 1 into rigid bodies: LG-domain (blue), the middle domain (gray), helix  $\alpha 12$  (green), helix  $\alpha 13$  (orange). The middle domain was decomposed into two separate bodies to allow for more flexibility. (B) Loops between secondary structure elements define structural elements which were connected by weak quadratic potentials (C) Rigid bodies and amino-acid ranges (numbers) which define the bodies. The average fluorophore positions of an example FRET-pair are shown as spheres within the accessible volumes of the dyes which are shown as semi-transparent green (donor) and red (acceptor) surfaces.

Our rigid body docking procedure was originally designed for protein nucleic acid complexes (Kalinin et al., 2012b). To obtain structural models of protein we used a flexible assembly of rigid bodies which were connected via their C- and the N-termini (see Fig. 7). Our rigid body docking approach did not consider the flexibility of individual sub-units. Therefore, we used a very weak repulsion potential that tolerated atomic overlaps to a certain degree. The potential was already described previously (Kalinin et al., 2012b). The decomposition of the protein

into rigid domains followed the known well described sub-domains (LG, middle,  $\alpha 12/\alpha 13$ ). We allowed for a reorganization of the middle domain to facilitate a potential rolling motion of the helices  $\alpha 12/\alpha 13$  along the LG-domain. The AV calculations used for in rigid body dockings were originally designed for fluorophores connected to the host-molecule by long linkers ( $\sim 20$  Å). The linkers used to attach the EPR-probe to proteins are significantly shorter and stiffer. To incorporate the EPR-restraints we parameterized the EPR-labels by their linker-length, the linker-width and the label-radius. We benchmarked our used in the AV-calculations against the MTSSL-Wizard (Hagelueken et al., 2012) (details are given in the supplement). This allowed us to determine structural model which are in agreement with both: the FRET-restraints and the EPR-restraints.

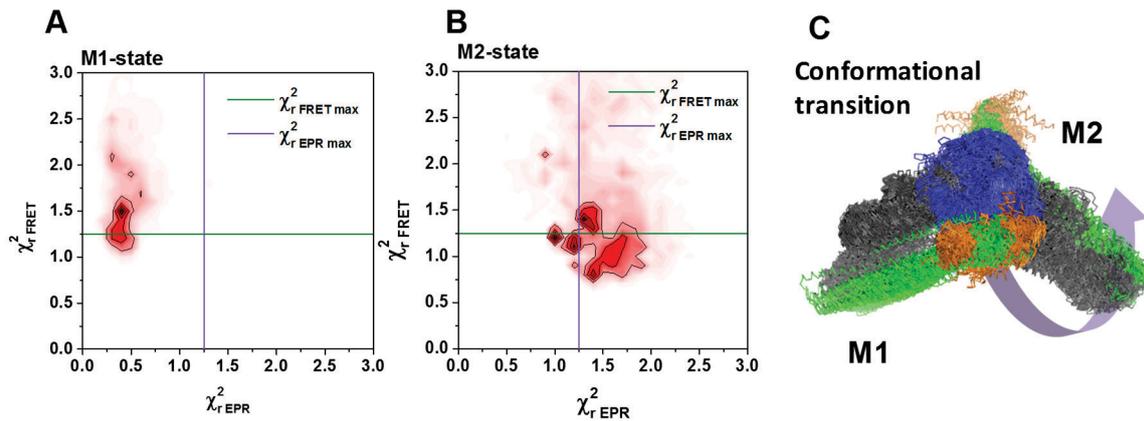
Initially all measured EPR-datasets were assigned to both anticipated conformational states and rigid body dockings were performed using the combined EPR- and FRET-datasets. The result of these dockings were inconsistent. For the majorly populated FRET-state (M1) the experimental constrains of the EPR-pairs C225C/K567C and C225C/Q577C were in good agreement between. In case of the minorly populated FRET-state (M2) C225C/K567C and C225C/Q577C disagreed with the found structural model (data not shown). Clearly the DEER-traces (see Fig. S5, supplement) show a pronounced initial drop and oscillate with a high frequency. However, given the length the DEER-traces of C225C/K567C (1  $\mu$ s) and C225C/Q577C (1.5  $\mu$ s) the presence of a long distance as predicted by the FRET-measurements was difficult to resolve. Therefore, the distance constraints of these measurements were only considered in the M1-state and not in the M2-state.

Below we only present the results of the FRET and EPR guided rigid body dockings. However, additional FRET-guided normal mode based geometric simulations (NMSim) followed by molecular dynamics (MD) simulations and screening of the obtained conformational ensemble by SAXS and FRET restraints are currently performed for further refinement of our results. To incorporate this additional information analogous to the FRET-data we are currently repeating our previous simulations: Rigid body docking, constraint guided NMSim and finally molecular dynamic simulations. Therefore, for consistency in this document solely results in rigid body docking are reported, even though other simulation and analysis techniques have been already successfully applied.

Using the combined distance constraints of the label based techniques (see Table 1 and Table 3) we determined a set of structural models which fulfilled the experimental constraints. We deliberately used an optimization technique, which was not strongly biased towards the

optimal solution. This allowed us to obtain conformational ensembles which are representative for the error-bounds of the experiments.

In Fig. 8 the result of the rigid body dockings for the M1- and the M2-state are shown. Given the modeled structures frequency histograms were constructed using reduced  $\chi^2$  of the FRET- and the EPR-constraints. This plot can be utilized to visualize the mutual selectivity of the methods. Overall the FRET-measurements allowed for a stronger discrimination of structures. The discrimination of conformers by EPR data was weaker, because the conformational states were not been clearly assigned to distances and the distribution width of the inter-spin distance distribution was taken as uncertainty. Nevertheless, the EPR-measurements allowed us to discriminate a fair amount of potential conformers of the M2-state. After selecting a certain number of structures using the  $\chi^2$  as a measure of the disagreement between the structural model and the experimental constraints we determined the allowed conformational ensembles for both detected states. We compared the ensembles by overlaying the LG-domains. To stress the differences among the found structural models, we first aligned the two conformational ensembles separately and aligned the LG-domains of these ensembles in the second step. As can be seen in Fig. 8C this alignment visualizes the conformational transition from between the M1- and the M2-state. As previously suspected by the result of the molecular dynamics simulations followed by the principle component analysis the connector between the LG- and the middle domain is a pivot point for the conformational transition. To go from the major state to the minor state the helices  $\alpha_{12/13}$  “roll” along the LG-domain while the middle-domain rotates around the connector to the LG-domain. Starting from the M1-state, which is comparable to the crystal structure except for a middle domain which is slightly kinked towards the LG-domain, the movement of the helices  $\alpha_{12/13}$  stops on the opposing site of the LG-domain.



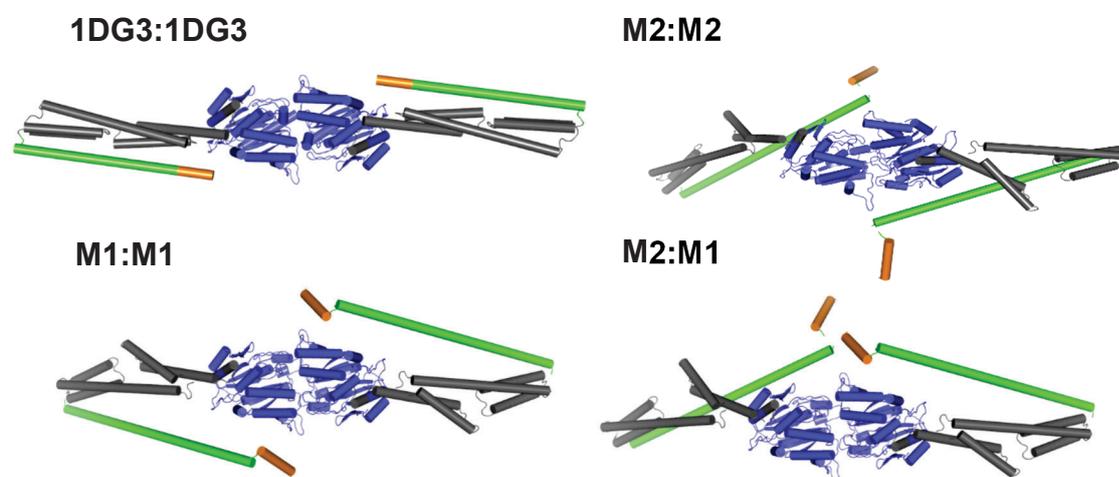
**Figure 8.** Conformational space sampled by rigid body dockings with FRET- and EPR constraints of the M1- (A) and the M2-state (B) FRET-state. Overall 100,000 structures were generated for each dataset and are displayed in a frequency histogram (from white to red being highest). For each structure the reduced  $\chi_r^2$  of the FRET experiments and the reduced  $\chi_r^2$  of the EPR-experiments were determined. By selecting structures below a certain threshold (in both cases a reduced  $\chi^2$  of 1.25) conformational ensembles in agreement with both experimental techniques were selected. The selected conformational ensemble of the major- and the minor FRET-state are displayed in form of a structural model (blue LG-, gray middle-, green helix  $\alpha 12$ , orange helix  $\alpha 13$ ). The two conformational ensembles were first internally aligned over all structures. Next, the LG-domain of the whole conformational ensembles were aligned to the common LG-domain. This alignment highlights the conformational transition from the M1- to the M2-state.

## 4 Discussion

We demonstrated by SAXS-, EPR- and FRET-measurements that the conformations of hGBP1 in solution deviate from the crystal structure. They are much better described by two distinct conformational ensembles M1 and M2 as limiting states which are in dynamic exchange. Hereby the conformers of the M1 ensemble are comparable to the crystal structure except for a middle domain which is slightly kinked to the LG domain. To perform the conformational transition, the helices  $\alpha 12/13$  change their relative position with respect to the LG-domain, this is accomplished by a rotation around a pivot point between the LG- and the middle domain. By this rotation the helices  $\alpha 12/13$  change their position from the “left” side to the “right” side of the LG-domain. Notably, the middle domain is kinked towards the LG-domain in both limiting states. The conformational transition among both limiting states is potentially very complex, which is consistent with the complex spectra of relaxation times indicating a rough energy landscape with several intermediates. Analogous to protein folding where Chung et al. (Chung et al., 2012) monitored the transition from the unfolded to the folded state and defined a transition path time, it would be intriguing to define an effective time for the conformational transition from M1 to M2. The conformational transition time would be definitively a convolute between all observed relaxation times (Figure 5, Table 2) and is expected to be in sub-millisecond time range. Hence, the complicated structural

rearrangement might be a reason why essentially no conformational dynamics was detected by NSE measurements.

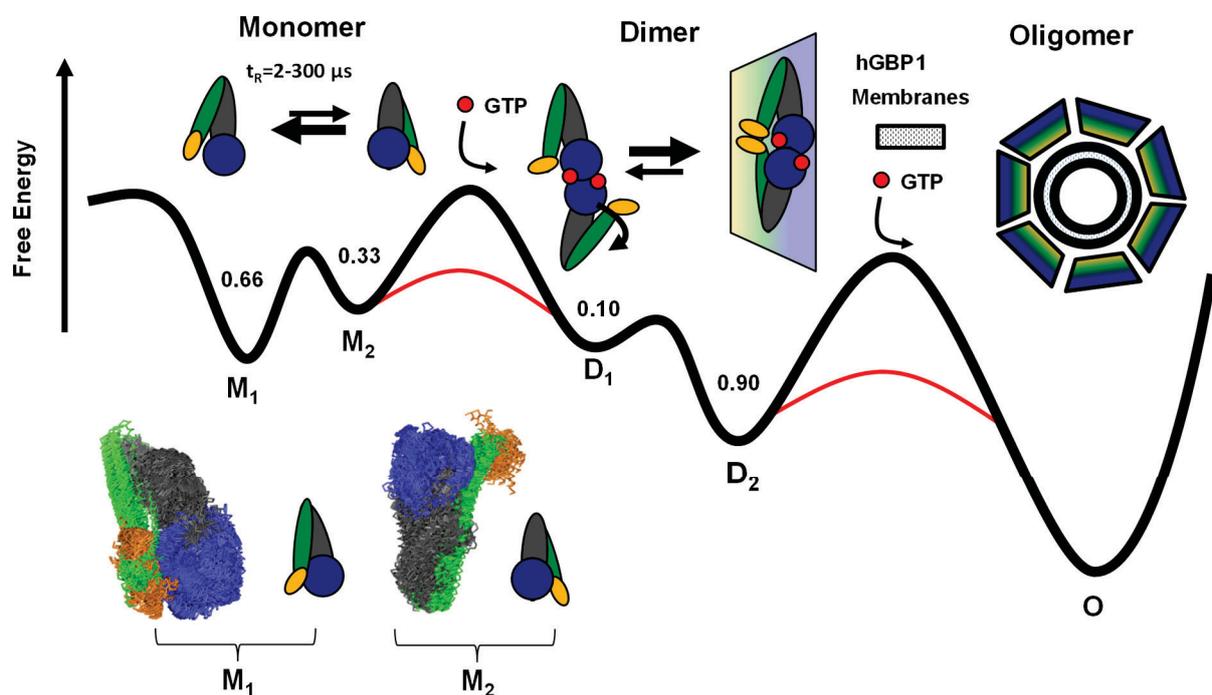
This structural information on the two conformational states M1 and M2 can be combined with pre-knowledge on the dimeric state of hGBP1 to discuss whether they could have a functional relevance in solution. We recently showed by FRET- and EPR-measurements that the interface of the hGBP1 dimer is described by a head-to-head dimer (see Supplement B) (Vopel et al., 2014). Additionally, we found that helix  $\alpha$ 13 dimerizes. Given the full length crystal structures the dimerization of helix  $\alpha$ 13 could not be explained. This is illustrated in detail in Fig. 9 where we used the best representative of both solution structures to construct all relevant dimer configurations using the confirmed head-to-head interface of the LG-domain. Overall 4 distinct conformations have to be considered (M1:M1, M1:M2, M2:M1, M2:M2). In the crystal structure the helices  $\alpha$ 12/13 are not centered with respect to the LG-domain. Hence, if the full length crystal structure (1DG3:1DG3) is preserved the  $\alpha$ 13 helices are separated by the LG-domains and cannot associate. Considering the resolved solution structures this is not necessarily the case because the helices  $\alpha$ 12/13 might be located at two distinct sites of the LG-domain.



**Figure 9.** Possible dimeric states constructed using the confirmed head-to-head interface to the LG-domain, the full-length crystal structure (1DG3), and the both found conformation states in solution. The LG-domains are colored in blue, the middle-domains are colored in gray, the helix  $\alpha$ 12 is colored in green and the helix  $\alpha$ 13 is colored in orange.

Due to symmetry reasons only three of these four possible dimers are depicted in Fig. 9. Considering the symmetric dimers M1:M1 and M2:M2, the helices  $\alpha$ 13 are also clearly separated by both LG-domains like in 1DG3:1DG3 even though the helices  $\alpha$ 13 are detached from the helices  $\alpha$ 12. However, in the case that a dimer is formed by two distinct conformers the helices  $\alpha$ 13 can associate in parallel manner as detected previously (Vopel et al., 2014).

The fFCS- and NSE-measurements suggest that the conformational transition is probably described by a complex diffusive process on a rough energy landscape. Interestingly, we detected no significant differences in the relaxation time spectrum of hGBP1 with bound substrate in absence of a second hGBP1 as a binding partner (holo-hGBP1). The measurement on the holo form of hGBP1 stressed that the conformational dynamics is independent of GTP. This indicates that GTP solely serves as activator for the LG-domain. The previous findings (Kravets et al., 2012, Vopel et al., 2014, Kravets et al., in revision) and new findings in this manuscript led to a common model which describes the reaction pathway of hGBP1 from the monomer to large-scale oligomers associated to membranes (see Fig. 10).



**Figure 10.** Possible reaction pathway from a monomer over a dimer to large oligomer. This reaction pathway summarizes current experimental findings (Kravets et al., 2012, Vopel et al., 2014, Kravets et al., in revision) (Supplement C). In solution two distinct conformations (M1 and M2) of hGBP1 are found. The current structural models of these conformers is depicted below. The transition between M1 and M2 can be described by a partial rotation of the middle-domain around the LG-domain. M1 and M2 are aligned via their LG domains. For simplicity hGBP1 is depicted as in a reduced representation. The hGBP1 monomer is in dynamic exchange following a complex relaxation time spectrum, at room-temperature M1 is the major state. The species fractions are given as numbers in the wells. Upon addition of GTP the LG-domain is activated for dimerization. The conformational dynamics is independent on the presence of a substrate. The formation of a conformational homo-dimer demands for a rearrangement of the middle domain as in a hetero-dimer helix  $\alpha 13$  associates. In presence of a membrane and multiple GBPs large membrane associated complexes are found.

In absence of substrate and additional GBPs, hGBP1 is found in two distinct conformational states. The characteristic relaxation times for the exchange between these two conformers range from  $2\mu\text{s}$  up to  $300\mu\text{s}$ . Upon addition of GTP the LG-domain is capable of binding and the conformational dynamics is unchanged. If two hGBP1s with bound GTP collide, a head-to-head dimer either in a M1:M1, M2:M2 or a M1:M2 configuration is formed. However, the

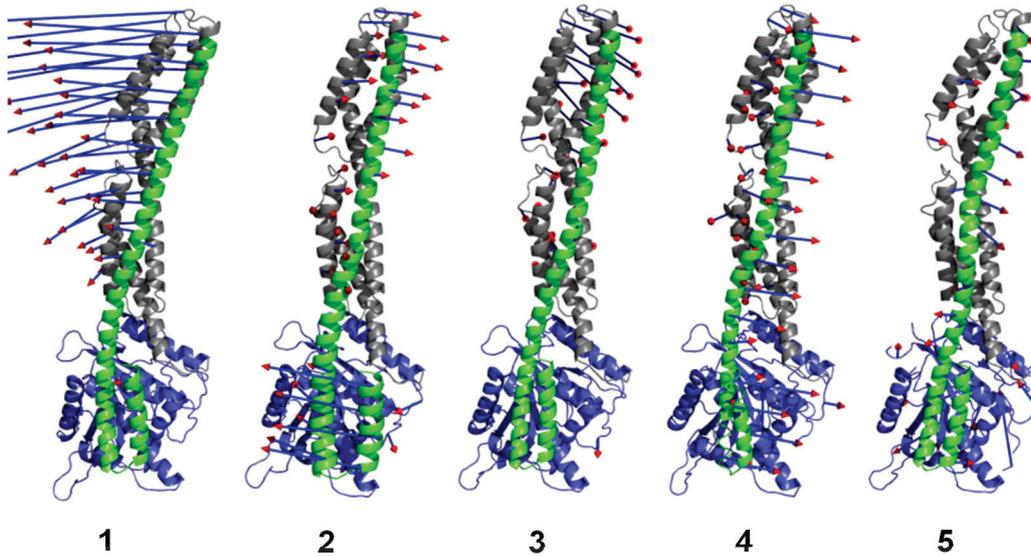
M1:M2 dimer has a higher stability, as the helices  $\alpha 13$  of the two subunits associate which shifts the equilibrium towards the M1:M2 dimers (Vopel et al., 2014). In presence of additional GTP and membranes large multimeric complexes consisting of several hundred monomers can be formed (Supplement C, Fig. 7D). In living cells these complexes directly attack endoplasmic parasites (Kravets et al., in revision).

To conclude, the conformational equilibrium found for the apo form of hGBP1 might have no direct functional relevance for monomeric hGBP1. However, the conformational flexibility of hGBP1 is essential to promote a key step in multimerization - the association of the helices  $\alpha 13$  with the membrane anchor - so that the dimer is further stabilized and the affinity to member insertion is increased.

In a more general view one can conclude that the exchange between distinct protein conformers may be often encoded in the design of the enzyme from the beginning. Thus, the conformational flexibility of an enzyme is already characteristic for the apo-form although this property is only relevant for a later stage of its functional cycle for example in complex with ligands, substrates and other proteins, respectively. Considering for example the movement the substrate-dependent conformational transitions in the fingers subdomain of a DNA polymerase (Rothwell et al., 2013), it is obvious that these opening and closing movements are essential for catalyzing polymerization under ambient conditions. The rule, that functionally relevant conformational equilibria may be predefined by the protein design, applies also for other steps in protein function, in this case oligomerization of hGBP1 as shown in this study.

## 5 Supplement

### 5.1 Principle component analysis



**Figure S1.** Molecular picture of the principle component analysis. The length of the vectors represent the relative magnitude of motion

### 5.2 Rigid body docking

In the used rigid body docking procedure in the optimization step the distance between the average fluorophore positions is used. The distance between the mean positions is defined by:

$$R_{mp} = \left| \left\langle \overrightarrow{R_D^{(i)}} \right\rangle - \left\langle \overrightarrow{R_A^{(j)}} \right\rangle \right| = \left| \frac{1}{n} \sum_{i=1}^n \overrightarrow{R_D^{(i)}} - \frac{1}{m} \sum_{j=1}^m \overrightarrow{R_A^{(j)}} \right| \quad (12)$$

Here  $R_D^{(i)}$  are the coordinates of the  $n$  possible donor positions and  $R_A^{(j)}$  are the coordinates of the  $m$  acceptor positions. The above average is faster calculated than mean donor acceptor distance given by:

$$\langle R_{DA} \rangle = \left| \left\langle \overrightarrow{R_D^{(i)}} - \overrightarrow{R_A^{(j)}} \right\rangle \right| = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \left| \overrightarrow{R_D^{(i)}} - \overrightarrow{R_A^{(j)}} \right| \quad (13)$$

Using the mean position of the dyes instead of the accessible volumes reduces the complexity of the rigid body docking and hence increases its speed (Kalinin et al., 2012c).

Neither the average transfer efficiency obtained by fluorescence intensity histograms nor the donor-acceptor distance distribution related obtained by the FRET-induced donor-decay relate directly to the distance between the mean dye positions (Wozniak et al., 2008, Sindbert et al., 2011a). However, the distance between the mean positions  $R_{mp}$  and the mean donor-acceptor

distance  $\langle R_{DA} \rangle$  can be related by a transfer function (Kalinin et al., 2012c). In the first iteration steps we make use of such transfer functions and to accelerate our rigid body dockings. Later the accessible volumes of the dyes are update and the model distances  $\{R_{\text{model}}\}$  are calculated using the updated accessible volumes.

In the rigid body docking a set of potentials are considered: the experimental constrains, clashes between rigid bodies and finally bonds between the rigid bodies. Minimization of sum of the experimental squared weighted deviations is equivalent to the energy minimization of a network of rigid bodies connected by  $N$  springs. We consider weights that depend on the sign of the deviation to account for asymmetric errors:

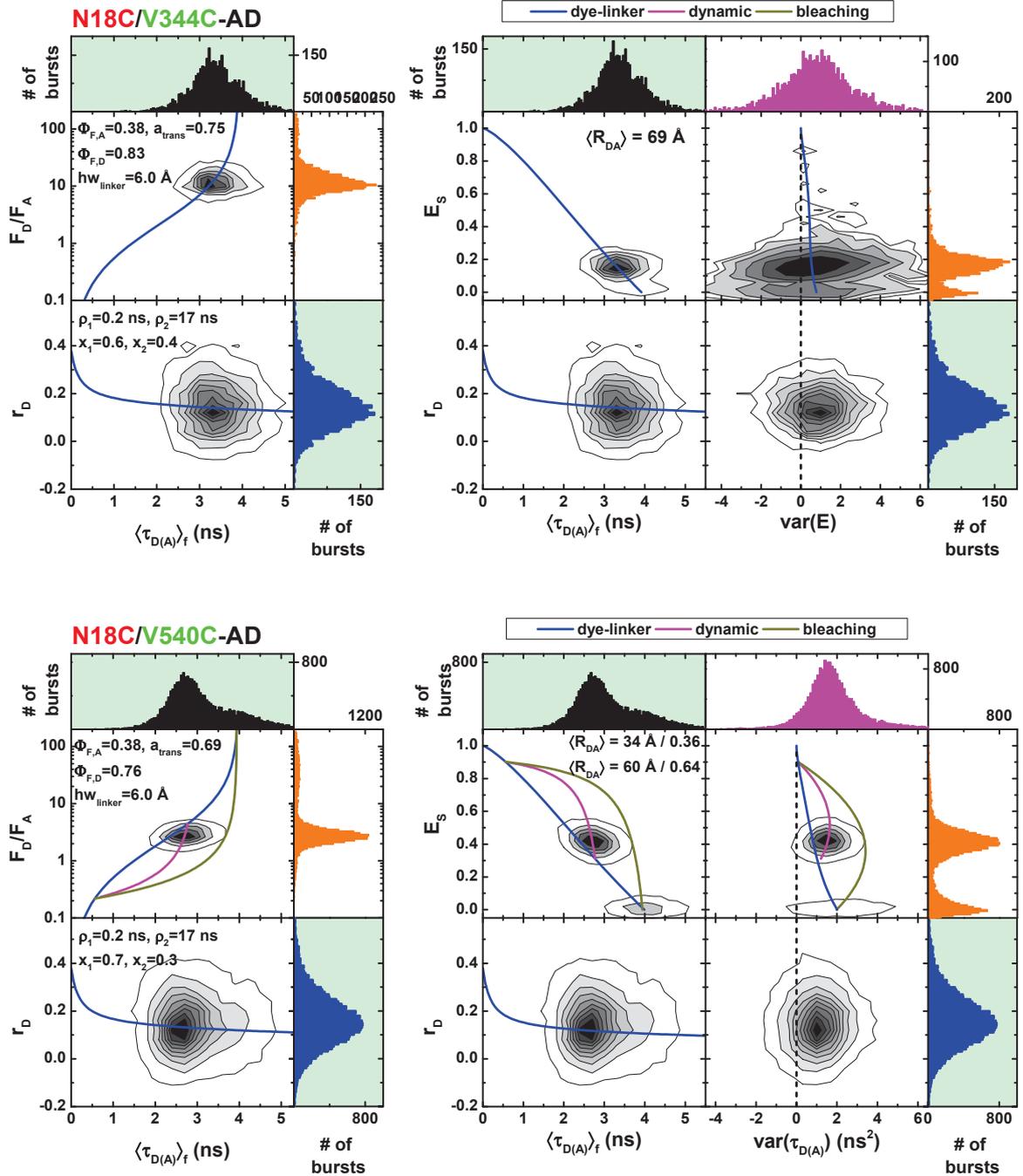
$$\chi_{Exp}^2 = \sum_{i=1}^N \left( \frac{R_{DA}^{(i)} - R_{\text{model}}^{(i)}}{\Delta R_{\pm}^{(i)}} \right)^2 \quad (14)$$

$R_{DA}^{(i)}$ , and  $R_{\text{model}}^{(i)}$  correspond to the experimental distance and the model distance, respectively, while  $\Delta R_{\pm}^{(i)}$  is the uncertainty of the experimental distance which depends on the sign of the deviation. In addition to the experimental restraints we use an additional clash potential as a penalty function. As our simple model does not allow for bending of the rigid domains we deliberately use a simple very weak potential given by:

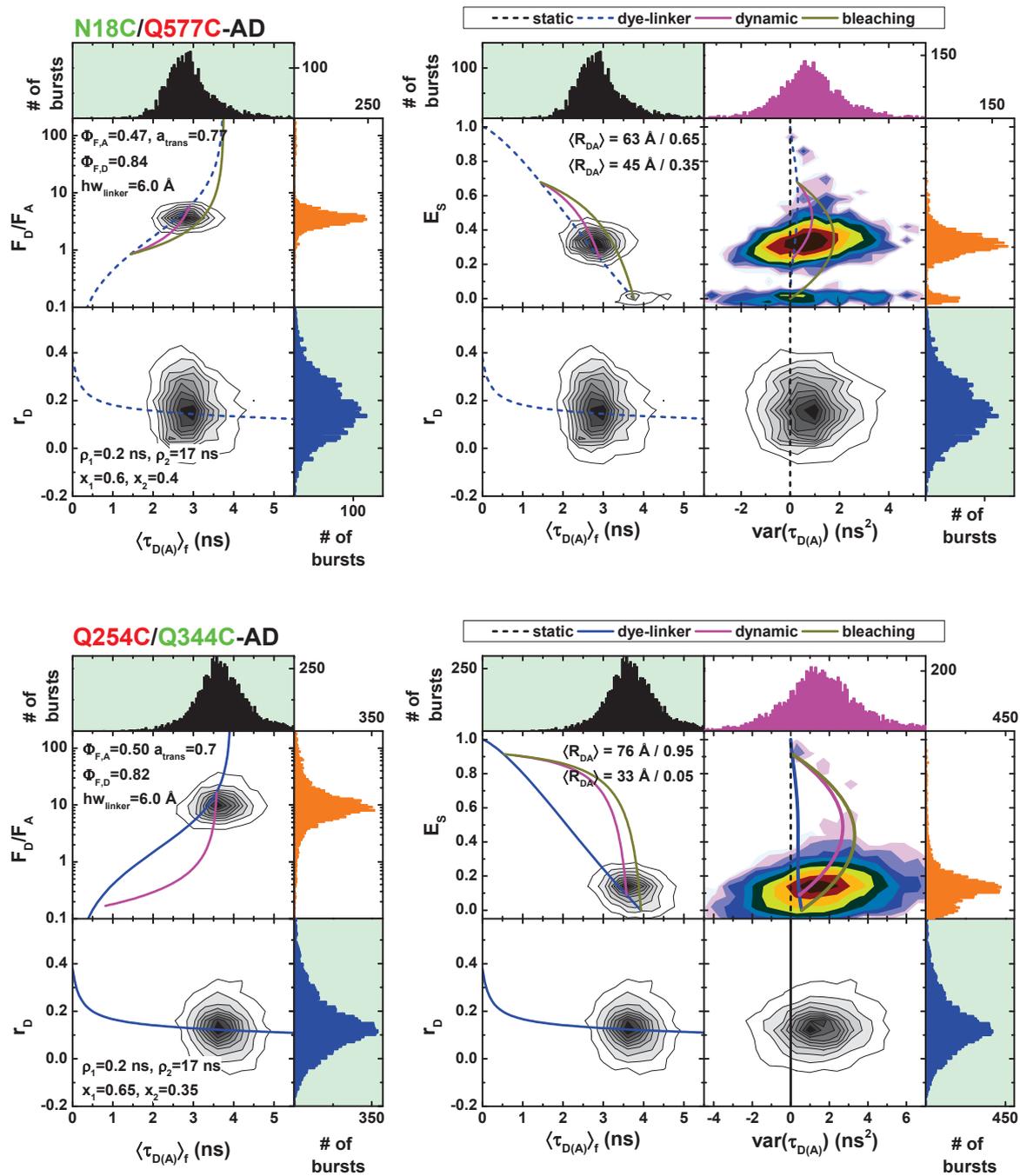
$$\chi_{\text{Clash}}^2 = \sum_{i,j} \begin{cases} 0 & , r_{ij} \geq r_{wi} + r_{wj} \\ (r_{wi} + r_{wj} - r_{ij})^2 / r_{\text{ctol}}^2 & , r_{ij} < r_{wi} + r_{wj} \end{cases} \quad (15)$$

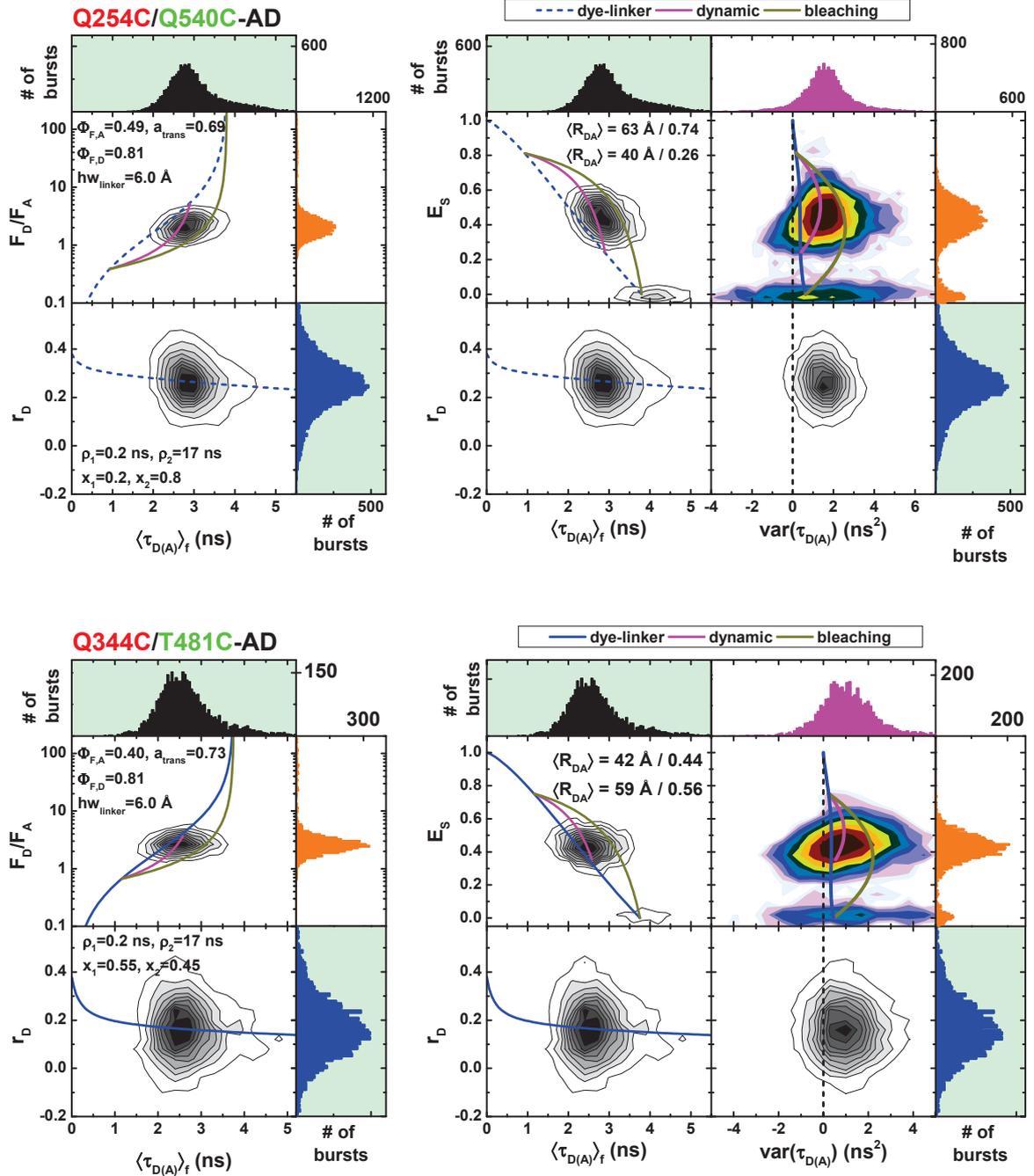
where  $r_{ij}$  is the distance between atoms  $i$  and  $j$  which belong to different subunits,  $r_{wi}$  and  $r_{wj}$  are their van der Waals radii, and  $r_{\text{ctol}}$  is the pre-defined clash tolerance. Here we used a clash tolerance of 6 Å. Finally, we introduce quadratic bonds of an equilibrium distance of 1.5 Å and weak spring constants  $k=0.25 \text{ 1/\AA}^2$  between the N- and C-terminus of each rigid body. As we use reduced energies spring constant is given in the unusual units  $[1/\text{\AA}^2]$ .

### 5.3 MFD-Histograms

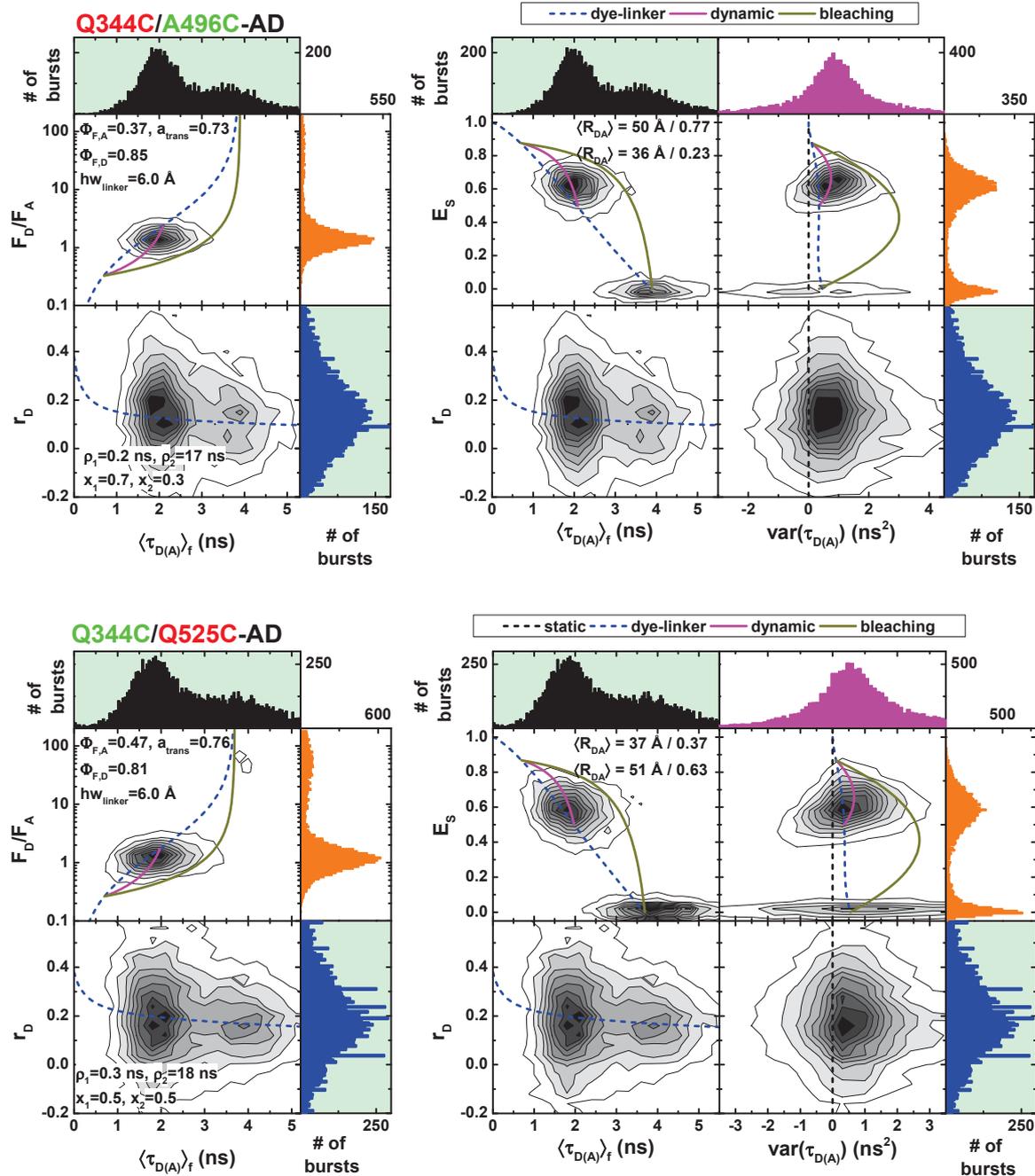


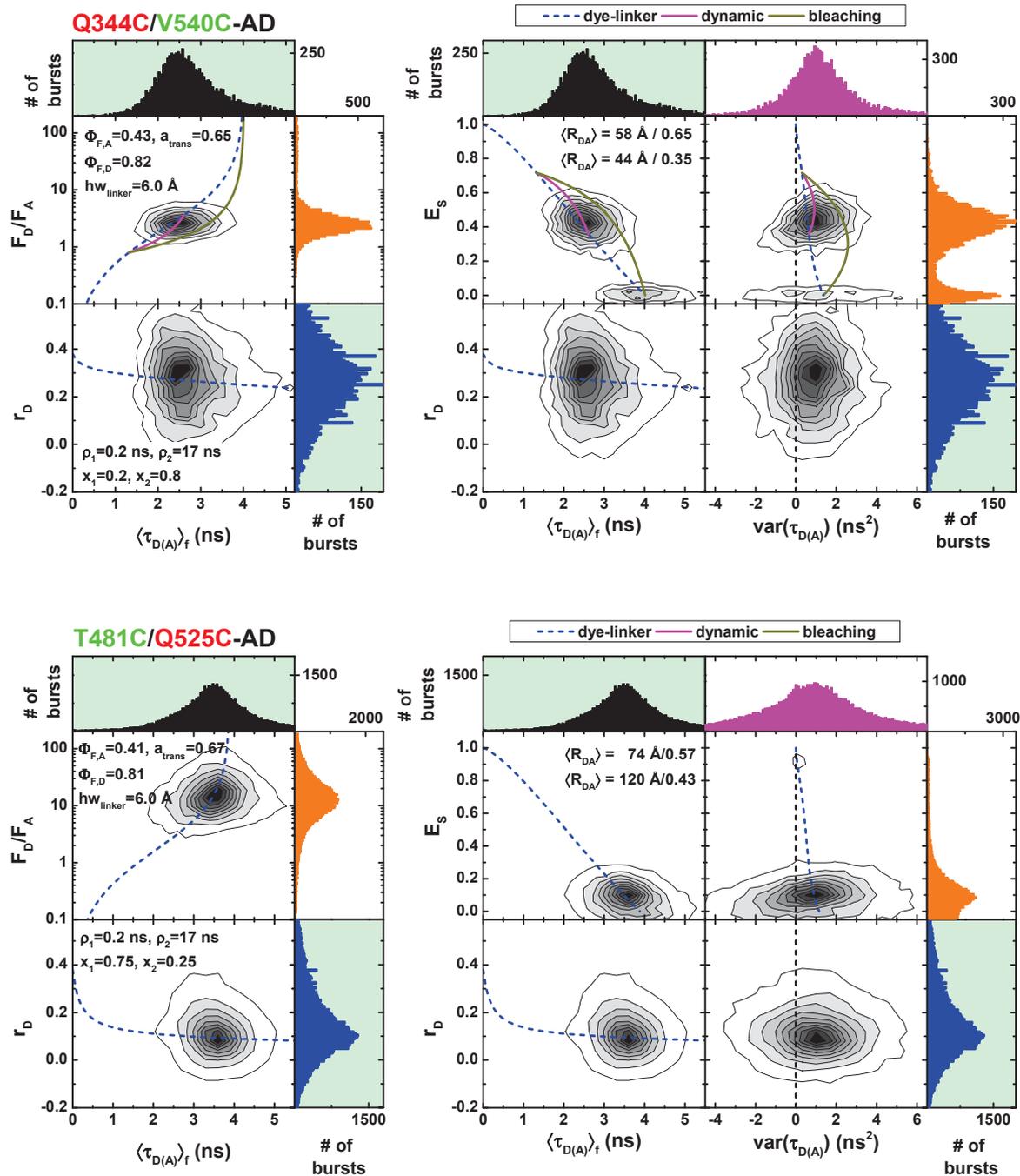
## F - Mapping the motions and structure of a state necessary for oligomerization of a large GTPase

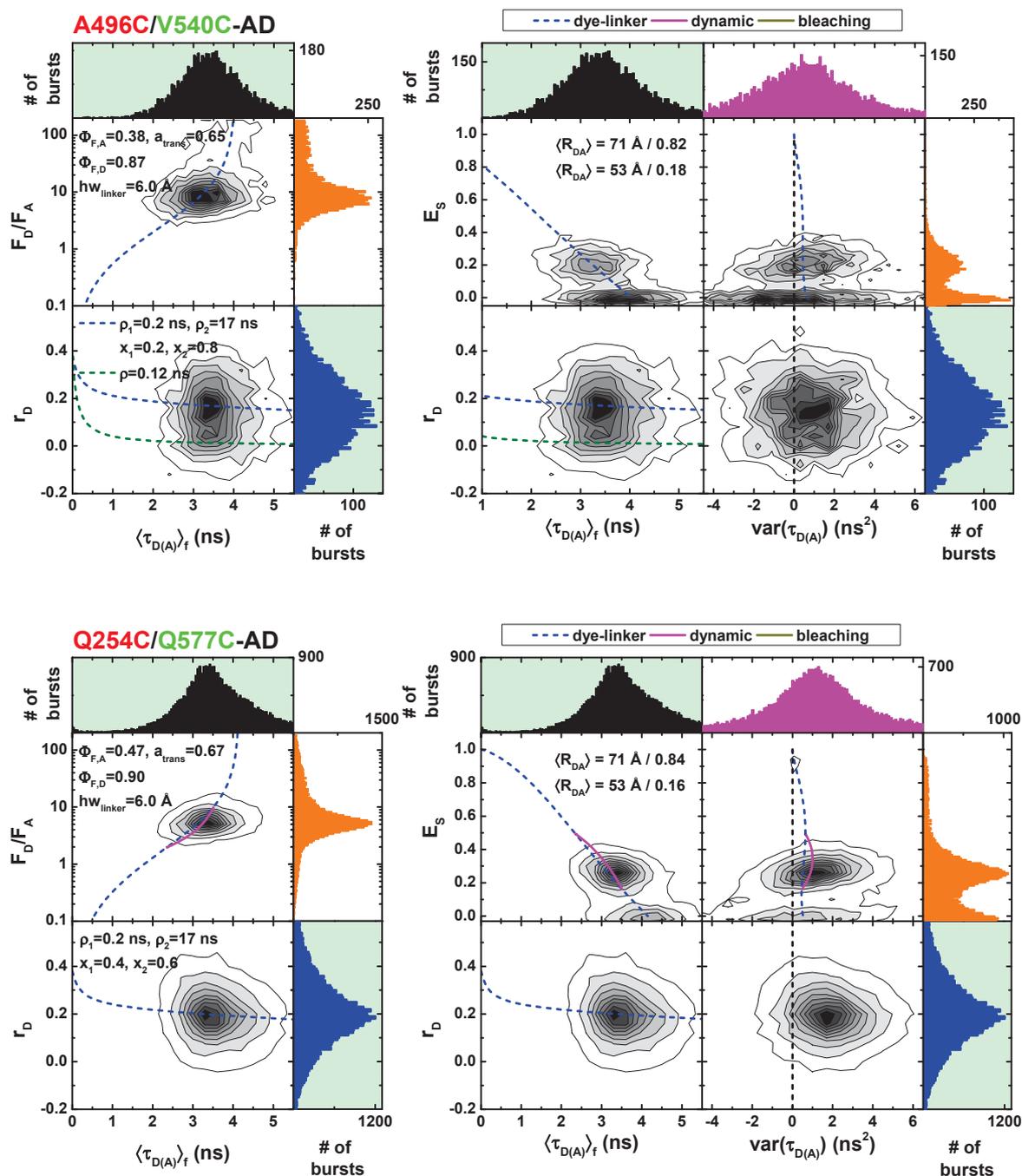




## F - Mapping the motions and structure of a state necessary for oligomerization of a large GTPase



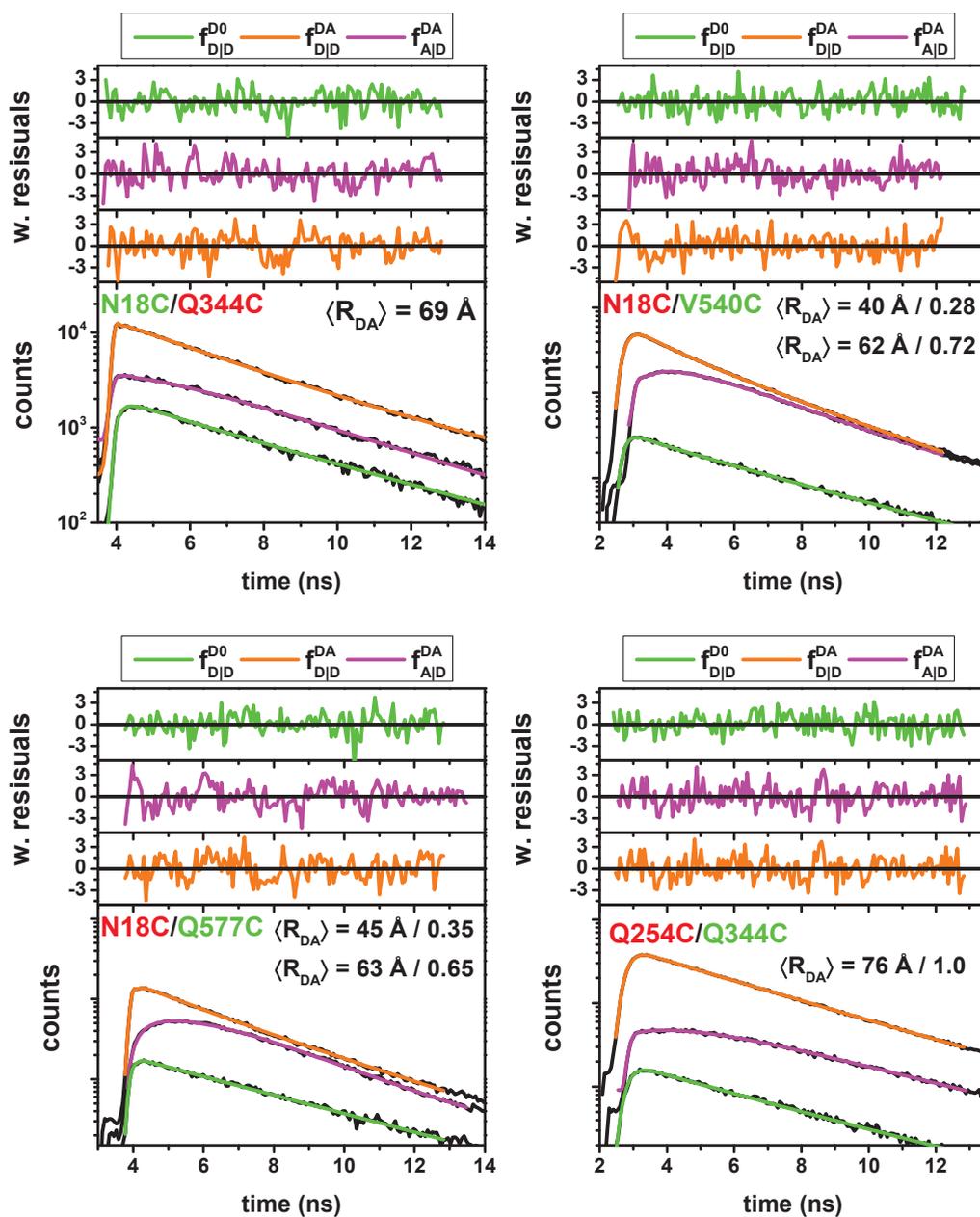


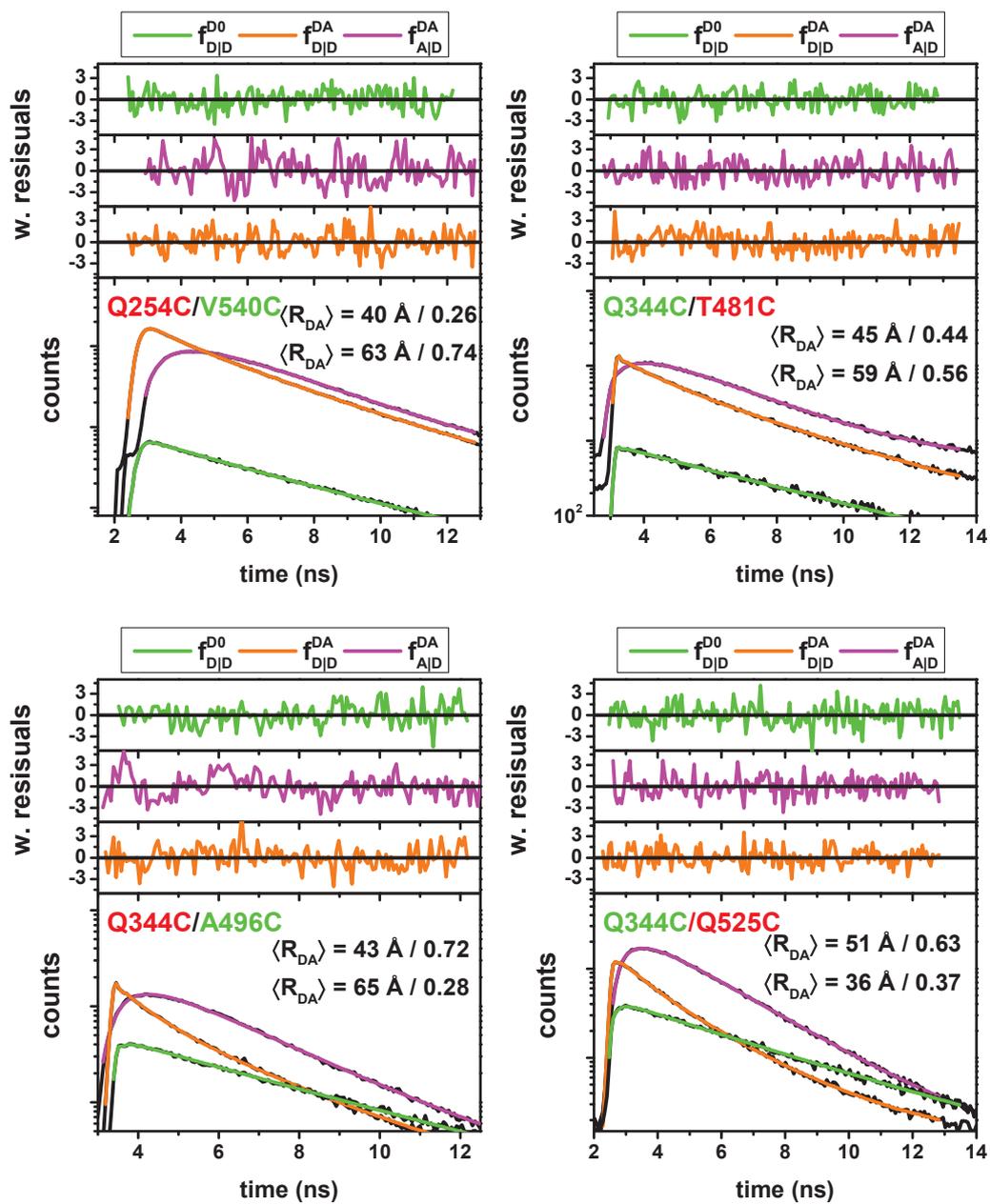


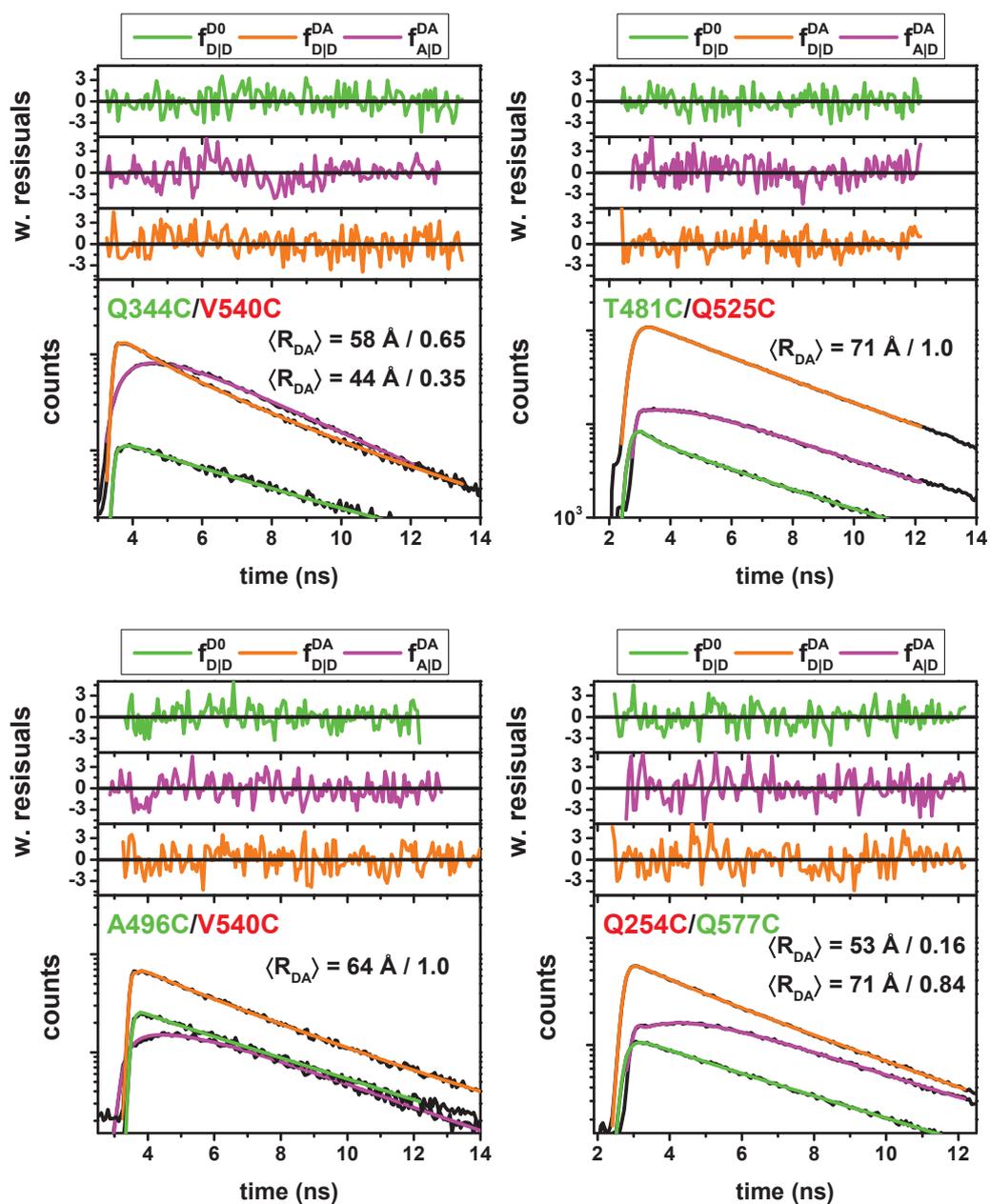
**Figure S2:** Dashed blue lines are either linker-dynamics corrected static FRET-lines or Perrin-equations. In Case two FRET-states were identified the dynamic FRET-line between these states is shown as magenta solid line. The dark-yellow line describes of acceptor bleaching from the high-FRET state. The FRET-data is displayed as the fluorescence intensity-ratio, the steady-state transfer-efficiency in dependence of the mean fluorescence averaged lifetime of the donor in presence of the acceptor  $\langle \tau_{D(A)} \rangle_f$  or the variance of the donor-acceptor lifetime  $var(\tau_{D(A)})$ . The color of the FRET-pair name indicated the most probable position of the donor (green) and acceptor dye (red), as inset the fluorescence quantum yield of the acceptor  $\Phi_{F,A}$  and the donor  $\Phi_{F,D}$  are shown. The fraction of the acceptor Alexa647 in cis-conformation was determined by FCS and is shown in the inset.

## 5.4 Sub-Ensemble TCPCSC

The donor fluorescence intensity decay and the FRET-sensitized acceptor emission were globally fitted.







**Figure S3:** Sub-ensemble fluorescence intensity decay analysis of the single-molecule measurements presented in Fig. 1. In green the fluorescence intensity decay of the donor-only fraction of the respective sample are shown. The donor-only fraction was selected by the acceptor intensity. In orange the time-resolved fluorescence intensity of the donor in presence of the acceptor is shown. In magenta the FRET-sensitized acceptor emission is displayed. The means of Gaussian distributed donor-acceptor distanced were determined by global analysis of all three decays. The analysis of the FRET-sensitized acceptor emission additionally requires the distribution of the lifetimes in the acceptor sample. The acceptor lifetimes were determined in separate measurements of acceptor only labeled samples (see below).

**Table S1.** Global fits time-resolved donor-fluorescence and FRET-sensitized acceptor emission of sub-ensemble data

FRET-sample	$R_{DA,1}$	$x_{RDA,1}$	$R_{DA,2}$	$x_{RDA,2}$	$x_{Donly}$
N18C/Q344C	69.3	1.00	-	-	0.18
N18C/V540C	60.0	0.50	34.1	0.50	0.37
N18C/Q577C	63.3	0.65	45.1	0.35	0.17
Q254C/Q344C	76.1	1.00	-	-	0.21
Q254C/V540C	63.4	0.74	39.3	0.26	0.31
Q344C/T481C	45.0	0.44	59.3	0.56	0.17
Q344C/A496C	47.0	0.77	36.0	0.23	0.45
Q344C/Q525C	51.5	0.63	36.1	0.37	0.40
Q344C/V540C	57.8	0.65	43.8	0.35	0.24
T481C/Q525C	70.6	1.00	-	-	0.20
A496C/V540C	63.9	1.00	-	-	0.37
Q254C/Q577C	70.8	0.84	52.9	0.16	0.34

## 5.5 Ensemble TCSPC

### Acceptor-only

**Table S2.** Fluorescence lifetimes of Alexa647N determined by ensemble TCSPC

Acceptor	$x_1$	$\tau_1$ [ns]	$x_2$	$\tau_2$ [ns]	$x_3$	$\tau_3$ [ns]	$\langle \tau \rangle_x$ [ns]	$\Phi_{F,A}$
N18C	0.39	1.85	0.49	1.22	0.12	0.10	1.33	0.40
Q254C	0.58	2.23	0.42	1.42			1.89	0.57
Q344C	0.58	2.06	0.42	1.09			1.75	0.52
T481C	0.43	1.89	0.57	1.32			1.57	0.43
A496C	0.43	1.21	0.57	1.88			1.59	0.48
Q525C	<b>0.65</b>	<b>1.93</b>	<b>0.35</b>	<b>1.08</b>			<b>1.63</b>	<b>0.49</b>
V540C	0.65	2.33	0.35	1.43			2.02	0.60
Q577C	0.49	2.06	0.51	1.42			1.73	0.52

The quantum yield was calculated using the species averaged lifetime and the lifetime and quantum yield of the free dye 1.0 ns  $QY=0.32$  as a reference.

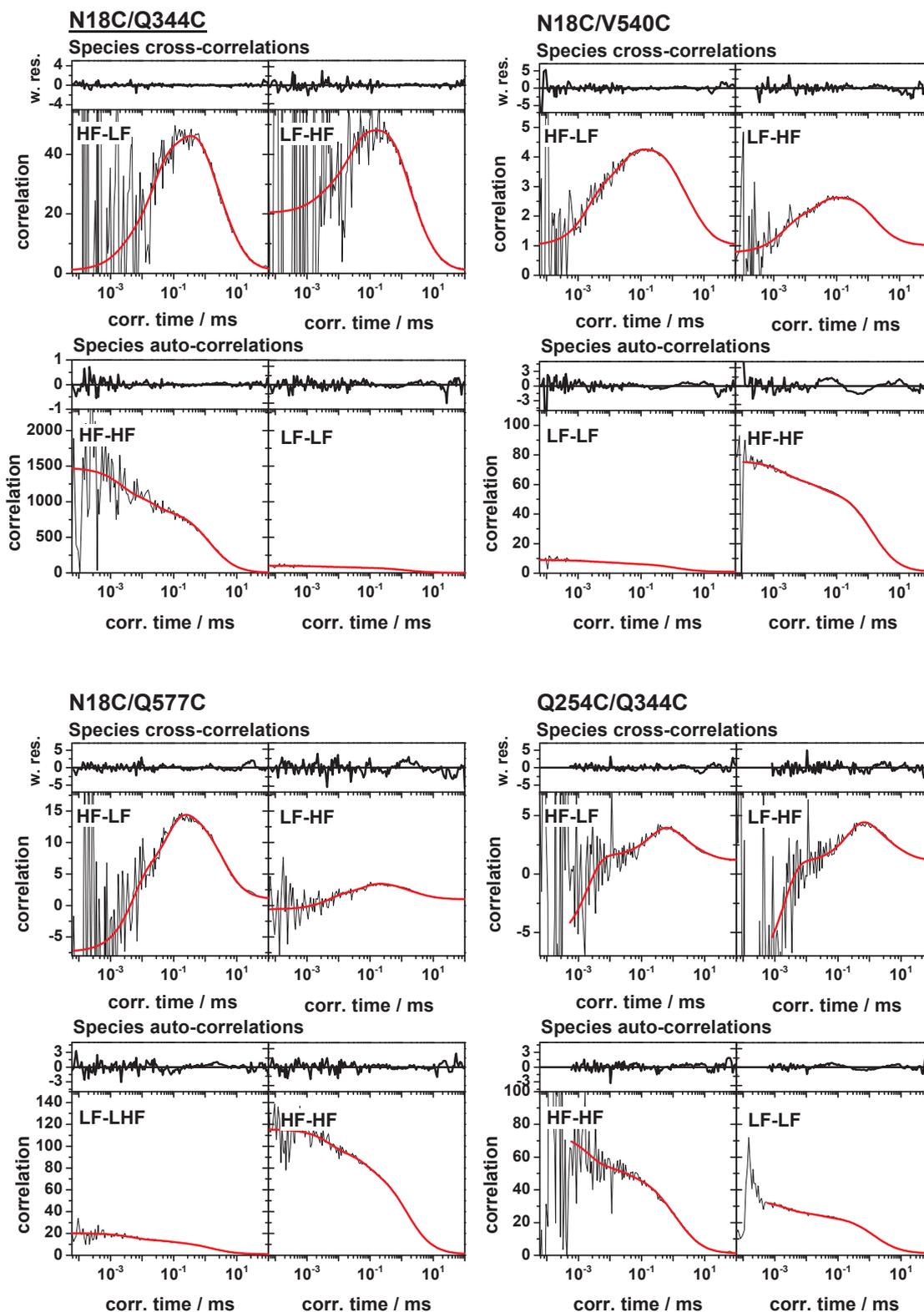
### Donor-only

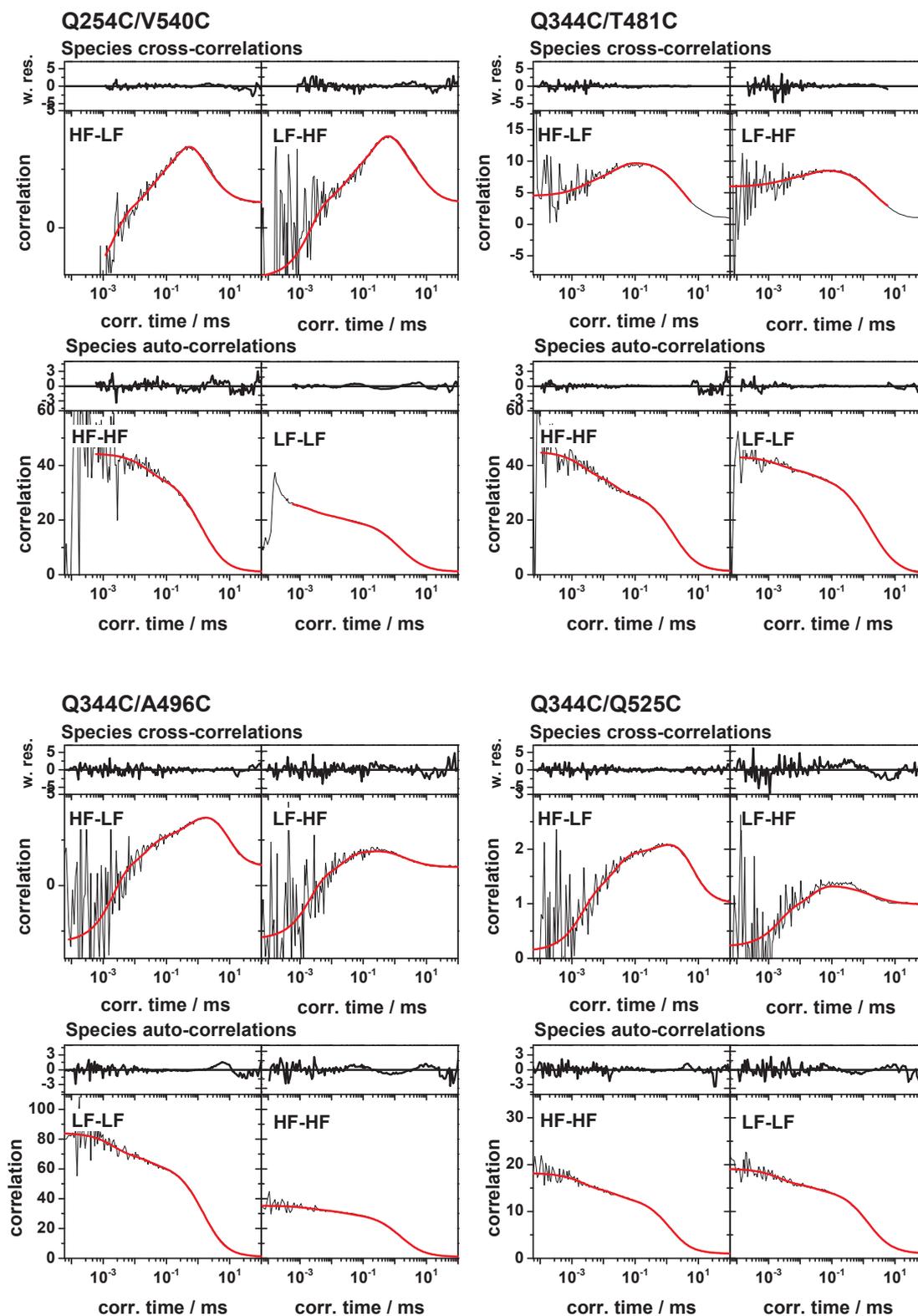
**Table S3.** Fluorescence lifetimes of Alexa488 determined by ensemble TCSPC

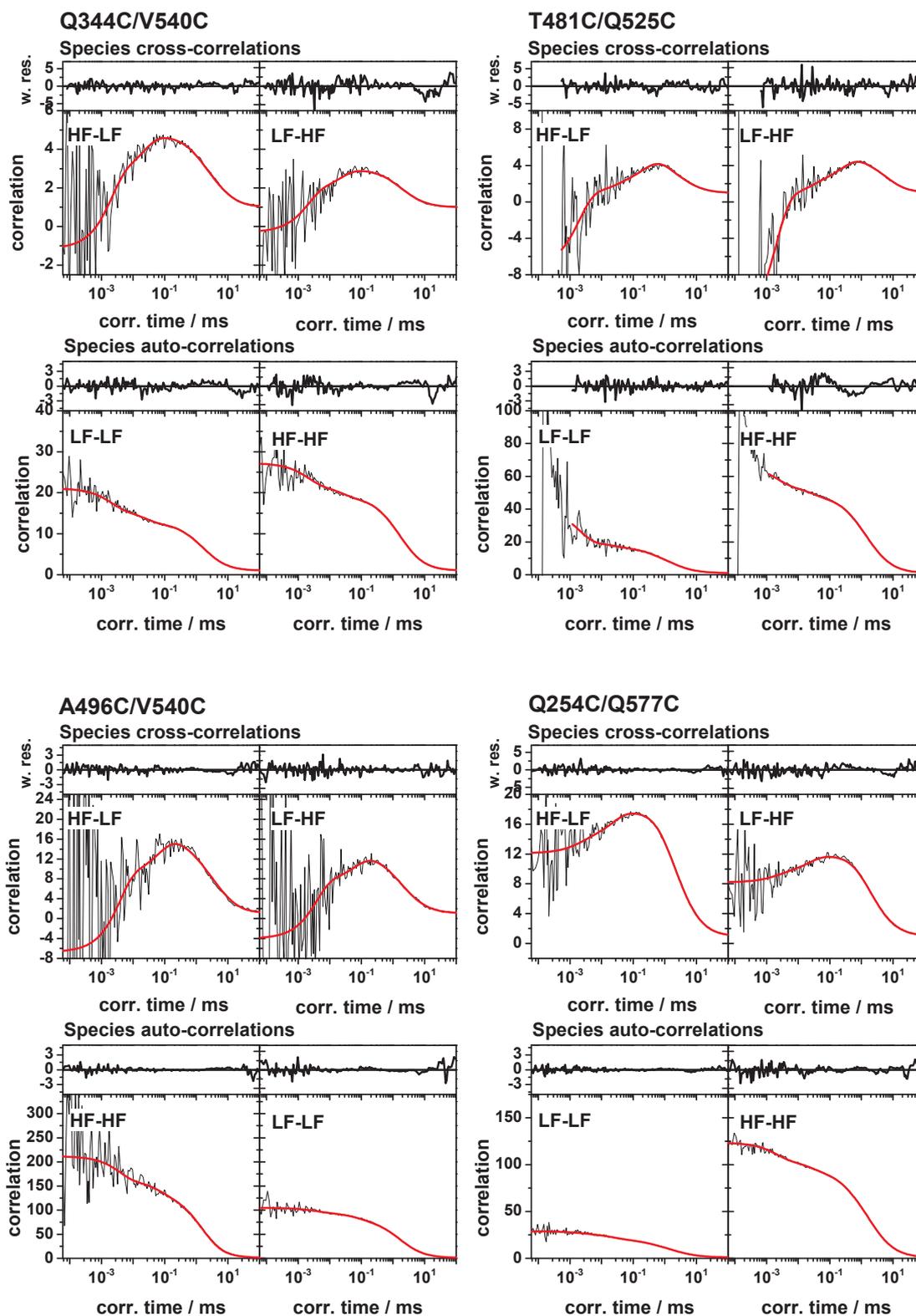
Donor	$x_1$	$\tau_1$ [ns]	$x_2$	$\tau_2$ [ns]	$\langle \tau \rangle_x$ [ns]	$\Phi_{F,D}$
N18C	0.82	4.15	0.18	1.35	3.65	0.82
Q254C	0.69	3.60	0.31	0.53	2.65	0.59
Q344C	0.94	3.80	0.06	1.00	3.63	0.81
T481C*	0.93	3.78	0.07	1.07	3.59	0.81
A496C*	0.84	3.89	0.16	1.14	3.44	0.77
Q525C	0.80	3.51	0.20	0.66	2.94	0.66
V540C*	0.85	4.00	0.15	1.86	3.67	0.82
Q577C	0.91	4.15	0.09	1.49	3.91	0.88

The quantum yield was calculated using the species averaged lifetime and the lifetime and quantum yield of the free dye 4.1 ns  $\Phi_{F,D}=0.92$  as a reference.

### 5.6 Filtered FCS







**Figure S4:** Global analysis of all 12 hGBP1 FRET-variants by lifetime filtered correlation. Overall all 12 FRET-sample were fitted by a global model with cross-correlation times of 297.6  $\mu$ s, 22.6  $\mu$ s and 2.0  $\mu$ s. In the analysis the cross-correlation times were global among all dataset whereas the amplitudes of the correlation times were shared among each FRET-pair. The low-FRET (LF) to high-FRET (HF) cross-correlation was additionally fitted by a relaxation time that accounts for bleaching of the HF-state.

### 5.7 Error estimation

The estimated error reported in Table 1 includes statistical uncertainties, potential systematic errors of the donor references and uncertainties of the accessible volumes due to the differences of the donor and acceptor linker-length.

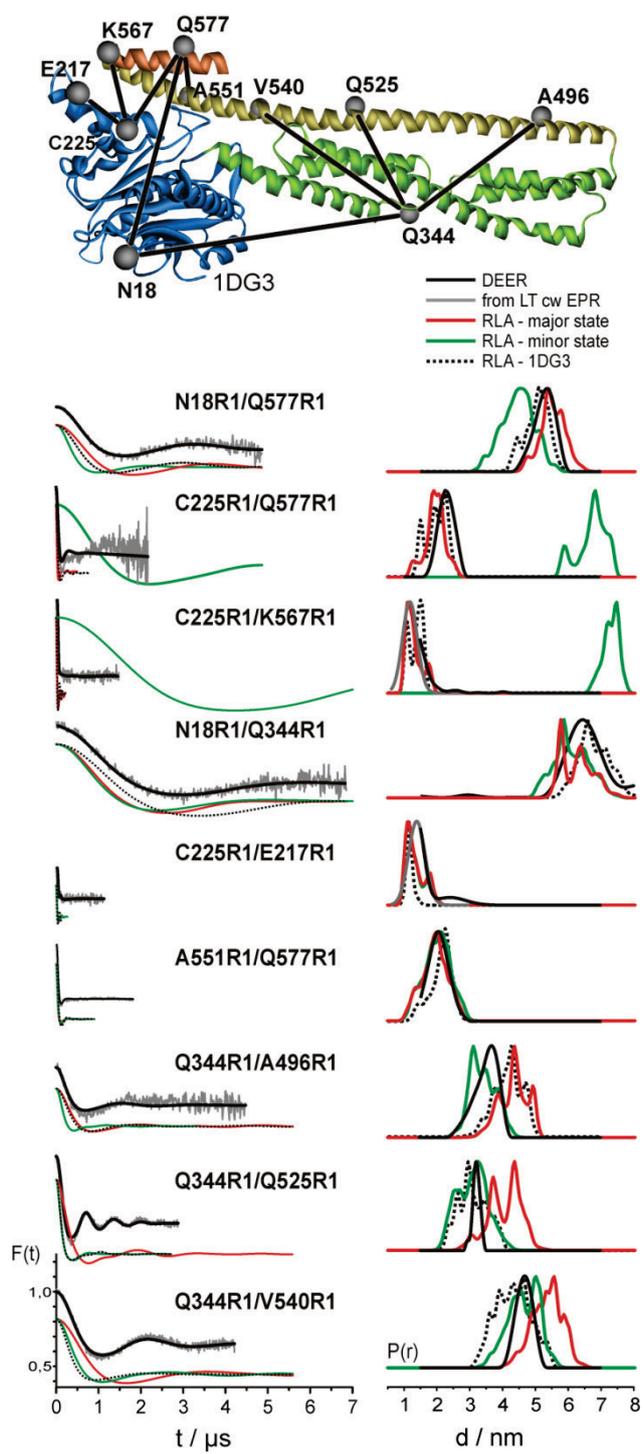
The statistical errors were determined by sampling the parameter space of the model function as described in the main text. Systematic errors were estimated by assuming that a relative error of the donor-reference lifetime. For a given donor-acceptor distance  $R_{DA}$  and Förster-radius  $R_0$  a relative error of the donor-acceptor distance can be estimated as previously described (see Chapter B) and is given by:

$$\frac{\Delta R_{DA}}{R_{DA}} = \left( 1 - \left( 1 + \frac{R_{DA}}{R_0} \right)^6 \delta\tau_0 \right)^{-1/6}$$

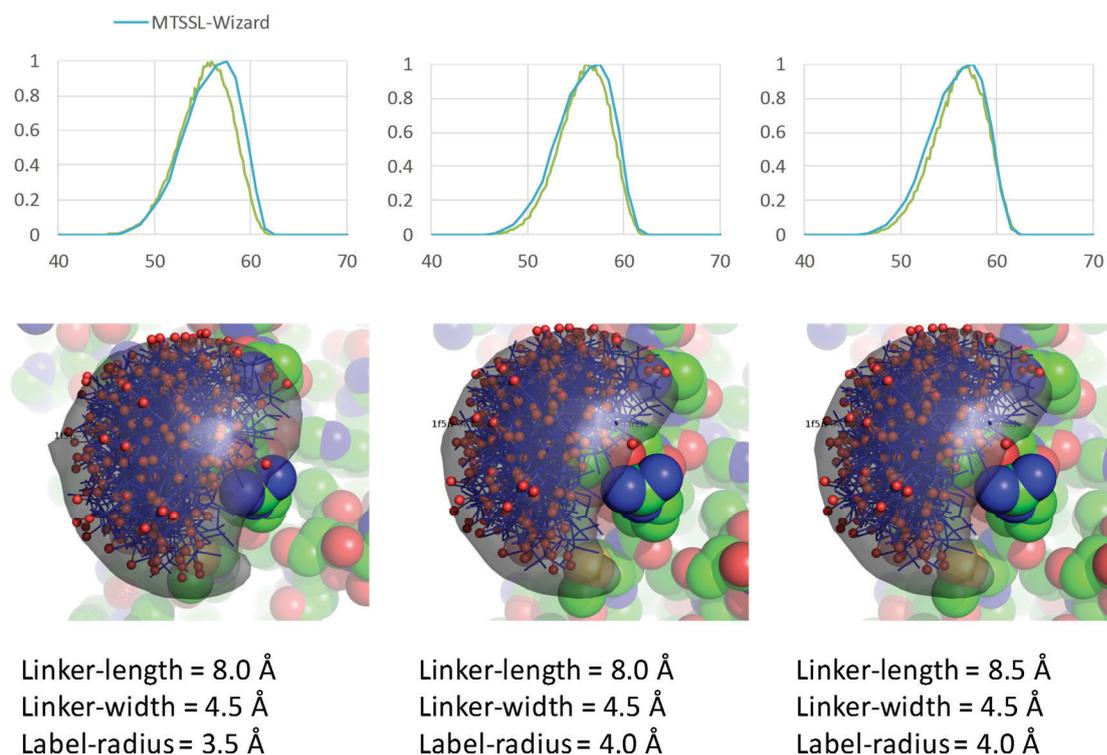
Here,  $\delta\tau_0$  is the relative error of the reference which was assumed to be 4%. The uncertainties regarding the orientation factor were estimated as described in Supplement A. Potential systematic error due to difference in the linker length of the donor and the acceptor fluorophore were considered by multiple accessible volume calculations as described in Chapter B. In average this results in an uncertainty of  $\sim 0.8 \text{ \AA}$ .

### 5.8 EPR-measurements

**Figure S5.** (top) Measured EPR-distance network. (left) Experimental DEER-traces and simulated DEER-traces based on a rotamer library analysis of in different structural models (right) Inter-spin distance distribution as determined by Tikhonov regularization of the experimental DEER-trace.



## 5.9 Parametrization of MTSSL for Accessible volume calculations



**Figure S2.** Parametrization of the EPR-MTSSL label for accessible volume calculations. Top the distance distributions for the spin-pair N18C/Q577C of the hGBP1 crystal structure (PDB-ID: 1DG3) as calculated by the MTSSL-Wizard (Hagelueken et al., 2012) is overlaid by the distance distribution as calculated by accessible volume calculations with the parameter set as provided below. For visual comparison the rotamers are overlaid with the accessible volume calculated for the labeling position N18C.

To parameterize the MTSSL-label we used the variant N18C/Q577C as reference as optimized the simulated linker-length, the “dye”-radius and the linker-width until the distance distribution as determined by the AV-calculations agrees with the distance distributions as determined by the MTSSL-Wizard (Hagelueken et al., 2012). The best agreement was found using a linker-length of 8.5 Å, a linker-width of 4.5 Å and a label-radius of 4.0 Å. All rigid body dockings were performed using this parameter set.

## References

- AHMED, A. & GOHLKE, H. 2006. Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. *Proteins-Structure Function and Bioinformatics*, 63, 1038-1051.
- AHMED, A., RIPPMANN, F., BARNICKEL, G. & GOHKE, H. 2011. A Normal Mode-Based Geometric Simulation Approach for Exploring Biologically Relevant Conformational Transitions in Proteins. *Journal of Chemical Information and Modeling*, 51, 1604-1622.
- AHMED, A., VILLINGER, S. & GOHLKE, H. 2010. Large-scale comparison of protein essential dynamics from molecular dynamics simulations and coarse-grained normal mode analyses. *Proteins-Structure Function and Bioinformatics*, 78, 3341-3352.

- ANDERSON, S. L., CARTON, J. M., LOU, J., XING, L. & RUBIN, B. Y. 1999. Interferon-induced guanylate binding protein-1 (GBP-1) mediates an antiviral effect against vesicular stomatitis virus and encephalomyocarditis virus. *Virology*, 256, 8-14.
- BAS, D. C., ROGERS, D. M. & JENSEN, J. H. 2008. Very fast prediction and rationalization of pK(a) values for protein-ligand complexes. *Proteins-Structure Function and Bioinformatics*, 73, 765-783.
- BENNETT, M. J., SCHLUNEGGER, M. P. & EISENBERG, D. 1995. 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci*, 4, 2455-68.
- BRITZEN-LAURENT, N., BAUER, M., BERTON, V., FISCHER, N., SYGUDA, A., REIPSCHLAGER, S., NASCHBERGER, E., HERRMANN, C. & STURZL, M. 2010. Intracellular Trafficking of Guanylate-Binding Proteins Is Regulated by Heterodimerization in a Hierarchical Manner. *Plos One*, 5.
- CHUNG, H. S., MCHALE, K., LOUIS, J. M. & EATON, W. A. 2012. Single-Molecule Fluorescence Experiments Determine Protein Folding Transition Path Times. *Science*, 335, 981-984.
- D.A. CASE, J. T. B., R.M. BETZ, D.S. CERUTTI, T.E. CHEATHAM, III, T.A. DARDEN, R.E. DUKE, T.J. GIESE, H. GOHLKE, A.W. GOETZ, N. HOMEYER, S. IZADI, P. JANOWSKI, J. KAUS, A. KOVALENKO, T.S. LEE, S. LEGRAND, P. LI, T. LUCHKO, R. LUO, B. MADEJ, K.M. MERZ, G. MONARD, P. NEEDHAM, H. NGUYEN, H.T. NGUYEN, I. OMELYAN, A. ONUFRIEV, D.R. ROE, A. ROITBERG, R. SALOMON-FERRER, C.L. SIMMERLING, W. SMITH, J. SWAILS, R.C. WALKER, J. WANG, R.M. WOLF, X. WU, D.M. YORK AND P.A. KOLLMAN. 2015. *Amber* [Online]. Available: <http://ambermd.org/> [Accessed].
- DARDEN, T., YORK, D. & PEDERSEN, L. 1993. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics*, 98, 10089-10092.
- EGGELING, C., BERGER, S., BRAND, L., FRIES, J. R., SCHAFFER, J., VOLKMER, A. & SEIDEL, C. A. M. 2001. Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *Journal of Biotechnology*, 86, 163-180.
- FELEKYAN, S., KALININ, S., SANABRIA, H., VALERI, A. & SEIDEL, C. A. M. 2012. Filtered FCS: Species Auto- and Cross-Correlation Functions Highlight Binding and Dynamics in Biomolecules. *Chemphyschem*, 13, 1036-1053.
- FELEKYAN, S., SANABRIA, H., KALININ, S., KUHNEMUTH, R. & SEIDEL, C. A. M. 2013. Analyzing Forster Resonance Energy Transfer with Fluctuation Algorithms. *Fluorescence Fluctuation Spectroscopy (Ffs)*, Pt B, 519, 39-85.
- FOREMAN-MACKEY, D., HOGG, D. W., LANG, D. & GOODMAN, J. 2013. emcee: The MCMC hammer. *Publications of the Astronomical Society of the Pacific*, 125, 306-312.
- GAMBIN, Y., VANDELINDER, V., FERREON, A. C. M., LEMKE, E. A., GROISMAN, A. & DENIZ, A. A. 2011. Visualizing a one-way protein encounter complex by ultrafast single-molecule mixing. *Nature Methods*, 8, 239-U77.
- GHOSH, A., PRAEFCKE, G. J. K., RENAULT, L., WITTINGHOFER, A. & HERRMANN, C. 2006. How guanylate-binding proteins achieve assembly-stimulated processive cleavage of GTP to GMP. *Nature*, 440, 101-104.
- GILL, S. C. & VONHIPPEL, P. H. 1989. Calculation of Protein Extinction Coefficients from Amino-Acid Sequence Data. *Analytical Biochemistry*, 182, 319-326.
- GOODMAN, J. & WEARE, J. 2010. Ensemble Samplers with Affine Invariance. *Communications in Applied Mathematics and Computational Science*, 5, 65-80.
- HAGELUEKEN, G., WARD, R., NAISMITH, J. H. & SCHIEMANN, O. 2012. MtsslWizard: In Silico Spin-Labeling and Generation of Distance Distributions in PyMOL. *Applied Magnetic Resonance*, 42, 377-391.

- HAMELBERG, D., MONGAN, J. & MCCAMMON, J. A. 2004. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *Journal of Chemical Physics*, 120, 11919-11929.
- HASHIMOTO, K., NISHI, H., BRYANT, S. & PANCHENKO, A. R. 2011. Caught in self-interaction: evolutionary and functional mechanisms of protein homooligomerization. *Physical Biology*, 8.
- HOFIG, H., GABBA, M., POBLETE, S., KEMPE, D. & FITTER, J. 2014. Inter-dye distance distributions studied by a combination of single-molecule FRET-filtered lifetime measurements and a weighted accessible volume (wAV) Algorithm. *Molecules*, 19, 19269-19291.
- HOPKINS, C. W., LE GRAND, S., WALKER, R. C. & ROITBERG, A. E. 2015. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation*, 11, 1864-1874.
- ITSUI, Y., SAKAMOTO, N., KAKINUMA, S., NAKAGAWA, M., SEKINE-OSAJIMA, Y., TASAKA-FUJITA, M., NISHIMURA-SAKURAI, Y., SUDA, G., KARAKAMA, Y., MISHIMA, K., YAMAMOTO, M., WATANABE, T., UEYAMA, M., FUNAOKA, Y., AZUMA, S. & WATANABE, M. 2009. Antiviral Effects of the Interferon-Induced Protein Guanylate Binding Protein 1 and Its Interaction with the Hepatitis C Virus NS5B Protein. *Hepatology*, 50, 1727-1737.
- ITSUI, Y., SAKAMOTO, N., KUROSAKI, M., KANAZAWA, N., TANABE, Y., KOYAMA, T., TAKEDA, Y., NAKAGAWA, M., KAKINUMA, S., SEKINE, Y., MAEKAWA, S., ENOMOTO, N. & WATANABE, M. 2006. Expressional screening of interferon-stimulated genes for antiviral activity against hepatitis C virus replication. *Journal of Viral Hepatitis*, 13, 690-700.
- JACOBS, D. J., RADER, A. J., KUHN, L. A. & THORPE, M. F. 2001. Protein flexibility predictions using graph theory. *Proteins-Structure Function and Genetics*, 44, 150-165.
- JORGENSEN, W. L., CHANDRASEKHAR, J., MADURA, J. D., IMPEY, R. W. & KLEIN, M. L. 1983. Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics*, 79, 926-935.
- KALININ, S., FULLE, S., KÜHNEMUTH, R., SINDBERT, S., FELEKYAN, S., GOHLKE, H. & SEIDEL, C. A. M. 2015. Diffusion with traps: experiment, simulation, and theory to describe fluorophore dynamics in biomolecular FRET.
- KALININ, S., PEULEN, T., SINDBERT, S., ROTHWELL, P. J., BERGER, S., RESTLE, T., GOODY, R. S., GOHLKE, H. & SEIDEL, C. A. M. 2012a. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat Methods*, 9, 1218-25.
- KALININ, S., PEULEN, T., SINDBERT, S., ROTHWELL, P. J., BERGER, S., RESTLE, T., GOODY, R. S., GOHLKE, H. & SEIDEL, C. A. M. 2012b. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat Methods*, 9, 1218-25.
- KALININ, S., PEULEN, T., SINDBERT, S., ROTHWELL, P. J., BERGER, S., RESTLE, T., GOODY, R. S., GOHLKE, H. & SEIDEL, C. A. M. 2012c. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nature Methods*, 9, 1218-25.
- KALININ, S., VALERI, A., ANTONIK, M., FELEKYAN, S. & SEIDEL, C. A. M. 2010a. Detection of structural dynamics by FRET: a photon distribution and fluorescence lifetime analysis of systems with multiple states. *J Phys Chem B*, 114, 7983-95.
- KALININ, S., VALERI, A., ANTONIK, M., FELEKYAN, S. & SEIDEL, C. A. M. 2010b. Detection of Structural Dynamics by FRET: A Photon Distribution and Fluorescence

- Lifetime Analysis of Systems with Multiple States. *Journal of Physical Chemistry B*, 114, 7983-7995.
- KOLLNER, M. & WOLFRUM, J. 1992. How Many Photons Are Necessary for Fluorescence-Lifetime Measurements. *Chemical Physics Letters*, 200, 199-204.
- KOSHIOKA, M., SASAKI, K. & MASUHARA, H. 1995. Time-Dependent Fluorescence Depolarization Analysis in 3-Dimensional Microspectroscopy. *Applied Spectroscopy*, 49, 224-228.
- KRAVETS, E., DEGRANDI, D., MA, Q., PEULEN, T., FELEKYAN, S., KÜHNEMUTH, R., WEIDTKAMP-PETERS, S., SEIDEL, C. A. M. & PFEFFER, K. in revision. Guanylate binding proteins (GBPs) directly attack *T. gondii* via supramolecular complexes. *eLife*.
- KRAVETS, E., DEGRANDI, D., WEIDTKAMP-PETERS, S., RIES, B., KONERMANN, C., FELEKYAN, S., DARGAZANLI, J. M., PRAEFCKE, G. J. K., SEIDEL, C. A. M., SCHMITT, L., SMITS, S. H. J. & PFEFFER, K. 2012. The GTPase Activity of Murine Guanylate-binding Protein 2 (mGBP2) Controls the Intracellular Localization and Recruitment to the Parasitophorous Vacuole of *Toxoplasma gondii*. *Journal of Biological Chemistry*, 287, 27452-27466.
- KUNZELMANN, S., PRAEFCKE, G. J. K. & HERRMANN, C. 2006. Transient kinetic investigation of GTP hydrolysis catalyzed by interferon-gamma-induced hGBP1 (human guanylate binding protein 1). *Journal of Biological Chemistry*, 281, 28627-28635.
- LEE, C. S., KIM, Y. J., KIM, W. & MYUNG, S. C. 2011. Guanylate Cyclase Activator YC-1 Enhances TRAIL-Induced Apoptosis in Human Epithelial Ovarian Carcinoma Cells via Activation of Apoptosis-Related Proteins. *Basic & Clinical Pharmacology & Toxicology*, 109, 283-291.
- MAUS, M., COTLET, M., HOFKENS, J., GENSCHE, T., DE SCHRYVER, F. C., SCHAFFER, J. & SEIDEL, C. A. M. 2001. An experimental comparison of the maximum likelihood estimation and nonlinear least squares fluorescence lifetime analysis of single molecules. *Analytical Chemistry*, 73, 2078-2086.
- MORÉ, J. J., GARBOW, B. S. & HILLSTROM, K. E. 1980. User Guide for MINPACK-1. ANL-80-74, Argonne National Laboratory.
- PHILLIPS, D., DRAKE, R. C., OCONNOR, D. V. & CHRISTENSEN, R. L. 1985. Time Correlated Single-Photon Counting (Tcspc) Using Laser Excitation. *Analytical Instrumentation*, 14, 267-292.
- PIERCE, L. C. T., SALOMON-FERRER, R., DE OLIVEIRA, C. A. F., MCCAMMON, J. A. & WALKER, R. C. 2012. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *Journal of Chemical Theory and Computation*, 8, 2997-3002.
- PRAEFCKE, G. J. K., GEYER, M., SCHWEMMLE, M., KALBITZER, H. R. & HERRMANN, C. 1999a. Nucleotide-binding characteristics of human guanylate-binding protein 1 (hGBP1) and identification of the third GTP-binding motif. *Journal of Molecular Biology*, 292, 321-332.
- PRAEFCKE, G. J. K. & MCMAHON, H. T. 2004. The dynamin superfamily: Universal membrane tubulation and fission molecules? *Nature Reviews Molecular Cell Biology*, 5, 133-147.
- PRAEFCKE, G. J. K., WITTINGHOFER, A. & HERRMANN, C. 1999b. Mechanism of nucleotide binding and hydrolysis of human guanylate binding protein 1 (hGBP1). *Molecular Biology of the Cell*, 10, 315a-315a.
- PRAKASH, B., PRAEFCKE, G. J. K., RENAULT, L., WITTINGHOFER, A. & HERRMANN, C. 2000a. Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins. *Nature*, 403, 567-571.

- PRAKASH, B., RENAULT, L., PRAEFCKE, G. J. K., HERRMANN, C. & WITTINGHOFER, A. 2000b. Triphosphate structure of guanylate-binding protein 1 and implications for nucleotide binding and GTPase mechanism. *Embo Journal*, 19, 4555-4564.
- ROTHWELL, P. J., ALLEN, W. J., SISAMAKIS, E., KALININ, S., FELEKYAN, S., WIDENGREN, J., WAKSMAN, G. & SEIDEL, C. A. M. 2013. dNTP-dependent Conformational Transitions in the Fingers Subdomain of KlenTaq1 DNA Polymerase INSIGHTS INTO THE ROLE OF THE "NUCLEOTIDE-BINDING" STATE. *Journal of Biological Chemistry*, 288, 13575-13591.
- RYCKAERT, J. P., CICCOTTI, G. & BERENDSEN, H. J. C. 1977. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *Journal of Computational Physics*, 23, 327-341.
- SCHAEFER, D. W. 1973. Dynamics of number fluctuations: motile microorganisms. *Science*, 180, 1293-5.
- SCHAFFER, J., VOLKMER, A., EGGELING, C., SUBRAMANIAM, V., STRIKER, G. & SEIDEL, C. A. M. 1999. Identification of single molecules in aqueous solution by time-resolved fluorescence anisotropy. *Journal of Physical Chemistry A*, 103, 331-336.
- SCHWEMMLE, M. & STAEHELI, P. 1994. The interferon-induced 67-KDa guanylate-binding protein (hGBP1) is a GTPase that converts GTP to GMP. *Journal of Biological Chemistry*, 269, 11299-11305.
- SINDBERT, S., KALININ, S., HIEN, N., KIENZLER, A., CLIMA, L., BANNWARTH, W., APPEL, B., MULLER, S. & SEIDEL, C. A. M. 2011a. Accurate Distance Determination of Nucleic Acids via Forster Resonance Energy Transfer: Implications of Dye Linker Length and Rigidity. *Journal of the American Chemical Society*, 133, 2463-2480.
- SINDBERT, S., KALININ, S., NGUYEN, H., KIENZLER, A., CLIMA, L., BANNWARTH, W., APPEL, B., MÜLLER, S. & SEIDEL, C. A. M. 2011b. Accurate distance determination of nucleic acids via Förster resonance energy transfer: implications of dye linker length and rigidity. *J Am Chem Soc*, 133, 2463-80.
- SISAMAKIS, E., VALERI, A., KALININ, S., ROTHWELL, P. J. & SEIDEL, C. A. M. 2010a. Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods in Enzymology*, 475, 456-514.
- SISAMAKIS, E., VALERI, A., KALININ, S., ROTHWELL, P. J. & SEIDEL, C. A. M. 2010b. Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods in Enzymology*, 475, 455-514.
- SYGUDA, A., BAUER, M., BENSCHIED, U., OSTLER, N., NASCHBERGER, E., INCE, S., STURZL, M. & HERRMANN, C. 2012. Tetramerization of human guanylate-binding protein 1 is mediated by coiled-coil formation of the C-terminal  $\alpha$ -helices. *Febs Journal*, 279, 2544-2554.
- VOPEL, T., HENGSTENBERG, C. S., PEULEN, T. O., AJAJ, Y., SEIDEL, C. A. M., HERRMANN, C. & KLARE, J. P. 2014. Triphosphate Induced Dimerization of Human Guanylate Binding Protein 1 Involves Association of the C-Terminal Helices: A Joint Double Electron-Electron Resonance and FRET Study. *Biochemistry*, 53, 4590-4600.
- VOPEL, T., KUNZELMANN, S. & HERRMANN, C. 2009. Nucleotide dependent cysteine reactivity of hGBP1 uncovers a domain movement during GTP hydrolysis. *Febs Letters*, 583, 1923-1927.
- VOPEL, T., SYGUDA, A., BRITZEN-LAURENT, N., KUNZELMANN, S., LUDEMANN, M. B., DOVENGERDS, C., STURZL, M. & HERRMANN, C. 2010. Mechanism of

- GTPase-Activity-Induced Self-Assembly of Human Guanylate Binding Protein 1. *Journal of Molecular Biology*, 400, 63-70.
- WANG, J. M., CIEPLAK, P. & KOLLMAN, P. A. 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, 21, 1049-1074.
- WEHNER, M., KUNZELMANN, S. & HERRMANN, C. 2012. The guanine cap of human guanylate-binding protein 1 is responsible for dimerization and self-activation of GTP hydrolysis. *Febs Journal*, 279, 203-210.
- WELLS, S. A., MENOR, S., HESPENHEIDE, B. & THORPE, M. F. 2005. Constrained geometric simulation of diffusive motion in proteins. *Physical Biology*, 2, S127-S136.
- WOZNIAK, A. K., SCHRODER, G. F., GRUBMULLER, H., SEIDEL, C. A. M. & OESTERHELT, F. 2008. Single-molecule FRET measures bends and kinks in DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 18337-18342.
- YU, E. W. & KOSHLAND, D. E., JR. 2001. Propagating conformational changes over long (and short) distances in proteins. *Proc Natl Acad Sci U S A*, 98, 9517-20.

---

# **Supplement A**

**A toolkit and benchmark study for FRET-restrained high-precision structural modeling**

# A toolkit and benchmark study for FRET-restrained high-precision structural modeling

Stanislav Kalinin<sup>1</sup>, Thomas Peulen<sup>1,5</sup>, Simon Sindbert<sup>1,5</sup>, Paul J Rothwell<sup>1,2</sup>, Sylvia Berger<sup>1</sup>, Tobias Restle<sup>2,3</sup>, Roger S Goody<sup>2</sup>, Holger Gohlke<sup>4</sup> & Claus A M Seidel<sup>1</sup>

**We present a comprehensive toolkit for Förster resonance energy transfer (FRET)-restrained modeling of biomolecules and their complexes for quantitative applications in structural biology. A dramatic improvement in the precision of FRET-derived structures is achieved by explicitly considering spatial distributions of dye positions, which greatly reduces uncertainties due to flexible dye linkers. The precision and confidence levels of the models are calculated by rigorous error estimation. The accuracy of this approach is demonstrated by docking a DNA primer-template to HIV-1 reverse transcriptase. The derived model agrees with the known X-ray structure with an r.m.s. deviation of 0.5 Å. Furthermore, we introduce FRET-guided 'screening' of a large structural ensemble created by molecular dynamics simulations. We used this hybrid approach to determine the formerly unknown configuration of the flexible single-strand template overhang.**

In recent years, single-molecule fluorescence spectroscopy has come of age. Single-molecule FRET (smFRET) serves as a 'spectroscopic ruler'<sup>1</sup> by reporting distance information between donor and acceptor fluorophores placed within a certain proximity. The method is providing important insights into the structural heterogeneity and function of biomolecules under *in vitro*<sup>2</sup> and *in vivo*<sup>3</sup> conditions. Yet a common misconception about FRET is that it is too inaccurate for structural modeling. The uncertain fluorophore positions with respect to their attachment points and the orientation dependence of FRET efficiency (the 'κ<sup>2</sup> problem') are considered fundamental limitations. We argue that explicitly modeling dye behavior<sup>4-7</sup> is the key to increasing the spatial resolution of FRET with flexibly linked dyes, thus making it a truly quantitative tool. This requires calculating the distribution of dye positions while considering the structure of the biomolecule. In contrast, the conversion of FRET data into distances between the labeling sites results in an unnecessary loss of accuracy.

Several approaches to derive FRET-restrained structures have been published<sup>8-17</sup>. In particular, several groups<sup>12,13,15</sup> have successfully developed and applied FRET-guided rigid-body docking with restraints obtained from experiments on immobilized

molecules. A complementary probabilistic data analysis approach termed the 'nano-positioning system'<sup>10,11,18</sup> aims to find fixed fluorophore positions consistent with experimental data. However, many questions remain. Although mean dye position displacement due to flexible linkers is well understood<sup>4,6,13,19</sup>, other issues have received less attention. The effects of averaging FRET efficiency over donor-acceptor distance distributions, the potential impact of the spatial arrangement of structural units on dye distributions, and missing information on the quality of structural models as influenced by the uncertainties of 'input' FRET data are all important considerations. There is also little evidence as to the accuracy of FRET-restrained three-dimensional (3D) modeling, which could be gained through comparison to known structures. (To our knowledge, FRET-derived models were quantitatively tested against independent structural data only once<sup>20</sup>, using low-resolution FRET data.) Finally, a productive combination of FRET and state-of-the-art *in silico* modeling to generate candidate model structures is needed.

We have established a comprehensive toolkit and pipeline for FRET-restrained high-precision structural modeling that considers the mobility and structural heterogeneity of biomolecules. To validate our approach, we performed a benchmark study with the DNA primer template in complex with a human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT), showing that our model is within a 0.5-Å r.m.s. deviation from the crystal structure. FRET-guided screening of a large structural ensemble created by molecular dynamics (MD) simulations allowed us to characterize the previously unresolved flexible single-strand template overhang.

Software for FRET-restrained positioning and screening (FPS) is freely available in the **Supplementary Software** and at <http://www.mpc.hhu.de/software>.

## RESULTS

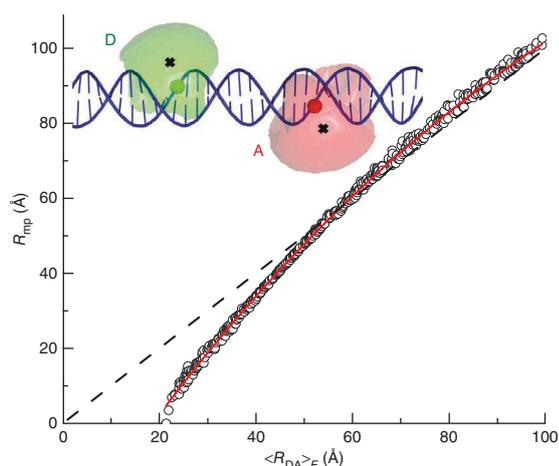
### Using a fast approximation of dye position distributions

There is ample theoretical<sup>5,7,21,22</sup> and experimental<sup>23</sup> evidence for the existence of donor-acceptor distance distributions due to dye linkers (**Supplementary Fig. 1**). **Figure 1** illustrates

<sup>1</sup>Lehrstuhl für Molekulare Physikalische Chemie, Heinrich-Heine-Universität (HHU), Düsseldorf, Germany. <sup>2</sup>Max-Planck-Institut für Molekulare Physiologie, Dortmund, Germany. <sup>3</sup>Institut für Molekulare Medizin, Universität zu Lübeck, Lübeck, Germany. <sup>4</sup>Institut für Pharmazeutische und Medizinische Chemie, HHU, Düsseldorf, Germany. <sup>5</sup>These authors contributed equally to this paper. Correspondence should be addressed to S.K. ([stanislav.kalinin@uni-duesseldorf.de](mailto:stanislav.kalinin@uni-duesseldorf.de)) or C.A.M.S. ([cseidel@hhu.de](mailto:cseidel@hhu.de)).

RECEIVED 22 APRIL; ACCEPTED 5 OCTOBER; PUBLISHED ONLINE 11 NOVEMBER 2012 DOI:10.1038/NMETH.2222

## Supplement A



**Figure 1** | Effect of dye position distributions on FRET efficiency. The distance between mean dye positions ( $R_{mp}$ ) is plotted against the distance formally calculated from the FRET efficiency,  $\langle R_{DA} \rangle_E$ . An  $\langle R_{DA} \rangle_E$ -to- $R_{mp}$  conversion function (red line) was generated by fitting a third-order polynomial to  $\langle R_{DA} \rangle_E/R_{mp}$  value pairs (open circles) calculated for a set of randomly oriented accessible volumes (AVs) of Alexa 488 and Cy5 ( $R_0 = 52$  Å) for dsDNA. The r.m.s. deviation between the data and the polynomial approximation is 0.9 Å over the whole range of  $\langle R_{DA} \rangle_E$  and 0.6 Å for  $30$  Å  $< \langle R_{DA} \rangle_E < 70$  Å. The black dashed line serves as the reference for direct correlation with a slope of 1. The difference between  $R_{mp}$  and  $\langle R_{DA} \rangle_E$  can reach 10 Å in the range accessible to FRET (~30–70 Å). The DNA schematic shows the space accessible to Alexa 488 (green) and Cy5 (red) attached to a dsDNA via a C6 linker, calculated using the AV algorithm. The mean positions of the dyes are depicted as black crosses; the attachment atoms of the linkers, as spheres. The distance between the dyes' attachment points (the C5 atoms) is 43.6 Å, whereas the distance between the dyes' mean positions ( $R_{mp}$ ) is 52.6 Å. D, donor and A, acceptor dyes.

the space accessible to the donor (D, Alexa 488) and acceptor (A, Cy5) fluorophores attached to DNA using standard C6 linkers. The mean dye positions are clearly far from attachment points and the accessible space is large, requiring that FRET observables be averaged over a distribution of donor-acceptor distances. The distance between mean dye positions,  $R_{mp}$ , and that calculated from the FRET efficiency,  $\langle R_{DA} \rangle_E$  (Online Methods), exhibit a large difference of up to 10 Å, or ~30% (Fig. 1). This confirms that considering distance distributions is essential for quantitative FRET, particularly when  $R_{mp}$  is below the Förster radius  $R_0$ .

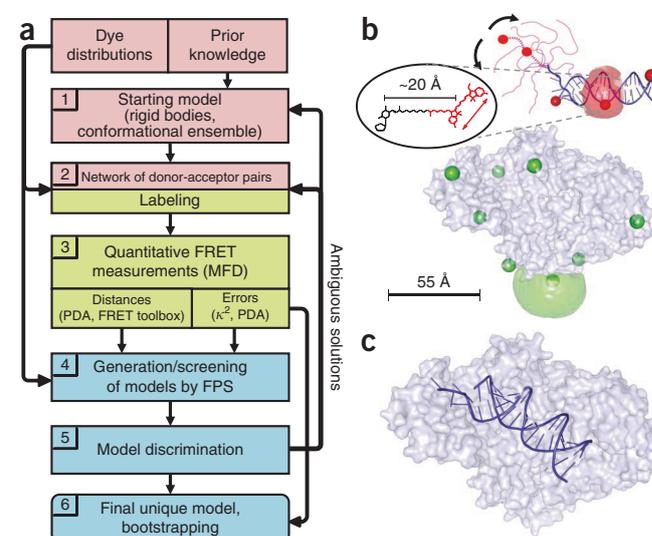
However, the expected FRET efficiency can only be calculated with high precision if the spatial distributions of donor and acceptor are accurately predicted. Dye behavior can be realistically modeled using MD simulations<sup>5,7,13</sup>, but these are time consuming. As a fast alternative with no sampling problems, in many cases a geometric accessible volume (AV) algorithm<sup>11,24</sup> can give a reasonable approximation<sup>23</sup> in under 1 s of computational time, making the screening of large structural ensembles feasible. The assumptions of the AV approach can be verified by analyzing complementary fluorescence parameters from a multiparameter fluorescence detection (MFD) experiment, such as fluorescence lifetime distribution and anisotropy, to rule out immobile dyes. Moreover, the FRET pair can be calibrated using molecules of known structure, such as dsDNA. We argue that systematic errors related to the AV approach are likely to average out if many donor-acceptor distances are measured, and we propose using the AV algorithm (Supplementary Note 1) for FRET-restrained positioning and

screening. That way, ångström precision can be achieved when positioning labeled macromolecules, even if a single FRET measurement does not provide the distance between two labeling sites with high accuracy.

### Workflow for FRET-restrained positioning and screening

Six steps are needed to generate a FRET-restrained structural model (Fig. 2a). (i) Generate a starting model using known structures, homology modeling or *ab initio* modeling. (ii) Design a network of dye positions from the starting model, taking the distributions of the coupled dyes into account. (iii) Measure FRET by single-molecule MFD (smMFD), thereby avoiding ensemble averaging; analyze photon distributions and time-resolved anisotropies of the dyes for rigorous data analysis and error estimation of donor-acceptor distances. (iv) Search for and evaluate possible structural models on the basis of their agreement with the FRET data by FPS. (We used two complementary approaches: discovering possible arrangements by rigid-body docking of known substructures and screening models in a large structure library.) (v) Rank models according to their violation of FRET restraints and cluster by similarity to judge their uniqueness and determine the corresponding confidence levels. (vi) Determine model precision by bootstrapping.

**Figure 2** | FRET-restrained positioning and screening (FPS). (a) FRET-restrained high-precision structural modeling comprises three main parts: experimental design (pink shading), sample measurement and analysis (green), and generation and validation of structural models (blue). (b) Separate protein (gray) and dsDNA (blue) structures from the 1ROA crystal structure, and cartoon of the unresolved ssDNA overhang (magenta). Naming and details of the eight donor labeling positions (small green spheres) and five acceptor positions (small red spheres) are in the Online Methods, Supplementary Figures 1 and 2 and Supplementary Table 1. The acceptor dyes (Cy5) are attached via C6 linkers (shown in ellipse). AV clouds are generated as in Supplementary Note 1 and are shown for labeling positions p51E194C (large green volume) and dp(10) (large red volume). Parameters used for generation of AV clouds are given in Supplementary Note 1. (c) Structural model of the RT:dp/dt complex obtained by rigid-body docking.



We applied this workflow to HIV-1 RT in complex with a 19/35 DNA primer/DNA template (dp/dt) (**Fig. 2b,c**). RT is a heterodimer composed of a 66-kD chain (p66) and a 51-kD chain (p51) that transcribes viral RNA into dsDNA<sup>25</sup>. We characterized both the rigid double-stranded and flexible single-stranded DNA parts of the complex (**Fig. 2b**). Several crystal structures exist for the productive complex in the open educt state (P-E)<sup>26,27</sup>, the state immediately before incorporation of the next nucleotide<sup>28</sup>, thus allowing us to determine the accuracy of FPS for dsDNA (**Fig. 2c**). The ssDNA template overhang was previously unresolved by X-ray crystallography<sup>25,29,30</sup>. However, by combining FPS and MD simulations, we found a well-defined configurational space of the overhang preferentially interacting with one region of the protein.

### Detailed workflow using HIV-1 RT case study

Here we demonstrate FRET-restrained high-precision structural modeling by applying our workflow to model dsDNA and ssDNA in the RT:dp/dt complex.

**Step 1: starting models.** The complex partners and the labeling positions are illustrated in **Figure 2b**. As prior knowledge, we chose the crystal structure with a 2.8-Å resolution from ref. 26 (PDB ID: 1R0A; justification given in **Supplementary Note 2**), where the RT:dp/pt complex is in the open educt state (P-E)<sup>28</sup>.

To test the accuracy of FRET-restrained modeling, we separated the dp/dt from the protein and applied FPS. Alternatively, we docked the protein from 1R0A with a canonical B-DNA and relaxed the resulting complex structure by MD simulations. To determine the template overhang conformation missing in the crystal structure in the final step, we generated a starting model for the MD simulations by attaching the single strand to the crystallized DNA such that it projects straight out from the protein.

**Step 2: network of donor-acceptor pairs.** We chose eight donor and five acceptor label positions on the enzyme and the primer/template DNA (**Supplementary Table 1**), respectively (**Fig. 2b**). Overall, 36 independent smFRET measurements were performed for the RT:dp/dt complex.

**Step 3.1: quantitative smFRET measurements by smMFD.** smMFD experiments avoid ensemble averaging by analyzing single-molecule events. Distance information is usually deduced from the FRET efficiency ( $E$ )<sup>31,32</sup>, which can be calculated either from the donor and acceptor fluorescence,  $F_D$  and  $F_A$ , or from donor fluorescence lifetimes in the presence ( $\tau_{D(A)}$ ) and absence ( $\tau_{D(0)}$ ) of the acceptor<sup>33</sup>.

$$E = \frac{F_A}{(\gamma' F_D + F_A)} = 1 - \frac{\tau_{D(A)}}{\tau_{D(0)}}, \quad \text{with } \gamma' = \Phi_{F_A} / \Phi_{F_D(0)} \quad (1)$$

In equation (1), the correction factor  $\gamma'$  accounts for fluorescence quantum yields  $\Phi_F$  of the donor and acceptor. In MFD, all fluorescence parameters are acquired simultaneously<sup>33</sup>, which enables a multidimensional analysis. The correlated FRET analysis by equation (1) helps to avoid most pitfalls of ensemble FRET measurements, such as incomplete labeling, fluorophore quenching and the inability to resolve multiple FRET states<sup>9,28,33</sup>.

In **Figure 3a**, 2D burst-frequency histograms of the  $F_D/F_A$  signal ratio and donor anisotropy ( $r_D$ ) versus  $\tau_{D(A)}$  are presented

for the complex RT(p51K173C):dp(10)/dt. In agreement with ref. 28, three complex types were found: dead-end (D-E), productive complex in the product state (P-P) and productive complex in the educt state (P-E; see **Supplementary Note 3**). The observed populations followed the theoretically expected dependencies between  $\tau_{D(A)}$  and  $F_D/F_A$  (**Supplementary Note 4**) or  $r_D$ , respectively. This indicates that no substantial dye quenching took place, which could have resulted in errors in the recovered  $R_{DA}$ , and that no long-lived immobile dye population was present (as characterized by a long mean rotation correlation time  $\rho$ ). Our 2D FRET analysis of all other data sets can be found in **Supplementary Data 1**.

Here we calculate  $E$  from intensities (equation (1)) because  $F_D$  and  $F_A$  obey well-defined statistics<sup>34,35</sup>, whereas the lifetime information ensures that the observed effects are due to FRET. Using both 2D analysis and photon distribution analysis (PDA; see step 3 below; analysis in the 1D  $F_D/F_A$  histogram in **Fig. 3a**)<sup>34,36</sup> gives unsurpassed sensitivity for characterization of FRET populations derived from smFRET experiments. For the template overhang, we used distances obtained from sub-ensemble time-correlated single-photon counting (seTCSPC) data (**Supplementary Note 5** and **Supplementary Data 2**).

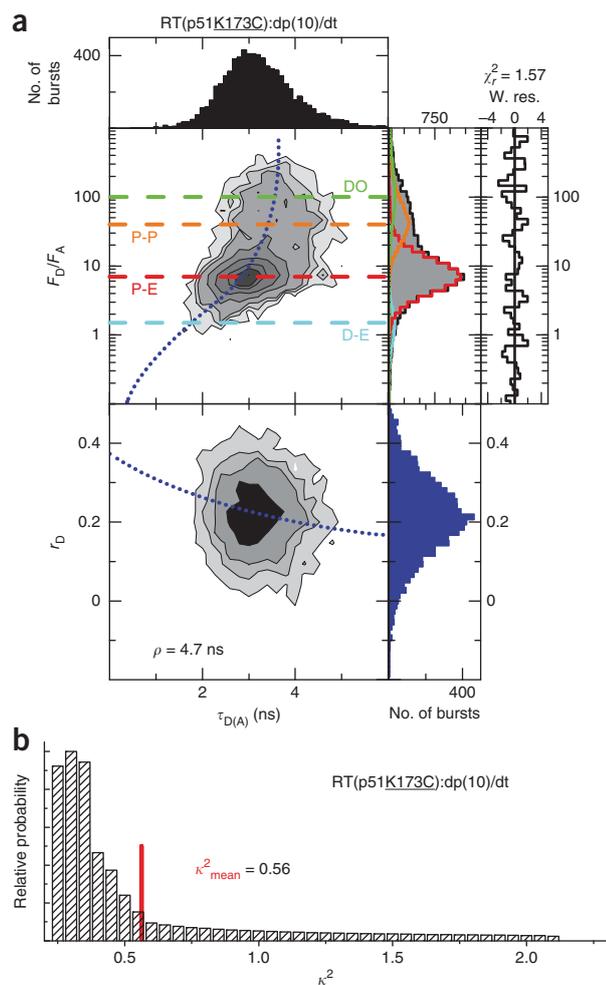
**Step 3.2: input data for FPS—distances and uncertainties.** PDA simultaneously provides mean distances  $\langle R_{DA} \rangle_E$  and uncertainties ( $\Delta R_{DA}$ ) by explicitly taking photon statistics into account. PDA needs fewer free parameters than the traditional approach of fitting multiple Gaussian peaks. It provides meaningful fit-quality parameters that allow one to justify the chosen fit model. To estimate the errors of fitted parameters,  $\Delta R_{DA}(E)$ , we explored the parameter space for sets of variables providing acceptable fits (**Supplementary Note 4**). If  $E$  distributions are fitted by Gaussians using general-purpose fitting software, error estimation is usually difficult because the proper model function and the s.d. of data points are unavailable.

Uncertainties in the mutual orientation of donor and acceptor ( $\kappa^2$  errors,  $\Delta R_{DA}(\kappa^2)$ ) can be estimated by analyzing anisotropy decays<sup>37</sup> accessible in MFD (**Supplementary Note 6** and **Supplementary Data 3**). A typical distribution of possible  $\kappa^2$  values compatible with experimental anisotropy decays is shown in **Figure 3b**. Because of the weak dependence of  $R_0$  on  $\kappa^2$  (ref. 32), this broad distribution results in only 10% uncertainty in the distance. As a consequence, correcting  $\langle R_{DA} \rangle_E$  values for the mean  $\kappa^2$  for all allowed orientations of D and A (**Fig. 3b**) yields almost the same final structures as the assumption of  $\langle \kappa^2 \rangle = 2/3$  (**Supplementary Note 6**). The overall uncertainties are determined following well-known error-propagation rules (Online Methods). This procedure was applied to all distances measured for the dsDNA and for the template overhang (**Supplementary Tables 2** and **3**).

For generating or judging FRET-based structure models, the knowledge of  $\langle E \rangle$  and, thus,  $\langle R_{DA} \rangle_E$  is sufficient. These parameters can be calculated for any putative structure using AV or MD models of dye position distributions and are directly comparable with experimental data. In practice, it is useful to convert  $\langle R_{DA} \rangle_E$  into  $R_{mp}$  (**Fig. 1**) to avoid repeated calculations or transformations of AVs during iterative structure optimization (**Supplementary Note 7**).

To calculate  $\langle R_{DA} \rangle_E$  from the AV model, we assumed static averaging of distances and dynamic reorientation (not necessarily

## Supplement A



resulting in  $\langle \kappa^2 \rangle = 2/3$ ) on the timescale of FRET. We have previously shown<sup>6,23</sup> that this approximation works well for dyes attached to DNA and RNA via long methylene linkers. Although the reorientation timescale is not as fast for donor dyes attached to RT (Fig. 3a), fast rotations of acceptor bound to DNA justify the assumption of dynamic  $\kappa^2$  averaging.

**Step 4a: rigid-body docking of dsDNA with FRET restraints.** To find the position and orientation of dsDNA (from crystal structure or B-DNA) with respect to RT in best agreement with FRET data, the weighted data-model deviation ( $\chi_E^2$ ) has to be minimized for the set of  $n$  distances

$$\chi_E^2 = \sum_{i=1}^n \frac{(R_{DA(i)} - R_{\text{model}(i)})^2}{\Delta R_{DA(i)}^2} \quad (2)$$

This optimization problem can be defined for  $\langle R_{DA} \rangle_E$  values or, more conveniently, for converted mean position distances  $R_{\text{mp}}$ , which are computed using a polynomial description (Fig. 1 and Supplementary Note 7). To solve this problem, we assumed partners in the complex to be rigid bodies and applied a rigid-body dynamics approach to dock them using FRET restraints. Although a large number of more sophisticated approaches exist (for example, refs. 38,39), to our knowledge none allow for the explicit modeling of fluorophores and averaging of measured quantities over distributions of donor-acceptor distances.

**Figure 3 | Distance measurements and error estimation.** (a) 2D burst-frequency histograms of  $F_D/F_A$  versus donor fluorescence lifetime  $\tau_{D(A)}$  (center left subpanel) and donor fluorescence anisotropy  $r_D$  versus  $\tau_{D(A)}$  (bottom left subpanel) for RT(p51K173C):dp(10)/dt. The number of molecules (bursts) is scaled from white (lowest) to black (highest). 1D histograms are shown as projections. The dotted blue line (center) represents the theoretical relationship between  $F_D/F_A$  and  $\tau_{D(A)}$  (static FRET line; Supplementary Note 4). The dotted blue line (bottom) represents the expected relationship between  $r_D$  and  $\tau_{D(A)}$  and is given by the Perrin equation  $r_D = r_0 / (1 + \tau_{D(A)} / \rho)$ , with mean rotational correlation time  $\rho = 4.7$  ns,  $r_0 = 0.374$ . The 1D  $F_D/F_A$  histogram is fitted (black line) by PDA using three FRET states with Gaussian distributed distances ( $\sigma_{\text{app}} = 0.09 \times \langle R_{DA} \rangle_E$ ; Supplementary Note 4), 55% P-E  $\langle R_{DA} \rangle_E$  (P-E) = 62 Å, 23% P-P  $\langle R_{DA} \rangle_E$  (P-P) = 88 Å and 4% D-E  $\langle R_{DA} \rangle_E$  (D-E) = 47 Å, and with 18% donor only (DO). The fit yields  $\chi_r^2 = 1.57$  (weighted residuals (w. res.) are plotted to the right of the 1D  $F_D/F_A$  histogram). FRET states: P-E, protein in educt state (red); P-P, protein in product state (orange); D-E, dead-end complex (cyan) (Supplementary Note 3). Dashed horizontal lines in the  $F_D/F_A$  versus  $\tau_{D(A)}$  plot correspond to burst distributions of individual FRET states at right. (b) Distribution of possible  $\kappa^2$  values for RT(p51K173C):dp(10)/dt derived from time-resolved anisotropy analyses. The uncertainty in  $\kappa^2$  results in an error of 10.0% in  $\langle R_{DA} \rangle_E$  (Supplementary Note 6).

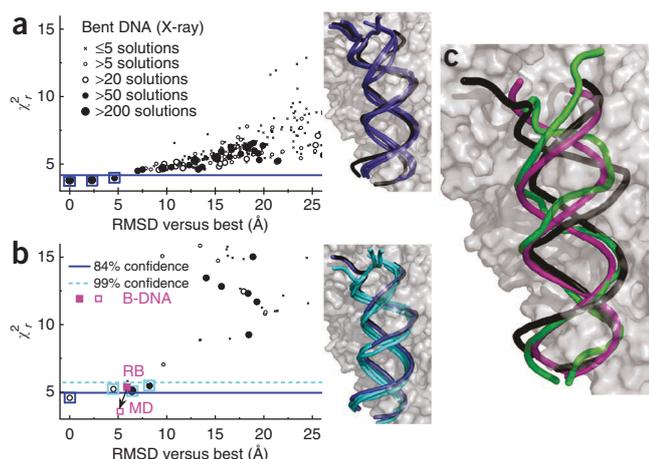
Here we estimate the coordinates of mean dye positions by AV simulations and then fix them with respect to the labeled substructure. If a distance is measured between certain donor and acceptor dyes, this is accounted for by adding a ‘spring’ connecting the dyes’ mean positions that has an equilibrium length of  $R_{\text{mp}}$  and a strength derived from the corresponding  $\Delta R_{DA}$  (see Supplementary Note 8 for implementation). Relaxing this system is equivalent to minimizing the  $\chi_E^2$  parameter given by equation (2). Two steps are distinguished in the rigid-body dynamics approach.

‘Search’: in the first step, we generate a large number of complexes starting from random configurations of the binding partners, excluding those with steric clashes. To prevent clashes between RT and DNA, we introduce strong repulsive forces between atoms approaching each other by a distance smaller than the sum of their van der Waals radii. In this way, the positioning is guided by an overall quality parameter (reduced  $\chi_r^2$ ; see Online Methods) that accounts for violations of FRET restraints and of van der Waals radii.

‘Refinement’: in the second step, AVs are recalculated, accounting for possible interactions (steric clashes) between the dyes and parts of the biomolecule structure that the dyes are not attached to. The resulting mean dye positions are used to reoptimize the structure.

**Step 4b: screening of structural ensembles.** An alternative strategy for finding a structure compatible with FRET data is to generate a large ensemble of putative structures (for example, by extracting conformations from an MD trajectory) and to filter the results with respect to agreement with FRET data by calculating  $\chi_r^2$  for each structure. Structures with a low  $\chi_r^2$  and good configurational quality in terms of MD are then selected. An obvious advantage of this approach is that state-of-the-art algorithms for conformational sampling can be used instead of rigid-body docking. Below we screened in two ways: (i) by refining a docked structure (B-DNA) that has been relaxed by MD simulations or (ii) by searching for a template overhang structure within a large ensemble, in the absence of prior knowledge.

**Step 5: analysis of docking/screening results.** We clustered structures obtained by docking into groups of solutions with similar



**Figure 4** | Cluster analysis of docking results. **(a,b)** Rigid-body docking of RT and bent dp/dt (both from crystal structure 1ROA). Left, cluster analysis of obtained solutions for initial search phase with 6-Å clash tolerance **(a)** and second refinement iteration with 1-Å clash tolerance **(b)**. Horizontal lines represent confidence thresholds as defined by  $\chi_r^2 < \chi_{r,\min}^2 + 0.378$  (blue solid line,  $\sim 1.4 \sigma$ ) and  $\chi_r^2 < \chi_{r,\min}^2 + 1.09$  (cyan dashed line,  $\sim 2.6 \sigma$ ). The thresholds are derived from the  $\chi^2$  distribution with 14 d.f. ( $\chi^2 < 19.3$  and  $\chi^2 < 29.2$ , respectively). Right, DNA from X-ray structure (black; obscured in **b** owing to agreement with docked DNA) overlaid with FPS solutions (blue, cyan) indicated in cluster plots by colored squares. The cluster plot in **b** also shows the solutions for rigid-body docking B-DNA (RB, filled magenta square) and refinement of B-DNA via MD simulations (MD, open magenta square). The improvement with respect to agreement with FRET data through MD refinement is indicated by the black arrow. **(c)** Overlay of DNA from crystal structure (1ROA, black) with docked B-DNA (magenta) and MD relaxed B-DNA (green).

$\chi_r^2$  values and low r.m.s. deviation within the group (**Fig. 4** and **Supplementary Note 9**). After a coarse search step (**Fig. 4a**; 6-Å clash tolerance), three groups of solutions were found below an 84% confidence threshold with respect to the best possible fit with  $\chi_{r,\min}^2$  (**Fig. 4a,b** and Online Methods). In relation to the X-ray structure, the r.m.s. deviation values for these groups were 6.3 Å, 5.0 Å and 9.7 Å, respectively. After one refinement iteration consisting of recalculating AVs and reducing the clash tolerance to 2 Å, only one solution remained below the threshold (1.3-Å r.m.s. deviation; data not shown). A second refinement with 1-Å clash tolerance further separated the best structure from the other solutions, with an r.m.s. deviation value of only 0.5 Å with respect to the X-ray structure (**Fig. 4b**). The agreement was remarkably good considering the dye linker lengths of  $\sim 20$  Å. Even if we required a significance level of  $>99\%$  (**Fig. 4b**) and accept the solutions with r.m.s. deviations of  $\sim 4$ –8 Å, the deviation from the X-ray structure was still much smaller than the sum of dye linker lengths, which justifies the term ‘high-precision FRET’. Compared to typical  $\kappa^2$  uncertainties ( $\sim 9\%$ , **Supplementary Tables 2** and **3**), the resulting errors were also small, most likely because of averaging toward  $\langle \kappa^2 \rangle = 2/3$  when multiple donor-acceptor distances were considered; the same applies to possible systematic errors of the AV approach. Accounting for clashes had a clear effect on the selectivity, but even with unrealistically mild restrictions (as in the search step, **Fig. 4a**) FRET models already agreed well with the known structure (**Supplementary Note 10**).

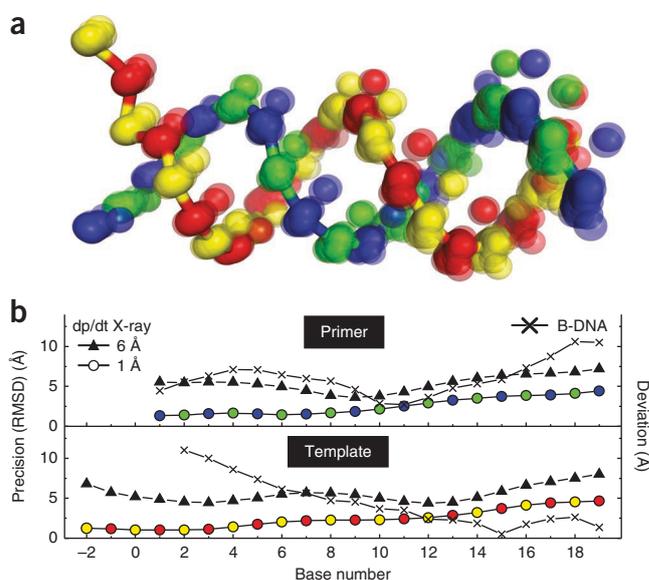
As an independent validation, we applied rigid-body docking to straight B-DNA and the protein structure from 1ROA (**Fig. 4c** and **Supplementary Note 11**). Clearly, rigid-body docking could not account for bending of DNA; however, the resulting model

agreed reasonably well with the X-ray structure (r.m.s. deviation = 5.9 Å). At the same time, because of its high  $\chi_r^2$  value of 5.36, this model could be rejected with  $\sim 95\%$  confidence (**Fig. 4b**). This confirmed the high resolution power of FPS: FRET ‘senses’ that the shape of straight B-DNA is not optimal.

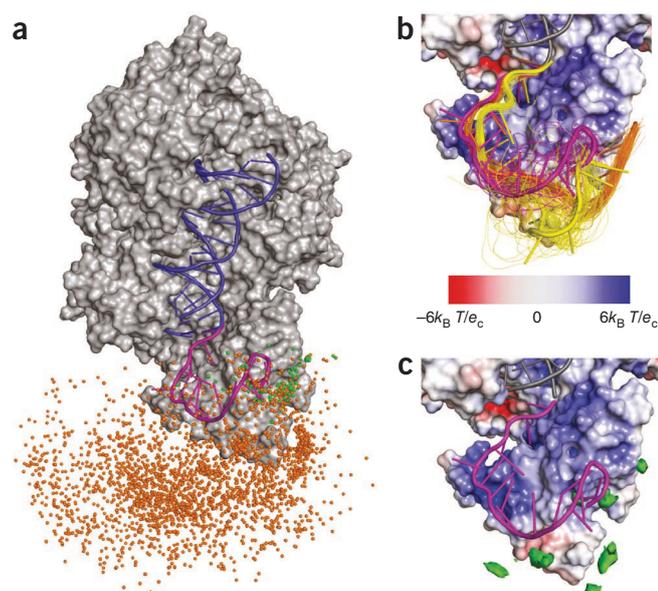
To find out whether the DNA becomes bent on binding, the docked RT:B-DNA complex was used as a starting model for relaxation by MD (**Supplementary Note 12**). Ten MD trajectories were screened for agreement with 20 FRET distances measured for the crystallized part of the complex. Screening of MD data revealed a clear correlation between  $\chi_r^2$  for FRET data and the MD agreement with the X-ray structure (**Supplementary Note 11**). In **Figure 4c**, a representative of the 50 best MD structures is shown. Relaxation by MD simultaneously improved the agreement with the X-ray structure (r.m.s. deviation = 5.2 Å, not accounting for the three bases of the single-stranded template overhang) and with FRET data (**Fig. 4b**), demonstrating the usefulness of combining MD simulations with high-precision FRET.

*Step 6: estimation of precision and accuracy.* The uncertainties of FRET-restrained modeling include possible alternative solutions (local  $\chi_r^2$  minima) with data-model deviations and uncertainties

**Figure 5** | Bootstrapping of docked dsDNA dp/dt. **(a)** Ensemble of structures generated with perturbed distances and a 1-Å clash tolerance. For better comparison, the phosphorus (P) atoms of the DNA backbone are alternately colored yellow/red or blue/green for the primer and template strands, respectively. The DNA is oriented as in **b**. **(b)** Uncertainty of P atom positions calculated for each P atom using equation (4) after the initial search with a 6-Å clash tolerance (triangles) and after two refinement steps with 2-Å (not shown) and 1-Å (circles) clash tolerances, respectively. The average r.m.s. deviation values for all P are 5.9 Å and 2.9 Å, respectively. The deviations between the docked B-DNA and the DNA of the crystal structure are also shown (crosses).



## Supplement A



**Figure 6** | Structure of ssDNA dp/dt overhang obtained by screening of MD trajectories. **(a)** Ensemble of ssDNA overhang structures generated by all-atom MD simulations in explicit solvent (orange; 2,855 conformations) and conformations filtered by sub-ensemble time-correlated single-photon counting (seTCSPC) using a confidence interval of  $1\sigma$  (green; 150 conformations). Dots represent the N1 atom of the nucleobase of the nucleotide dt(-15). The structure that best fits the FRET data is shown as a magenta cartoon ( $\chi_r^2 = \chi_E^2/16 = 0.88$ ). **(b)** Conformations within a confidence interval of  $1\sigma$  of the seTCSPC data. The PDB file with all conformers is given as **Supplementary Data 4**. The electrostatic potential of the protein as determined by adaptive Poisson-Boltzmann solver (APBS) has been mapped onto the protein surface using a color scale. **(c)** Green isopleths show regions of preferred residence of the N1 atom of the nucleobase of nucleotide dt(-15) as determined from MD simulations. The isopleths encompass regions with at least 2% of the maximal residence likelihood. Note the overall good agreement between these regions and the location of the  $1\text{-}\sigma$  confidence interval determined by seTCSPC **(b)**.

similar to those of the best solution itself. Assuming that a unique solution had been found at this point, we applied a procedure similar to bootstrapping<sup>40</sup> to estimate the precision for the best structure.

The model distance set calculated for the optimized structure (**Fig. 4b**) was perturbed by simultaneously adding normally distributed random numbers with a mean of 0 and an s.d. given by the experimental errors  $\{\Delta R_{DA}\}$  (step 3.2) to all model distances. The structure was then reoptimized with the perturbed distance restraints. The procedure was repeated 100 times, yielding 100 perturbed structures for the double-stranded dp/dt part (**Fig. 5a**). This set of structures represents a distribution of possible positions of all backbone atoms consistent with experimental data. For each phosphorus (P) atom, the uncertainty was calculated (**Fig. 5b**) from the resulting structural ensemble (Online Methods, equation (4)). For the best FRET model, we estimated an average precision of 2.9-Å r.m.s. deviation for all P atoms of the dsDNA. The X-ray structure (0.5-Å r.m.s. deviation from the FRET structure, all dsDNA atoms) was well within these uncertainty limits. On the other hand, the model with docked B-DNA showed large deviations from the best solution and could be excluded (**Fig. 5b**). The solutions obtained at earlier stages of rigid-body docking were clearly less precise and less accurate, showing that the refinement steps are essential.

For MD-relaxed B-DNA, the uncertainty is given by the largest r.m.s. deviation value for any of the 50 structures below the  $\chi_r^2$  threshold with respect to the best solution (**Supplementary Note 11**). This estimation yielded 4.1-Å r.m.s. deviation for P atoms. Thus, given the differences between the structures (**Fig. 4b**) and the precision of individual models, the FRET (2.9 Å for rigid-body docked dsDNA, **Fig. 5**; 4.1 Å for MD-relaxed B-DNA) and the X-ray (2.8 Å; **Supplementary Note 2**) structures could not be distinguished within the limits of precision. However, it is clear that the B-DNA was bent by binding to a significant extent (**Fig. 4b**).

Alternatively, cross-validation or similar tests<sup>41</sup> can be performed. However, in contrast with X-ray or NMR data, the redundancy of FRET data is usually low. New labeling positions are often chosen to distinguish between ambiguous solutions (**Fig. 2a**), and discarding a few FRET restraints might make the position of a unit completely undefined. For this reason, bootstrapping is preferred over procedures in which some data points are completely discarded. For RT, both error estimation methods work well (**Supplementary Note 13**).

### Extension to flexible parts of the complex

We generated a conformational ensemble of the template overhang (**Fig. 2b**), applying all-atom MD simulations in explicit solvent<sup>42</sup>. Ten trajectories (**Fig. 6a**) were filtered using 16 distances determined for the acceptor positions dp/dt(-6) and dp/dt(-15) on the template overhang (**Supplementary Table 4**). In **Figure 6a** the structure of the overhang that best fits the FRET data is depicted together with the approximate  $1\text{-}\sigma$  confidence interval (150 conformations) representing the N1 atom of the nucleobase of nucleotide dt(15). The conformational ensemble satisfying FRET data is represented by three major configurations in **Figure 6b** and **Supplementary Data 4** preferentially located in positively charged regions. The isopleths for the N1 atom determined by MD simulations (**Fig. 6c**) illustrate good agreement between these regions and the structures satisfying FRET restraints. The structures wind around the fingers domain, with the 5' end binding to RT in a loop structure in positively charged protein regions. Although we cannot exclude the existence of other minor conformer populations for template overhang, seTCSPC data showed no excessive broadening of donor-acceptor distance distributions measured for the template overhang as compared to the dsDNA part (**Supplementary Note 14** and **Supplementary Tables 4** and 5). Thus, a substantial population of free unbound overhang can be excluded.

Whereas interactions of the single-strand template region in close proximity to the primer terminus directly affect active site geometry and, thus, fidelity of the enzyme, the effects of template-RT interactions beyond position -6 are not immediately obvious. However, during reverse transcription, the enzyme has to pass regions with extended secondary and even short double-stranded structures. It is conceivable that extensive interactions between RT and the template relatively far upstream of the site of nucleotide incorporation help to resolve such obstacles.

### DISCUSSION

FPS improves the precision of FRET-derived structures dramatically. In combination with advanced computer simulations, it allows for a detailed molecular description of the proposed

structure models. For this hybrid approach, we established a procedure to characterize the uniqueness and precision of FRET-restrained models that is based on a precise spectroscopic estimation of 'input' uncertainties. With the single-molecule advantage of FRET, heterogeneous systems can be investigated easily. Our approach takes advantage of MFD data of freely diffusing molecules. This has two distinct advantages: (i) data are free of problems due to protein immobilization and surface artifacts, and (ii) statistics are better and single-molecule events are perfectly averaged.

Moreover, the FPS toolkit includes an interface for predicting donor-acceptor distances for a given structural model and labeling positions, which is indispensable for planning FRET experiments and comparing them with the experimental results.

Currently, we derive our knowledge of biomolecular structure largely from traditional methods such as X-ray crystallography that determine highly resolved but static models. However, biomolecules are dynamic and undergo intrinsic motions<sup>43,44</sup>.

With MFD, FRET has a key advantage: it allows for the observation of multiple biomolecular conformations in solution<sup>9</sup> with high time resolution determined by the fluorescence lifetime of the dyes on the order of a few nanoseconds. Moreover, together with filtered fluorescence correlation spectroscopy<sup>45</sup>, FRET harbors the potential to study conformational control of biomolecular function in complex systems and associate it with detailed dynamic structures without averaging.

Our application of FPS to the flexible single-strand template overhang of RT:dp/dt revealed a preferential structure with the 5' end of the overhang bound to the fingers domain of RT, corroborating the finding that overall dp/dt binding affinity to RT increases with the overhang length by a factor of 7 (ref. 29). There is biochemical evidence that the properly bound template overhang plays an important role in translocation of nucleic acids during processive DNA synthesis<sup>25</sup>, for example, by helping to resolve secondary structures within the substrate and by proper alignment of the primer terminus within the active site, thus affecting fidelity of DNA synthesis.

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

*Note: Supplementary information is available in the [online version of the paper](#).*

## ACKNOWLEDGMENTS

We dedicate this paper to the memory of our brilliant colleague, Robert M. Clegg, a pioneer in the application of FRET in the life sciences and a remarkable human being. We would like to thank E. Schweinberger, O. Kensch and B.M. Wöhl for assistance with the experiments, E. Hausteiner for analytical software, H. Sanabria for help with data visualization and A. Scheidig for helpful discussions. We acknowledge financial support from the Volkswagen Foundation (to R.S.G. and C.A.M.S.; grant no. I/74470), Max-Planck Society (to R.S.G.), German Federal Ministry of Education and Research (BMBF) (BioFuture grant no. 0311865) and German Science Foundation within SPP 1258 (grant no. SE 1195/12-2) (to C.A.M.S.) and 'Fit for Excellence' initiative of HHU (to H.G.). T.P. thanks the International Helmholtz Research School of Biophysics and Soft Matter (IHRS BioSoft), and S.S. thanks the NRW Research School of Biological Structures in Molecular Medicine and Biotechnology (BioStruct) for scholarships.

## AUTHOR CONTRIBUTIONS

S.K. and C.A.M.S. designed studies and assisted in data analysis and interpretation; P.J.R. and S.B. prepared samples and performed experiments; S.K. developed the software; T.P. and S.S. analyzed the data; H.G. performed MD simulations and

assisted with data interpretation; T.R. and R.S.G. supervised the biochemical experiments and assisted with data collection and interpretation; S.K., T.P., S.S., T.R., H.G. and C.A.M.S. wrote the manuscript; C.A.M.S. supervised the project.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/doi/10.1038/nmeth.2222>. Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Stryer, L. Fluorescence energy transfer as a spectroscopic ruler. *Annu. Rev. Biochem.* **47**, 819–846 (1978).
2. Ha, T. *et al.* Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci. USA* **93**, 6264–6268 (1996).
3. Sakon, J.J. & Wenginger, K.R. Detecting the conformation of individual proteins in live cells. *Nat. Methods* **7**, 203–205 (2010).
4. Merchant, K.A., Best, R.B., Louis, J.M., Gopich, I.V. & Eaton, W.A. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc. Natl. Acad. Sci. USA* **104**, 1528–1533 (2007).
5. Best, R.B. *et al.* Effect of flexibility and *cis* residues in single-molecule FRET studies of polyproline. *Proc. Natl. Acad. Sci. USA* **104**, 18964–18969 (2007).
6. Woźniak, A.K., Schröder, G., Grubmüller, H., Seidel, C.A.M. & Oesterhelt, F. Single-molecule FRET measures bends and kinks in DNA. *Proc. Natl. Acad. Sci. USA* **105**, 18337–18342 (2008).
7. Hoefling, M. *et al.* Structural heterogeneity and quantitative FRET efficiency distributions of polyprolines through a hybrid atomistic simulation and Monte Carlo approach. *PLoS ONE* **6**, e19791 (2011).
8. Mekler, V. *et al.* Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **108**, 599–614 (2002).
9. Margittai, M. *et al.* Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. *Proc. Natl. Acad. Sci. USA* **100**, 15516–15521 (2003).
10. Andrecka, J. *et al.* Single-molecule tracking of mRNA exiting from RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **105**, 135–140 (2008).
11. Muschielok, A. *et al.* A nano-positioning system for macromolecular structural analysis. *Nat. Methods* **5**, 965–971 (2008).
12. Choi, U.B. *et al.* Single-molecule FRET-derived model of the synaptotagmin 1-SNARE fusion complex. *Nat. Struct. Mol. Biol.* **17**, 318–324 (2010).
13. Brunger, A.T., Strop, P., Vrljic, M., Chu, S. & Wenginger, K.R. Three-dimensional molecular modeling with single molecule FRET. *J. Struct. Biol.* **173**, 497–505 (2011).
14. Sabir, T., Schroder, G.F., Toulmin, A., McGlynn, P. & Magennis, S.W. Global structure of forked DNA in solution revealed by high-resolution single-molecule FRET. *J. Am. Chem. Soc.* **133**, 1188–1191 (2011).
15. McCann, J.J., Zheng, L.Q., Chiantia, S. & Bowen, M.E. Domain orientation in the N-terminal PDZ tandem from PSD-95 is maintained in the full-length protein. *Structure* **19**, 810–820 (2011).
16. Balci, H., Arslan, S., Myong, S., Lohman, T.M. & Ha, T. Single-molecule nanopositioning: structural transitions of a helicase-DNA complex during ATP hydrolysis. *Biophys. J.* **101**, 976–984 (2011).
17. Boura, E. *et al.* Solution structure of the ESCRT-I complex by small-angle X-ray scattering, EPR, and FRET spectroscopy. *Proc. Natl. Acad. Sci. USA* **108**, 9437–9442 (2011).
18. Muschielok, A. & Michaelis, J. Application of the nano-positioning system to the analysis of fluorescence resonance energy transfer networks. *J. Phys. Chem. B* **115**, 11927–11937 (2011).
19. Olofsson, M., Kalinin, S., Zdunek, J., Oliveberg, M. & Johansson, L.B.Å. Tryptophan-BODIPY: A versatile donor-acceptor pair for probing generic changes of intraprotein distances. *Phys. Chem. Chem. Phys.* **8**, 3130–3140 (2006).
20. Knight, J.L., Mekler, V., Mukhopadhyay, J., Ebricht, R.H. & Levy, R.M. Distance-restrained docking of rifampicin and rifamycin SV to RNA polymerase using systematic FRET measurements: developing benchmarks of model quality and reliability. *Biophys. J.* **88**, 925–938 (2005).
21. Dolgikh, E., Roitberg, A.E. & Krause, J.L. Fluorescence resonance energy transfer in dye-labeled DNA. *J. Photochem. Photobiol. A Chem.* **190**, 321–327 (2007).
22. VanBeek, D.B., Zwier, M.C., Shorb, J.M. & Krueger, B.P. Fretting about FRET: Correlation between kappa and R. *Biophys. J.* **92**, 4168–4178 (2007).

## Supplement A

23. Sindbert, S. *et al.* Accurate distance determination of nucleic acids via Förster resonance energy transfer: implications of dye linker length and rigidity. *J. Am. Chem. Soc.* **133**, 2463–2480 (2011).
24. Cai, Q. *et al.* Nanometer distance measurements in RNA using site-directed spin labeling. *Biophys. J.* **93**, 2110–2117 (2007).
25. Götte, M., Rausch, J.W., Marchand, B., Sarafianos, S. & Le Grice, S.F.J. Reverse transcriptase in motion: Conformational dynamics of enzyme-substrate interactions. *Biochim. Biophys. Acta* **1804**, 1202–1212 (2010).
26. Peletskaya, E.N., Kogon, A.A., Tuske, S., Arnold, E. & Hughes, S.H. Nonnucleoside inhibitor binding affects the interactions of the fingers subdomain of human immunodeficiency virus type 1 reverse transcriptase with DNA. *J. Virol.* **78**, 3387–3397 (2004).
27. Huang, H., Chopra, R., Verdine, G.L. & Harrison, S.C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282**, 1669–1675 (1998).
28. Rothwell, P.J. *et al.* Multi-parameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 reverse transcriptase: primer/template complexes. *Proc. Natl. Acad. Sci. USA* **100**, 1655–1660 (2003).
29. Patel, P.H. *et al.* Insights into DNA polymerization mechanisms from structure and function analysis of HIV-1 reverse transcriptase. *Biochemistry* **34**, 5351–5363 (1995).
30. Upadhyay, A.K., Talele, T.T. & Pandey, V.N. Impact of template overhang-binding region of HIV-1 RT on the binding and orientation of the duplex region of the template-primer. *Mol. Cell. Biochem.* **338**, 19–33 (2010).
31. Lakowicz, J.R. *Principles of Fluorescence Spectroscopy* 2nd edn. (Springer, 2006).
32. Van Der Meer, B.W., Coker, G. III. & Chen, S.-Y.S. *Resonance Energy Transfer: Theory and Data*. (Wiley, 1994).
33. Sisamakos, E., Valeri, A., Kalinin, S., Rothwell, P.J. & Seidel, C.A.M. Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods Enzymol.* **475**, 455–514 (2010).
34. Antonik, M., Felekyan, S., Gaiduk, A. & Seidel, C.A.M. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J. Phys. Chem. B* **110**, 6970–6978 (2006).
35. Gopich, I.V. & Szabo, A. Single-macromolecule fluorescence resonance energy transfer and free-energy profiles. *J. Phys. Chem. B* **107**, 5058–5063 (2003).
36. Kalinin, S., Felekyan, S., Valeri, A. & Seidel, C.A.M. Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis. *J. Phys. Chem. B* **112**, 8361–8374 (2008).
37. Dale, R.E., Eisinger, J. & Blumberg, W.E. The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer. *Biophys. J.* **26**, 161–193 (1979).
38. Lyskov, S. & Gray, J.J. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res.* **36**, W233–W238 (2008).
39. van Dijk, M. & Bonvin, A. Pushing the limits of what is achievable in protein-DNA docking: benchmarking HADDOCK's performance. *Nucleic Acids Res.* **38**, 5634–5647 (2010).
40. Efron, B. Discussion: jackknife, bootstrap and other resampling methods in regression-analysis. *Ann. Stat.* **14**, 1301–1304 (1986).
41. Brünger, A.T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475 (1992).
42. Case, D.A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
43. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
44. Tokuriki, N. & Tawfik, D.S. Protein dynamism and evolvability. *Science* **324**, 203–207 (2009).
45. Felekyan, S., Kalinin, S., Sanabria, H., Valeri, A. & Seidel, C.A.M. Filtered FCS: species auto- and cross-correlation functions highlight binding and dynamics in biomolecules. *ChemPhysChem* **13**, 1036–1053 (2012).

## Supplement A

### ONLINE METHODS

All samples, applied methods and supporting data are described in detail in the supplementary information. All data used for FPS are compiled in the **Supplementary Tables 2–5**.

**Materials.** Sample preparation and labeling were done according to ref. 46. We used RT mutants containing single accessible cysteine positions that were labeled with Alexa 488 as donor dye. Cy5 was used as acceptor dye and was attached to different positions on a 19/35 DNA primer/DNA template. Further details are given in the **Supplementary Methods** and **Supplementary Figure 1**. The Alexa 488–Cy5 dye pair has a Förster radius  $R_0 = 52 \text{ \AA}$  for  $\langle \kappa^2 \rangle = 2/3$ .

The donor labeling positions (green spheres) are named according to the RT subunit (p51 or p66) and the position of the introduced cysteine mutation to which they were coupled: p51Q6C, p51K173C, p51E194C, p51K281C, p66Q6C, p66T27C, p66E194C and p66K287C. Five acceptor labeling positions on the primer/template DNA (red spheres) are named according to strand (dp or dt) and the position of the labeled nucleotide. The last paired nucleotides are referred to as position 1. For labeled positions on the template overhang, the position is negative with regard to position 1, and for the primer-coupled fluorophores, the number is positive (**Supplementary Fig. 2**). Thus, the labeled dp/dt complexes are named dp(1)/dt, dp(10)/dt, dp(19)/dt, dp/dt(-6) and dp/dt(-15) (**Supplementary Table 1**). As an example, FRET-pairs are named RT(p51E194C):dp/dt(-15).

**Experimental conditions.** The aqueous measurement buffer contained 10 mM KCl, 50 mM Tris·HCl, pH 7.8, 6 mM MgCl<sub>2</sub> and 400  $\mu$ M ascorbic acid. See **Supplementary Methods** for further details.

**Experimental methods.** The multiparameter fluorescence detection (MFD) is performed using an inverted confocal microscope with excitation by a pulsed laser at 496 nm. Fluorescence detection is performed with an additional pinhole defining a detection volume of 2 femtoliters (fl). Sample molecules diffusing through the detection volume cause a brief (~1-ms) burst of fluorescence. Dilute solutions of molecules (~50 pM) ensure that only single molecules are detected. The fluorescence signal is divided into parallel and perpendicular components and in wavelength ranges below and above 620 nm (green and red, respectively), and single photons are detected by four avalanche photodiodes. For each photon, the arrival time after the laser pulse, the time since the last photon, the polarization, and the wavelength range are recorded. Fluorescence bursts are distinguished from the background signal of 1–2 kHz by applying certain threshold intensity criteria<sup>47</sup>. See **Supplementary Methods** for further details.

**Analysis methods.** Distances and their uncertainties (**Supplementary Tables 2 and 3**) were determined by PDA and time-resolved sub-ensemble anisotropies. The positional distribution of certain species was analyzed by sub-ensemble time-correlated single-photon counting (**Supplementary Note 5** and **Supplementary Tables 4 and 5**).

**Toolkit for FPS.** Software is available for download at <http://www.mpc.hhu.de/software>. Details on FPS, especially on the generation

and screening of models, are given in **Supplementary Note 8**. The software can use all types of distance restraints, including, for example, FRET distances derived from surface smFRET or ensemble TCSPC experiments.

**Methods for rigid-body docking of dsDNA with FRET restraints in step 4.** (For more details see **Supplementary Note 8**.)

*Input distances for FPS.* Typically the mean FRET efficiency  $\langle E \rangle$  is measured in an experiment. However, because of different averaging of  $E$  and the D–A distance  $R_{DA}$ ,  $\langle E \rangle$  is not directly related to the distance between mean dye positions  $R_{mp}$  (refs. 6,23). However, a formal distance, referred to as donor-acceptor FRET-averaged distance  $\langle R_{DA} \rangle_E$ , is calculated by  $\langle E \rangle = 1/(1 + \langle R_{DA} \rangle_E^6/R_0^6)$ . The Förster radius  $R_0$  is calculated for  $\langle \kappa^2 \rangle = 2/3$  unless stated otherwise. The optimization problem can be defined for  $R_{mp}$  or  $\langle R_{DA} \rangle_E$  values. The first option is easier to implement in combination with rigid-body dynamics, whereas directly calculating deviations between experimental and model  $\langle R_{DA} \rangle_E$  is more appropriate for structure screening. PDA and similar techniques<sup>48,49</sup> directly provide  $\langle R_{DA} \rangle_E$ . To obtain  $R_{mp}$  values, an  $\langle R_{DA} \rangle_E$ -to- $R_{mp}$  conversion function can be generated (**Fig. 1** and **Supplementary Note 7**).

*Search.* In the first step, we generate a large number of rigid-body models that correspond to local minima of  $\chi_r^2$  (see equation (8.3) in **Supplementary Note 8**). For this, the optimization is started from a random configuration of the binding partners, excluding those with clashes between them. The Verlet algorithm<sup>50</sup> (**Supplementary Note 8**) is applied to model translational and rotational motions of units until the system is relaxed and certain convergence criteria are fulfilled. Although probably not the most efficient optimization algorithm, rigid-body dynamics is straightforward to use and allows exploration of the local minima. In addition, clashes between molecules are prevented by introducing strong repulsive forces between atoms approaching each other by a distance smaller than the sum of their van der Waals radii. In this way, rigid-body dynamics effectively minimizes the reduced chi-squared parameter that accounts for violations of FRET restraints and of van der Waals radii,  $\chi_r^2 = (\chi_E^2 + \chi_{clash}^2)/(n - p)$ , where  $n$  is the number of FRET restraints and  $p$  is the number of degrees of freedom—here, 6. Initially, clashes are allowed to some extent to ensure penetration of DNA into the nucleic acid binding cleft, and structures showing considerable clashes are filtered out later by a  $\chi_r^2$  threshold (see below). To ensure exhaustive sampling of the configurational space of the binding partners, docking with random initial conditions<sup>15</sup> is repeated many (here, 10<sup>4</sup>) times.

*Refinement.* The second step is to remodel the AVs accounting for possible interactions (steric clashes) between the dyes and parts of the biomolecule structure the dyes are not attached to. These modified AVs are then used to calculate new mean dye positions, which is followed by a reoptimization of the structure. At the same time, we reduce the clash tolerance to make clashes between complex partners practically forbidden. This procedure can be repeated several times for each structure until new iterations do not further improve the agreement with experimental data or change the solution significantly.

## Supplement A

**Analysis of docking/screening results.** Solutions are considered ambiguous if the respective  $\chi_r^2$  values do not differ significantly. We typically apply a threshold  $\chi_r^2 < \chi_{r,\min}^2 + [2/(n-p)]^{1/2}$ , which roughly corresponds to the variance of the chi-squared distribution of  $2 \times$  (degrees of freedom)<sup>51</sup> (blue lines in **Fig. 4a,b**). The fact that  $\chi_{r,\min}^2$  is often larger than 1 is attributed to systematic experimental errors and to possible limitations of the AV and/or rigid-body models. In this work, this problem is compensated for by oversampling. Other criteria defining different levels of significance can be applied here in a straightforward way.

**Estimation of precision.** In most of cases, a clear peak assignment to the P-P and P-E state is possible (**Supplementary Tables 2 and 3**). The overall ‘input’ distance uncertainties in equation (2) are then estimated following the propagation rules

$$\Delta R_{DA(\pm)}^2(E, \kappa^2) = \Delta R_{DA(\pm)}^2(E) + \Delta R_{DA}^2(\kappa^2) \quad (3a)$$

In equation (3a), terms indicated with a  $\pm$  subscript can be asymmetric with respect to  $R_{DA}$ . More advanced  $\kappa^2$  estimation procedures<sup>18,52</sup> can be also conveniently incorporated into our set of tools. In four cases, the peak assignment (see **Supplementary Tables 2 and 3**) was unclear owing to comparable amplitudes. In these cases, the errors in distances also include this uncertainty. If two peaks with  $R_{DA(1)}$  and  $R_{DA(2)}$  (with  $R_{DA(1)} < R_{DA(2)}$ ) have comparable amplitudes  $x_1$  and  $x_2$ , the amplitude-weighted average distance  $\langle R \rangle_{E,x} = (a_1 R_{DA(1)} + a_2 R_{DA(2)}) / (a_1 + a_2)$  is taken for modeling as  $R_{DA}$ , whereas the uncertainties are given by equation (3b)

$$\begin{aligned} \Delta R_{DA(+)}(E, \kappa^2, \text{assign}) &= R_{DA(2)} + \Delta R_{DA(2,+)}(E, \kappa^2) - \langle R \rangle_{E,x} \\ \Delta R_{DA(-)}(E, \kappa^2, \text{assign}) &= \langle R \rangle_{E,x} - (R_{DA(1)} - \Delta R_{DA(1,-)}(E, \kappa^2)) \end{aligned} \quad (3b)$$

Procedures to estimate various contributions to the input uncertainties are described in detail in **Supplementary Note 15**.

For each phosphorus atom, the position uncertainty (described by the r.m.s. deviation) is estimated from an ensemble of perturbed structures using equation (4)

$$\text{r.m.s.d.} = \left\langle \left| \vec{R}_{\text{best model}} - \vec{R}_{\text{perturbed model}} \right|^2 \right\rangle^{1/2} \quad (4)$$

MD simulations are described in **Supplementary Note 12**. The discrimination between models and the determination of quality parameters is described in **Supplementary Note 9**.

46. Rothwell, P.J. *Structural Investigations on HIV-1 RT Using Single Pair Fluorescence Energy Transfer* PhD thesis, Univ. Dortmund (2002).
47. Eggeling, C. *et al.* Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *J. Biotechnol.* **86**, 163–180 (2001).
48. Nir, E. *et al.* Shot-noise limited single-molecule FRET histograms: Comparison between theory and experiments. *J. Phys. Chem. B* **110**, 22103–22124 (2006).
49. Santoso, Y., Torella, J.P. & Kapanidis, A.N. Characterizing single-molecule FRET dynamics with probability distribution analysis. *ChemPhysChem* **11**, 2209–2219 (2010).
50. Martyna, G.J., Tobias, D.J. & Klein, M.L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).
51. Soong, T.T. *Fundamentals of Probability and Statistics for Engineers* (Wiley, 2004).
52. Isaksson, M., Norlin, N., Westlund, P.O. & Johansson, L.B.Å. On the quantitative molecular analysis of electronic energy transfer within donor-acceptor pairs. *Phys. Chem. Chem. Phys.* **9**, 1941–1951 (2007).

## Supplement A

### Nature Methods

# A toolkit and benchmark study for FRET-restrained high-precision structural modeling

Stanislav Kalinin<sup>1#</sup>, Thomas Peulen<sup>1,5</sup>, Simon Sindbert<sup>1,5</sup>, Paul J. Rothwell<sup>1/2</sup>, Sylvia Berger<sup>1</sup>, Tobias Restle<sup>2/3</sup>, Roger S. Goody<sup>2</sup>, Holger Gohlke<sup>4</sup>, Claus A. M. Seidel<sup>1#</sup>

<sup>1</sup>Lehrstuhl für Molekulare Physikalische Chemie, Heinrich-Heine-Universität, Universitätsstraße 1, 40225 Düsseldorf, Germany, <sup>2</sup>Max-Planck-Institut für molekulare Physiologie, Otto-Hahn-Str. 11, 44227 Dortmund, Germany, <sup>3</sup>Institut für Molekulare Medizin, Universität zu Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany, <sup>4</sup>Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität, Universitätsstr. 1, 40225 Düsseldorf, Germany

<sup>5</sup> contributed equally

# corresponding authors: stanislav.kalinin@uni-duesseldorf.de, cseidel@hhu.de.

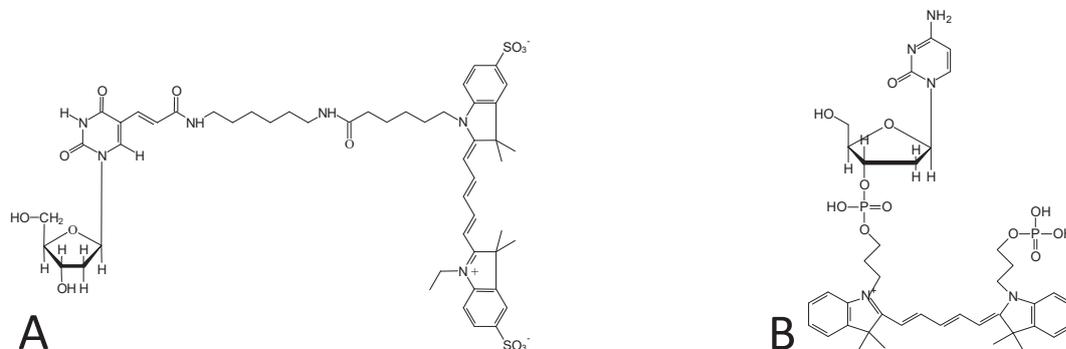
Supplementary Figure 1	Structures of modified nucleotides
Supplementary Figure 2	Secondary structure of dp/dt
Supplementary Table 1	dt/dp sequences
Supplementary Table 2	Distances $\langle R_{DA} \rangle_E$ of the major population resulting from PDA for dsDNA
Supplementary Table 3	Distances $\langle R_{DA} \rangle_E$ of the major population resulting from PDA for ssDNA
Supplementary Table 4	seTCSPC distances of FRET-populations for ssDNA
Supplementary Table 5	seTCSPC distances of FRET-populations for dsDNA
Supplementary Note 1	Dye distributions by the AV approach
Supplementary Note 2	Justification for the use of crystal structure 1R0A
Supplementary Note 3	Prior Knowledge: States expected to be observed for the HIV-1 RT:dp/dt complex
Supplementary Note 4	Distances through PDA
Supplementary Note 5	Sub-ensemble TCSPC (seTCSPC) of FRET data
Supplementary Note 6	Estimation of the $\kappa^2$ related distance uncertainties
Supplementary Note 7	Obtaining mean position distances: additional details
Supplementary Note 8	Rigid body docking: implementation details
Supplementary Note 9	Cluster analysis on dsDNA
Supplementary Note 10	Influence of the clash tolerance
Supplementary Note 11	FRET-guided positioning and screening for ds B-DNA
Supplementary Note 12	Structure ensemble generation via MD simulation
Supplementary Note 13	Overfitting tests (dsDNA)
Supplementary Note 14	FRET-guided screening for the template overhang
Supplementary Note 15	Peak assignment uncertainties

## Supplement A

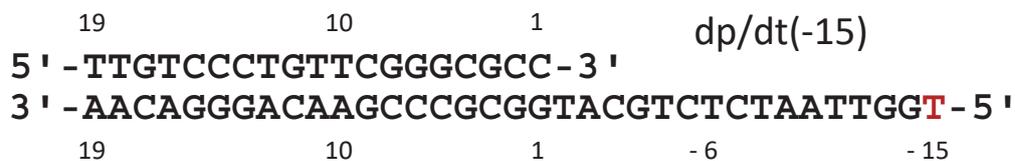
Supplementary Data 1	smFRET data sets
Supplementary Data 2	seTCSPC data sets
Supplementary Data 3	Obtained $\kappa^2$ distributions
Supplementary Data 4	PDB file for the FRET-restrained structural model of the RT:dp/dt complex including the single strand template overhang
Supplementary Methods	HIV-RT:dp/dt complex, Measurement conditions, and Multiparameter Fluorescence Detection (MFD)
Supplementary Software	Software for FRET-restrained positioning and screening with help file and test data

## Supplement A

### Supplementary Figures



**Supplementary Figure 1. Structures of the fluorescent modified nucleotides used.** (A) dThymine labeled with C6-aminolink with the NHS-ester of Cy5 used at positions dp(10), dp(19), dt(-6) and dt(-15). (B) 3' labeled dCytosine with the phosphoamidite derivative of Cy5 used for position dp(1).



**Supplementary Figure 2.** Secondary structure of DNA dp/dt(-15) with labeling position (red).

## Supplement A

### Supplementary Tables

**Supplementary Table 1.** Sequences and labeling positions (red) of primer (dp) and template (dt) strands.

sample	sequence and labeling position
dp	5'-d(TTGTCCCTGTTTCGGGCGCC)-3'
dp(1)	5'-d(TTGTCCCTGTTTCGGGCGCC)-3'
dp(10)	5'-d(TTGTCCCTGTTTCGGGCGCC)-3'
dp(19)	5'-d(TTGTCCCTGTTTCGGGCGCC)-3'
dt	5'-d(TGGTTAATCTCTGCATGGCGCCCGAACAGGGACAA)-3'
dt(-6)	5'-d(GGGTTAATCTCTGCATGGCGCCCGAACAGGGACAA)-3'
dt(-15)	5'-d(TGGTTAATCTCTGCATGGCGCCCGAACAGGGACAA)-3'

## Supplement A

**Supplementary Table 2.** Distances  $\langle R_{DA} \rangle_E$  (amplitudes in brackets) of the major population resulting from PDA used for structural modeling of dsDNA, corresponding  $\kappa^2$ -errors  $\Delta R_{DA}(\kappa^2)$ , fitted apparent widths of  $\langle R_{DA} \rangle_E$ -distributions  $\sigma_{app}$ , and average amplitude weighted distances  $\langle R \rangle_{E,x}$ , with corresponding overall absolute errors  $\Delta R_{DA(\pm)}(E, \kappa^2)$  (corresponds to  $\Delta R_{DA}(\kappa^2)$ ) or  $\Delta R_{DA(+,-)}(E, \kappa^2, \text{assign})$ , in the table referred to as  $\text{err}_{+/-}$ . (see equation 3a and 3b in Online Methods) and the used experimental setup.

sample	$\langle R_{DA} \rangle_E, \text{\AA}$ (fraction) <sup>[a]</sup>	$\Delta R_{DA}(\kappa^2)$	$\sigma_{app}/R$	$\langle R \rangle_{E,x}$ , $\text{\AA}$	$\text{err}_{+}$ , $\text{\AA}$	$\text{err}_{-}$ , $\text{\AA}$	setup <sup>[b]</sup>
RT(p66Q6C):dp(1)/dt	46(70%)	9.8%	7.8%	46	4.5	4.5	b
RT(p66Q6C):dp(10)/dt	46(58%)	9.9%	5.9%	46	4.5	4.5	b
RT(p66Q6C):dp(19)/dt	73(69%)	9.0%	8.5%	73	6.4	6.4	a
RT(p66T27C):dp(1)/dt	30(20%)	8.8%	9.0%	30	2.8	2.8	b
RT(p66T27C):dp(19)/dt	73(59%)	8.8%	8.0%	73	6.4	6.4	a
RT(p66E194C):dp(1)/dt	41(52%)	7.4%	7.2%	41	3.1	3.1	c
RT(p66E194C):dp(10)/dt	44(35%)	7.3%	7.9%	44	3.2	3.2	c
RT(p66E194C):dp(19)/dt	83(58%)	7.3%	9.2%	83	6.1	6.1	c
RT(p66K287C):dp(1)/dt	44(26%)	10.3%	9.0%	44	4.5	4.5	c
RT(p66K287C):dp(19)/dt	45(50%)	9.6%	9.1%	45	4.4	4.4	a
RT(p51Q6C):dp(1)/dt	68(45%) 76(36%)	9.3%	9.0%	71	10.5	9.8	b
RT(p51Q6C):dp(19)/dt	65(80%)	9.4%	7.3%	65	6.2	6.2	b
RT(p51K173C):dp(1)/dt	60(61%)	10.1%	7.1%	60	6.0	6	b
RT(p51K173C):dp(10)/dt	62(55%)	10.0%	8.5%	62	6.2	6.2	b
RT(p51K173C):dp(19)/dt	67(71%)	10.0%	9.0%	67	6.6	6.6	b
RT(p51E194C):dp(1)/dt	63(37%) 79(32%)	7.3%	9.0%	70.6	13.9	12.9	c
RT(p51E194C):dp(10)/dt	69(53%)	7.3%	9.0%	68	5.1	5.1	c
RT(p51E194C):dp(19)/dt	55(40%)	7.4%	5.0%	55	4.1	4.1	c
RT(p51K281C):dp(10)/dt	81(54%)	9.6%	9.0%	81	7.8	7.8	b
RT(p51K281C):dp(19)/dt	35(37%)	10.0%	6.2%	35	3.5	3.5	b

<sup>[a]</sup> In two cases the major peak could not be clearly assigned. In these cases the weighted mean was used for modeling.

<sup>[b]</sup>The setups have the following green/red-detection efficiency ratios: setup a:  $g_G/g_R=0.80$ ; setup b:  $g_G/g_R=0.54$ ; setup c:  $g_G/g_R=0.70$

## Supplement A

**Supplementary Table 3.** Distances  $\langle R_{DA} \rangle_E$  (amplitudes in brackets) of the major population resulting from PDA of ssDNA, corresponding  $\kappa^2$ -errors  $\Delta R_{DA}(\kappa^2)$ , fitted apparent widths of  $\langle R_{DA} \rangle_E$ -distributions  $\sigma_{app}$ , and average amplitude weighted distances  $\langle R \rangle_{E,x}$ , with corresponding overall absolute errors  $\Delta R_{DA(\pm)}(E, \kappa^2)$  (corresponds to  $\Delta R_{DA}(\kappa^2)$ ) or  $\Delta R_{DA(+,-)}(E, \kappa^2, \text{assign})$ , in the table referred to as  $\text{err}_{+/-}$  (see equations 3a and 3b in Online Methods) and the used experimental setup.

Sample	$\langle R_{DA} \rangle_E, \text{\AA}$ (fraction) <sup>[a]</sup>	$\Delta R_{DA}(\kappa^2)$	$\sigma_{app}/R$	$\langle R \rangle_{E,x}, \text{\AA}$	$\text{err}_+, \text{\AA}$	$\text{err}_-, \text{\AA}$	setup <sup>[b]</sup>
RT(p66Q6C):dp/dt(-6)	48(53%)	9.1%	10%	48	4.4	4.4	a
RT(p66Q6C):dp/dt(-15)	49(61%)	9.1%	10%	49	4.5	4.5	a
RT(p66T27C):dp/dt(-6)	34(7%) 42(4%)	8.8%	13%	37	5.9	8.8	a
RT(p66T27C):dp/dt(-15)	29(30%) 45(23%)	9.0%	7%	36	9.6	13.1	b
RT(p66E194C):dp/dt(-6)	44(34%)	7.2%	6%	44	3.2	3.2	c
RT(p66E194C):dp/dt(-15)	46(28%)	7.2%	7%	46	3.3	3.3	c
RT(p66K287C):dp/dt(-6)	36(34%)	10.0%	7%	36	3.6	3.6	a
RT(p66K287C):dp/dt(-15)	43(32%)	10.1%	13%	43	4.3	4.3	b
RT(p51Q6C):dp/dt(-6)	76(82%)	9.4%	11%	76	7.1	7.1	b
RT(p51Q6C):dp/dt(-15)	74(83%)	9.5%	7%	74	7.0	7.0	b
RT(p51K173C):dp/dt(-6)	62(63%)	10.0%	8%	62	6.2	6.2	b
RT(p51K173C):dp/dt(-15)	61(69%)	10%	9%	61	6.1	6.1	b
RT(p51E194C):dp/dt(-6)	68(53%)	7.2%	8%	68	4.9	4.9	c
RT(p51E194C):dp/dt(-15)	68(38%)	7.4%	7%	68	5.0	5.0	c
RT(p51K281C):dp/dt(-6)	76(76%)	10.0%	10%	76	7.6	7.6	a
RT(p51K281C):dp/dt(-15)	77(59%)	9.9%	9%	77	7.6	7.6	a

<sup>[a]</sup> In two cases the major peak could not be clearly assigned. In these cases the weighted mean was used for modeling.

<sup>[b]</sup> The setups have the following green/red-detection efficiency ratios: setup a:  $g_G/g_R=0.80$ ; setup b:  $g_G/g_R=0.54$ ; setup c:  $g_G/g_R=0.70$

## Supplement A

**Supplementary Table 4:** seTCSPC distances of FRET-populations of the major state in the single stranded overhang selected by the  $F_D/F_A$ -ratio.

Sample	$R_{DA}$ [Å]	$\sigma_{DA}$ [Å]	$x_{DOnly}$ [%]	$F_D/F_A$	$S_{\parallel}$	$S_{\perp}$	$\chi^2_r$
RT(p51Q6C):dp/dt(-6)	74.7	n.r.		all	0.1	0.1	2.4
RT(p51Q6C):p/t(-15)	74.3	n.r.		all	0.1	0.1	2.3
RT(p51K173C):dp/dt(-6)	62.3	n.r.		all	0.2	0.1	1.2
RT(p51K173C):p/t(-15)	62.5	n.r.		0.1-35	0.3	0.1	1.5
RT(p51E194C):dp/dt(-6)	69.6	n.r.	20	all	0.3	0.3	2.6
RT(p51E194C):p/t(-15)	75.2	n.r.		0.1-6	0.3	0.2	1.2
RT(p51K281C):dp/dt(-6)	70.2	n.r.	20	all	0.2	0.1	2
RT(p51K281C):p/t(-15)	72.1	n.r.		all	0.2	0.1	1.8
RT(p66Q6C):dp/dt(-6)	55.8	15		0.1-10	0.3	0.2	1.3
RT(p66Q6C):p/t(-15)	49.4	19.7		0.1-10	0.2	0.1	1.2
RT(p66T27C):dp/dt(-6)	43.3	18		0.1-6	0.5	0.3	1
RT(p66T27C):p/t(-15)	45.5	17		0.1-8	0.3	0.2	1.2
RT(p66E194C):dp/dt(-6)	58	11		0.1-12	0.7	0.3	1
RT(p66E194C):p/t(-15)	53.8	12		0.6-3.5	0.5	0.2	1.2
RT(p66K287C):dp/dt(-6)	37.2	11		0.1-6	0.2	0.2	1
RT(p66K287C):p/t(-15)	42.1	11		0.1-11	0.2	0.1	1.2

Meaning of the parameters: donor/acceptor fluorescence intensity ratio  $F_D/F_A$ , donor-acceptor distance  $R_{DA}$ , the half width of the fitted Gaussian donor-acceptor distance distribution  $\sigma_{DA}$ , the donor-only fraction  $x_{DOnly}$ , the scatter amplitude in the vertical and horizontal channel, respectively  $S_{\parallel}$ ,  $S_{\perp}$ . n.r. ( not resolved).

## Supplement A

**Supplementary Table 5.** seTCSPC of FRET-population selected by  $F_D/F_A$ -ratio located on the double-stranded DNA.

Sample	$R_{DA}$ [Å]	$\sigma_{DA}$ [Å]	$x_{DOnly}$ [%]	$F_D/F_A$	$s_{\parallel}$	$s_{\perp}$	$\chi^2_r$
RT(p51E194C):p(1)/t	75.3	n.r.		0.1-6	0.5	0.3	1.2
RT(p51E194C):p(10)/t	73.8	n.r.		9-55	0.4	0.2	1.1
RT(p51E194C):p(19)/t	62.2	6.6		9-74	0.3	0.2	1.3
RT(p51K173C):p(1)/t	64.9	n.r.		0.1-100	0.2	0.1	1.6
RT(p51K173C):p(10)/t	64.9	n.r.		0.1-40	0.3	0.1	1.4
RT(p66E194):p(1)/t	50.2	11.7		0.1-6	0.6	0.3	1.7
RT(p66E194):p(10)/t	52.5	19.9		0.1-16	0.4	0.2	2.3
RT(p66E194):p(19)/t	75.2	n.r.	20	5-1200	0.5	0.2	3.4
RT(p66K287C):p(1)/t	40.0	14.5		0.05-5	0.2	0.1	1.2
RT(p66K287C):p(19)/t	31.4	15.1		0.05-7	0.1	0.1	1.4
RT(p66Q6C):p(1)/t	45.9	11.2		0.1-5	0.2	0.1	1.3
RT(p66Q6C):p(10)/t	45.8	10.3		0.1-10	0.3	0.1	1.1
RT(p66Q6C):p(19)/t	77.1	15.3		0.2	0.3	0.1	2.2
RT(p66T27C):p(1)/t	45.8	12.5		0.05-5	0.6	0.4	1.1
RT(p66T27C):p(19)/t	83.0	n.r.		5-1000	0.3	0.1	1.6
RT(p51E194C):p(1)/t	75.3	n.r.		0.1-6	0.5	0.3	1.2

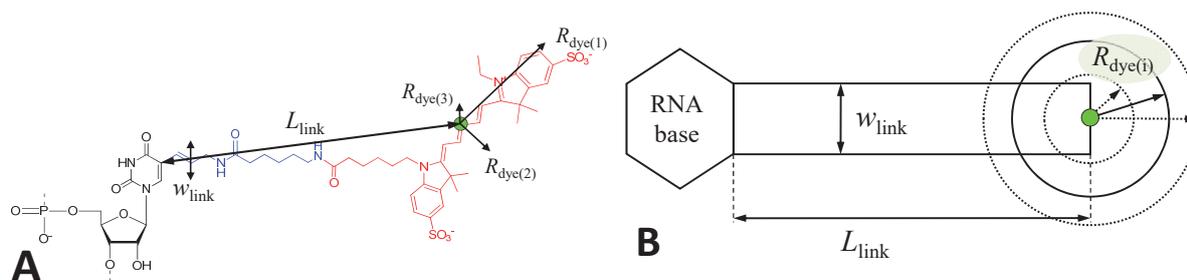
Meaning of the parameters: donor/acceptor fluorescence intensity ratio  $F_D/F_A$ , donor-acceptor distance  $R_{DA}$ , the half width of the fitted Gaussian donor-acceptor distance distribution  $\sigma_{DA}$ , the donor-only fraction  $x_{DOnly}$ , the scatter amplitude in the vertical and horizontal channel, respectively  $s_{\parallel}$ ,  $s_{\perp}$ . n.r. ( not resolved).

## Supplementary Notes

## Supplementary Note 1: Dye distributions by the AV approach.

We model dye distributions by the AV approach<sup>1,2</sup> according to<sup>3</sup>. The dyes are approximated by a sphere with an empirical radius of  $R_{\text{dye}}$ , where the central atom of the fluorophore is connected by a flexible linkage of a certain effective length  $L_{\text{link}}$  and width  $w_{\text{link}}$  to the nucleobases of the DNA (see Figure N1.1) or the  $C_{\beta}$  atoms for RT. In case of dp(1)/dt the linker was connected to the phosphorous atom of the DNA backbone. The overall length of the linkage is given by the actual length of the linker and the internal chemical structure of the dye. A geometric search algorithm finds all dye positions within the linkage length from the attachment point which do not cause steric clashes with the macromolecular surface. All allowed positions are considered as equally probable which allows one to define an accessible volume for the dye (AV). We used typical parameters for the linkage width ( $w_{\text{link}}=4.5$  Å) from<sup>2</sup>. The linkage lengths ( $L_{\text{link}}$ ) were estimated from the fully extended conformations of each linker using the Hyperchem software<sup>4</sup> and are listed in Table N1.1 and in the caption of Fig. 2 in the main text. To take the three quite different spatial dimensions of Cy5 (see Figure N1.1) into account, we used its real physical dimensions for each calculation of a position distribution and performed three independent AV simulations with three different radii  $R_{\text{dye}(i)}$  and superimposed them. Thus, the obtained position distribution represents an average weighted by the number of allowed positions. See Table N1.1 for values used for  $R_{\text{dye}(i)}$ ,  $L_{\text{link}}$  and  $w_{\text{link}}$ .

Before performing rigid body docking (see “Step 4a” in the main text), to suppress effects arising from labeling positions at the end of nucleic acids (dp(19)/dt, see Figure N1.1), we extend the dp/dt molecule extracted from the RT:dt/dp crystal structure (1R0A) and the B-DNA (see “Step 1” in the main text) by four base pairs (B-DNA form) at the 5’ end of the primer strand. In a previous study with linear dsRNA, this procedure has been shown to improve results<sup>3</sup>.



**Figure N1.1:** Molecular drawing (A) and sketch (B) for the AV parameters.  $R_{\text{dye}(i)}$ ,  $L_{\text{link}}$  and  $w_{\text{link}}$  indicated by arrows for an RNA nucleotide labeled with Cy5 (red).

## Supplement A

**Table N1.1:** Values used for  $R_{\text{dye}(i)}$ ,  $L_{\text{link}}$  and  $w_{\text{link}}$ .

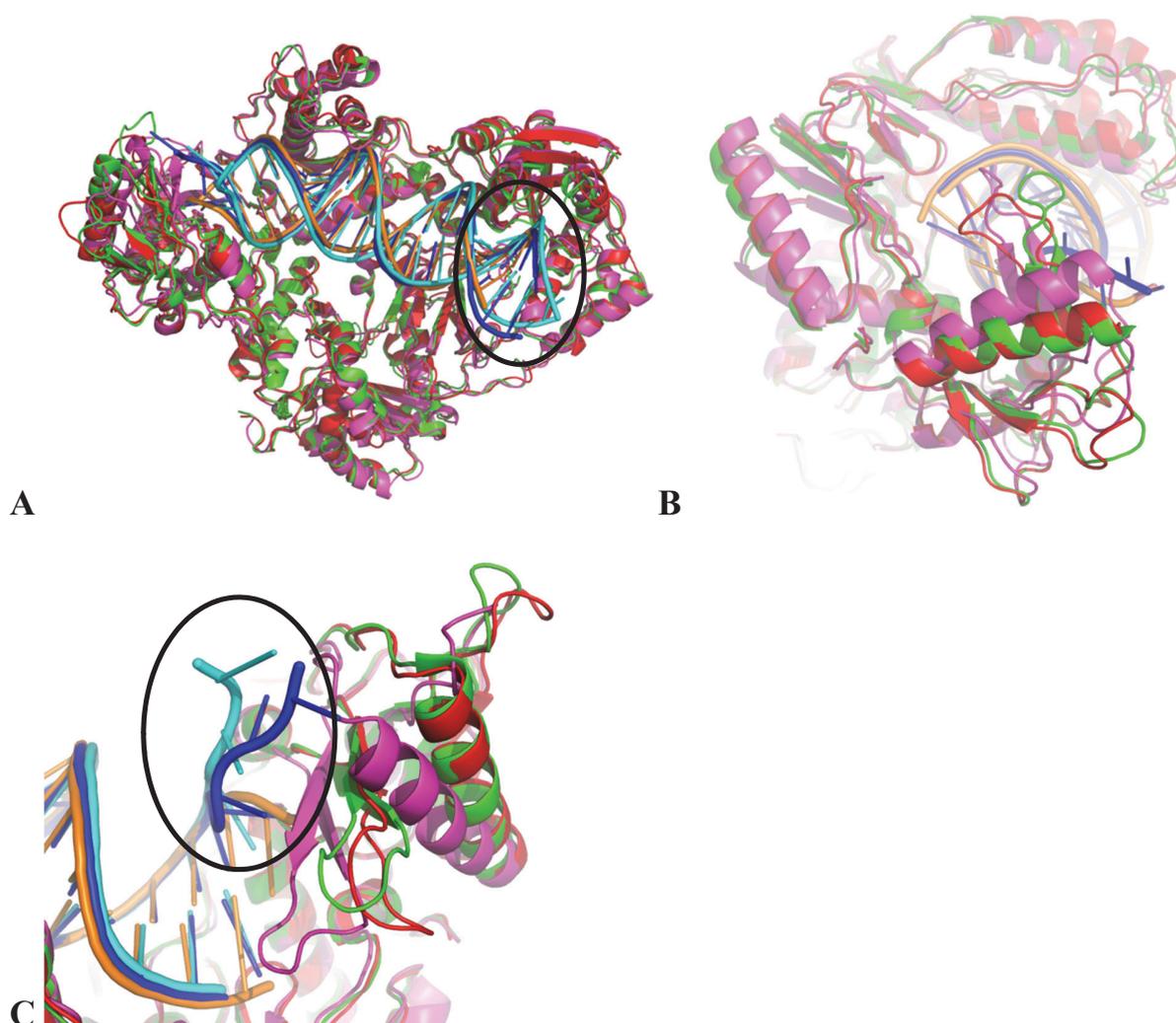
Donor positions	
$L_{\text{link}}$	20 Å
$w_{\text{link}}$	4.5 Å
$R_{\text{dye}}$	3.5 Å
Acceptor positions dp(10), dp(19), dt(-6) and dt(-15)	
$L_{\text{link}}$	23 Å
$w_{\text{link}}$	4.5 Å
$R_{\text{dye}(1)}$	11 Å
$R_{\text{dye}(2)}$	3 Å
$R_{\text{dye}(3)}$	1.5 Å
Acceptor position dp(1)	
$L_{\text{link}}$	8.5 Å
$w_{\text{link}}$	4.5 Å
$R_{\text{dye}(1)}$	11 Å
$R_{\text{dye}(2)}$	3 Å
$R_{\text{dye}(3)}$	1.5 Å

### AV algorithm: implementation details

AV simulations were performed similar to as described previously<sup>2</sup>. The AV algorithm finds fluorophore positions in space for which (1) the fluorophore does not clash with macromolecules and (2) it is possible to route the linker (approximated as a flexible pipe) from the linker attachment point to the fluorophore position. At first, a 3D grid is built around the linker attachment point with a default spacing of 1/5 of the smallest dye or linker dimension, but no less than 0.4 Å (default values; “AV grid (rel)” and “Min. grid” in FPS software, respectively). The total number of nodes is determined by the grid step and the maximum distance ( $L_{\text{link}}$ ) from the attachment atom placed at the origin. Second, positions that clash with the macromolecule are disregarded. For this, all grid nodes are checked for the proximity to the macromolecule considering the assumed dye radius, the known structure of the macromolecule, and known van der Waals radii of the atoms. Third, for each allowed position the shortest route to the attachment point is calculated as follows. Initially, to prevent the linker from being blocked close to the attachment point, all positions within  $0.5 \times w_{\text{link}}$  (“Allowed sphere” in FPS software) are allowed for the linker regardless of clashing. For each of these positions, all neighboring positions within 3 nodes (default; “Search nodes” in FPS software) are scanned. For these nodes, a possible route distance from the attachment point is calculated as a sum of the route distance calculated for the previous position, and the distance to that position. For each position the shortest possible distance is saved. This procedure is repeated starting from all positions for which the distance has been updated during the previous iteration, until no new positions are discovered. Finally, the position is considered populated if it is sterically allowed for the dye and the route distance is smaller than  $L_{\text{link}}$ .

## Supplement A

### Supplementary Note 2: Justification for the use of crystal structure 1R0A



**Figure N2.1:** Comparison of different crystal structures of HIV-RT:DNA complexes in the open P-E-state (2HMI<sup>5</sup>: red/orange and 1R0A<sup>6</sup>: green/blue) and the closed P-P state (1RTD<sup>7</sup>: magenta/cyan). Different views (A, B, C) are presented.

We use the crystal structure 1R0A for the extraction of rigid bodies for RT and dp/dt and for the validation of our approach by comparison to our docked structures (see main text). As can be seen in Figure N2.1B, the used crystal-structure 1R0A<sup>6</sup> is very similar to the crystal-structure 2HMI<sup>5</sup>. Both are in the open P-E-state. The main difference is that the bound DNA employed in the crystallization in 1R0A is longer (Figure N2.1 A and C). The double-stranded DNA (Figure N2.1A) and the single-stranded 5'-overhang (Figure N2.1C) used for crystallization are both longer than those used for 2HMI. In the single-molecule experiments the used DNA-fragment was bigger than the one used in the crystallization of 2HMI and contained always a 5'-overhang. Therefore, we docked the bodies of RT and dp/dt taken from 1R0A.

## Supplement A

The RMSD between these three structures calculated for P atoms of the nucleotides present in all structures is 2.8 Å, which gives an idea of the precision of the coordinates of the X-ray structures.

## Supplement A

### **Supplementary Note 3: Prior Knowledge: States expected to be observed for the HIV-1 RT:dp/dt complex**

Under the given measurement conditions the RT:dp/dt complexes adopts different conformational states designated as product complex in educt state (P-E), product-complex in product state (P-P) and a dead-end complex (D-E). The P-P and P-E are two species generally close in transfer efficiency representing different steps in the polymerization reaction<sup>8</sup>. The dead-end complex (D-E) is most likely not directly contributing to the enzymatic activity of the protein<sup>8</sup>. The measurements were performed in absence of dNTPs and pyrophosphate. Thus, all three states are present under the given measurement conditions. However, as the protein is preferentially in the educt-state (P-E), the state with the highest amplitude was assigned to the P-E the state and the less populated state to the P-P state. Depending on the relative orientation and distance of the donor and acceptor fluorophores the two states could not always be resolved as the change in transfer efficiency upon transition from P-E to P-P is not big enough. Hence, the states P-E and P-P were sometimes fitted with one state. Only a small fraction of the protein/DNA complex is in the enzymatically not contributing D-E complex. Thus an assignment of the P-E state is in most measurements unambiguous.

## Supplement A

### Supplementary Note 4: Distances through PDA

#### 4.1. Basic theory of PDA

#### 4.2. Species needed in PDA

#### 4.3. Linker corrected intensity-lifetime dependence

#### 4.4. Statistical uncertainties in PDA

#### 4.5. FRET populations are static on the ms timescale: sub-ensemble FCS

#### 4.1. Basic theory of PDA

As specified in the main text, we calculate distances by measuring the fluorescence intensities of D and A ( $F_D$  and  $F_A$ , respectively).  $R_{DA}$  is then given by combining equation (1) and (2) in the main text which yields equation (4.1):

$$R_{DA} = \left( \frac{\Phi_{FA} F_D}{\Phi_{FD(0)} F_A} \right)^{\frac{1}{6}} R_0 \quad (4.1)$$

or equivalently by equation (4.2):

$$R_{DA} = \left( \Phi_{FA} \frac{F_D}{F_A} \right)^{\frac{1}{6}} R_{0r} \quad (4.2)$$

Here, we use the reduced Förster radius  $R_{0r}$ <sup>9</sup>, which, in contrast to  $R_0$ , does not depend on  $\Phi_{FD(0)}$ . Throughout this work we used  $\Phi_{FA}=0.32$  ( $\Phi_{FA}=0.32$  due to the presence of ~ 20% cis-trans isomerization instead of  $\Phi_{FA}=0.4$  which is expected for ensemble measurements<sup>10</sup>) and  $R_{0r} = 53.97 \text{ \AA}$  (using  $R_{0r} = 53.97 \text{ \AA}$  is equivalent to  $R_0 = 52 \text{ \AA}$  and  $\Phi_{FD(0)} = 0.8$ ).  $F_D$  and  $F_A$  can be calculated from the signals measured in the green and red detection channels  $S_G$  and  $S_R$ , respectively, via equations (4.3) and (4.4):

$$F_D = \frac{F_G}{g_G} = \frac{S_G - \langle B_G \rangle}{g_G} \quad (4.3)$$

$$F_A = \frac{F_R}{g_R} = \frac{S_R - \alpha F_G - \langle B_R \rangle}{g_R} \quad (4.4)$$

where  $F_G$  and  $F_R$  are the fluorescence signals in the green and the red signal channels, respectively,  $\alpha$  is the crosstalk factor which is determined as the ratio between donor photons detected in the red channels

## Supplement A

and those detected in the green channels ( $\alpha = F_{R(D)}/F_{G(D)}$ ) for the D only labeled sample,  $g_G$  and  $g_R$  are the detection efficiencies in the green and red channels, respectively, and  $\langle B_G \rangle$  and  $\langle B_R \rangle$  are the mean background intensities in the green and red channels, respectively. For the determination of  $R_{DA}$  we only need to know the ratio between  $F_D$  and  $F_A$ , therefore, we only need to know the ratio between the detection efficiencies  $g_G/g_R$  (see below for the determination of  $g_G/g_R$ ).

To accurately predict the shape of  $S_G/S_R$  (or equivalently  $F_D/F_A$ ) histograms in the presence of FRET we use PDA, which explicitly takes into account shot noise, background contributions and additional broadening due to complex acceptor photophysics<sup>11-14</sup>. PDA calculates the probability of observing a certain combination of photon counts  $P(S_G, S_R)$

$$P(S_G, S_R) = \sum_{F_G+B_G=S_G, F_R+B_R=S_R} P(F)P(F_G, F_R|F)P(B_G)P(B_R) \quad (4.5)$$

The intensity distribution of the fluorescence only contribution to the signal,  $P(F)$ , is obtained from the total measured signal intensity distribution  $P(S)$  by deconvolution assuming that the background signals  $B_G$  and  $B_R$  obey Poisson distributions,  $P(B_G)$  and  $P(B_R)$ , with known mean intensities  $\langle B_G \rangle$  and  $\langle B_R \rangle$ .  $P(F_G, F_R | F)$  represents the conditional probability of observing a particular combination of green and red fluorescence photons,  $F_G$  and  $F_R$ , provided the total number of registered fluorescence photons is  $F$ , and can be expressed as a binomial distribution<sup>13</sup>. Subsequently,  $P(S_G, S_R)$  may be further manipulated to generate a theoretical histogram of any FRET-related parameter as discussed elsewhere<sup>12</sup>.

### 4.2. Species needed in PDA

Typically a model was used which accounts for two FRET-populations representing the educt-state (P-E) and the product-state (P-P), a D-only population, and, if necessary, populations accounting for impurities and for a dead-end complex (see Supplementary Note 3) were added as well (see Supplementary Data 1 for the assignment procedure of the states in MFD plots and in PDA). The FRET states were fitted using the same value for the global relative additional (beyond the shot noise) width  $\sigma_{app}$  (see above). In general, for a model with  $n$  FRET states and a D-only fraction,  $2n+1$  fit parameters are needed:  $n$  mean DA distances,  $n$  fractions, and an additional (beyond the shot noise) distribution width  $\sigma_{app}$  expressed as a fraction of the corresponding mean distance. This additional distribution width can be attributed mainly to complex acceptor photophysics and thus can be fitted globally as justified in<sup>15</sup>. As a result, PDA needs much fewer free parameters than the classical approach of fitting multiple Gaussian peaks, which requires up to  $3n+1$  parameters ( $n$  mean DA distances,  $n$  fractions,  $n+1$  peak widths).

### 4.3. Linker corrected intensity-lifetime dependence

To be able to convert model distances into probabilities of observing green photons, the detection efficiency ratio  $g_G/g_R$  is needed (see equations (4.3) and (4.4)). The smFRET measurements were performed over a period of more than one year. For the individual measurement sessions the  $g_G/g_R$ -ratio of the setup was calibrated for each measurement session by requiring that the linker-corrected static FRET line<sup>14</sup> goes through the observed FRET populations in the 2D histograms of  $F_D/F_A$  vs  $\tau_{D(A)}$  (see e.g. Fig. 3a main text, see Supplementary Tables 2 and 3 for the determined values for  $g_G/g_R$ ). The linker-corrected static FRET lines were calculated with the following equation:

## Supplement A

$$\begin{aligned}
 F_D/F_A &= \frac{\Phi_{FD(0)}}{\Phi_{FA}} \left/ \left( \frac{\langle \tau_{D(0)} \rangle_x}{\langle \tau_{D(A)} \rangle_x} - 1 \right) \right. \\
 &= \frac{\Phi_{FD(0)}}{\Phi_{FA}} \left/ \left( \frac{\langle \tau_{D(0)} \rangle_x}{c_3 \langle \tau_{D(A)} \rangle_f^3 + c_2 \langle \tau_{D(A)} \rangle_f^2 + c_1 \langle \tau_{D(A)} \rangle_f + c_0} - 1 \right) \right.
 \end{aligned} \tag{4.6}$$

whereas  $\langle \tau_{(i)} \rangle_x$  and  $\langle \tau_{(i)} \rangle_f$  are species and fluorescence averaged lifetimes, respectively. The polynomial coefficients in equation (4.6) are compiled in Table N4.1. They are obtained as described in <sup>14</sup> with one modification. In this work we also consider the fact that in most of cases the fluorescence decay of D dyes (Alexa488) bound to RT is multi-exponential. Thus, the formalism described in <sup>14</sup> has to be extended as follows. First, the donor fluorescence decay is formally fitted using a multi-exponential relaxation model (see Table N4.1). For each donor sub-species, we assume that its quantum yield is proportional to the corresponding fluorescence lifetime, and calculate the donor and acceptor fluorescence (equation (1) in the main text) as well as the lifetime distribution in the presence of FRET (equation (1) in the main text) as described <sup>14</sup>. The total donor and acceptor fluorescence intensities are weighed sums of those calculated for all donor sub-species, and the overall lifetime distribution is a superposition of species-specific lifetime distributions, weighted by the species fractions. From these data we can calculate  $\langle \tau \rangle_f$  and  $\langle E \rangle$  ( $\langle E \rangle = \langle F_A \rangle / [\gamma \langle F_D \rangle + \langle F_A \rangle]$ ) for any given  $\langle R_{DA} \rangle$ . This procedure is then repeated for a series of  $\langle R_{DA} \rangle$  values, yielding corresponding sets of  $\langle \tau \rangle_f$  and  $\langle E \rangle$ . This dependence is approximated with a polynomial function (equation (4.6)). A program for computing various FRET lines is available from the authors (see [www.mpc.hhu.de/software](http://www.mpc.hhu.de/software)).

## Supplement A

**Table N4.1.** Average fluorescence weighted lifetimes  $\langle \tau_{D(0)} \rangle_f$ , quantum yields  $\Phi_{FD(0)}$ , polynomial coefficients  $c_i$ , and estimated widths of donor/acceptor distribution  $\sigma$  used for the calculation of the static FRET-lines in Supplementary Data 1. In all measurements  $\Phi_{FA} = 0.32$ .

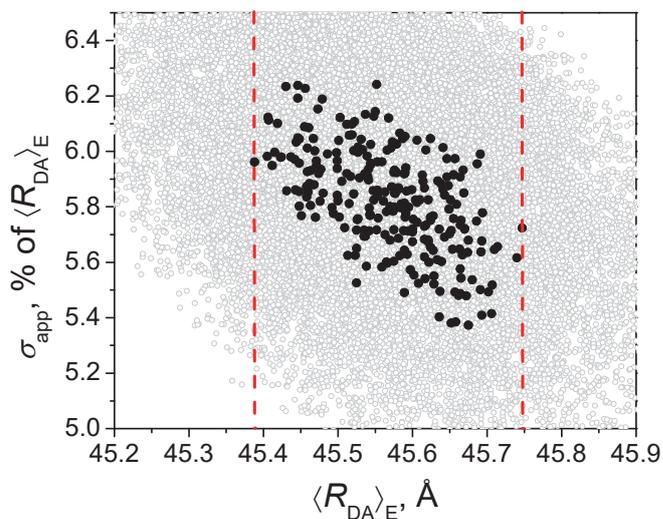
Sample	$\langle \tau_{D(0)} \rangle_f$	$\Phi_{FD(0)}$ <sup>[a]</sup>	$c_0$	$c_1$	$c_2$	$c_3$	$\sigma [A]$
RT(p66Q6C):dp/dt	3.558	0.635	-0.0560	0.6002	0.3064	-0.0534	6
RT(p66T27C):dp/dt	3.513	0.605	-0.0599	0.6471	0.2925	-0.0534	6
RT(p66E194C):dp/dt	3.870	0.755	-0.0425	0.4862	0.2927	-0.0406	12
RT(p66K287C):dp/dt	3.581	0.623	-0.0606	0.6392	0.2927	-0.0524	6
RT(p51Q6C):dp/dt	3.582	0.638	-0.0560	0.6028	0.3006	-0.0518	6
RT(p51K173C):dp/dt	3.585	0.638	-0.0564	0.6042	0.3012	-0.0521	6
RT(p51E194C):dp/dt	3.589	0.653	-0.0518	0.5695	0.3064	-0.0509	12
RT(p51K281C):dp/dt	3.754	0.682	-0.0545	0.5791	0.2932	-0.0473	6

<sup>[a]</sup> The quantum yields  $\Phi_{FD(0)}$  were calculated by  $\Phi_{FD(0)} = \Phi_{FD(0)}^{\text{ref}} / \tau_0^{\text{ref}} \left( \sum_i x_i \tau_i \right)$  where  $x_i$  and  $\tau_i$  are species fractions and lifetimes. As reference the known lifetimes and quantum yields of Alexa488 have been used ( $\tau^{\text{ref}}=4.0$  ns  $\Phi_{FD(0)}^{\text{ref}} = 0.8$ ).

#### 4.4. Statistical uncertainties in PDA

Confidence intervals estimation for multiple fit parameters is performed as follows. All free fit parameters are varied simultaneously in a random manner. The  $\chi_r^2$ -value is calculated at 100000 random points yielding 100-1000 points with  $\chi_r^2$ -values below  $\chi_{r,\text{min}}^2 + (2/N_{\text{bins}})^{1/2}$  (here  $N_{\text{bins}}$  is the number of histogram bins, and  $\chi_{r,\text{min}}^2$  is the reduced chi-squared of the best fit). The range where such fits are possible is assigned as  $1\sigma$  confidence interval (Figure N4.1). Whereas one could calculate  $\chi_r^2$  thresholds more strictly from the chi-squared distribution <sup>16</sup>, in practice  $\chi_{r,\text{min}}^2$  is often affected by experimental imperfections and can be considerably larger than one. For this reason, we prefer the simple test mentioned above which relates  $\chi_r^2$  values to that of the best fit.

## Supplement A

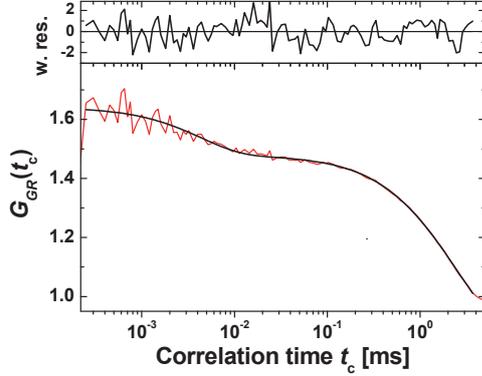


**Figure N4.1.** Confidence interval estimation in PDA. Here  $\langle R_{DA} \rangle_E$  of the P-E state is plotted against its apparent width ( $\sigma_{app}$ ) for RT(p66Q6C):dp(10)/dt. Black points represent sets of parameters which provide acceptable fits with  $\chi_r^2 < \chi_{r,min}^2 + (2/N_{bins})^{1/2}$ , where  $\chi_{r,min}^2$  is the reduced chi-squared for the best fit, and  $N_{bins}$  is the number of FRET histogram bins. Gray dots correspond to  $\chi_r^2 < \chi_{r,min}^2 + 1$ , making it clear that pre-defined search intervals are sufficiently large. Red lines show resulting  $1\sigma$  confidence interval for  $\langle R_{DA} \rangle_E$ .

It is clear from Figure N4.1 that in our case the statistical uncertainties ( $< \pm 0.2\text{\AA}$ ) are very small compared to those due to  $\kappa^2$  (see further). However, photon statistics becomes more relevant when minor and/or overlapping FRET states are considered. In our experience, for minor FRET states with  $< 20\%$  equilibrium population, statistical errors can easily exceed those due to  $\kappa^2$  uncertainties (unpublished results). Moreover, if an assignment of the P-P and P-E state is unclear the errors were calculated as described below in Supplementary Note 15.

## Supplement A

### 4.5. FRET populations are static on the ms timescale: sub-ensemble FCS



**Figure N4.2:** Sub-ensemble FCS curve (red line, green-red cross-correlation) obtained for the sample RT(p51K173C):dp(10)/dt and fitted with equation (4.7) accounting for one bunching term additional to the diffusion term. See Table N4.2 for all fit parameters.

To check whether FRET states of interest interconvert on the experimental time scale, we studied the FRET labeled sample RT(p51K173C):dp(10)/dt by sub-ensemble FCS. In Figure N4.2 the green-red correlation curve is presented. Besides the diffusion term the FCS curve exhibits a bunching term at  $\sim 4 \mu\text{s}$ , probably due to triplet relaxation (see Table N4.2) but clearly no anti-bunching term is present. This is a clear indication that there are no FRET states interconverting within the time range of milliseconds.

$$G_{GR}(t_c) = A_0 + \frac{1}{N} \cdot \frac{1}{1 + \frac{t_c}{t_d}} \cdot \frac{1}{\sqrt{\left(1 + \frac{t_c}{(z_0/w_0)^2 \cdot t_d}\right)}} \cdot \left(1 - A_1 + A_1 \cdot e^{-\frac{t_c}{t_1}}\right) \quad (4.7)$$

**Table N4.2:** Fit parameters and their assigned meanings for the correlation curve in Figure N4.2.

Parameter	Meaning	Value
$A_0$	Offset	0.67
$N/(1-A_1)$	Number of all molecules	0.86
$t_d$ [ms]	Diffusion time	2.67
$z_0/w_0$	Axial ratio	1.24
$A_1$	amplitude bunching term	0.17
$t_1$ [ $\mu\text{s}$ ]	timescale bunching term	4.35

## Supplement A

### Supplementary Note 5: Sub-ensemble TCSPC (seTCSPC) of FRET data

Due to local quenching the fluorescence decay of the donor in the absence of FRET is often multi-exponential (equation (5.1))

$$F_{D(0)}(t) = \sum_i x_D^{(i)} \exp(-t / \tau_{D(0)}^{(i)}) \quad (5.1)$$

Furthermore, it is usually assumed that for given DA distance and orientation, FRET rate ( $k_{\text{FRET}}$ ) is independent of the donor lifetime. This is strictly true if quenching does not change the donor radiative lifetime, which is generally reasonable. In the presence of FRET, the donor fluorescence decay can be then expressed as:

$$F_D(t) = \sum_i x_D^{(i)} \int_{R_{DA}} p(R_{DA}) \exp\left(-\frac{t}{\tau_{D(0)}^{(i)}} \left[1 + (R_0 / R_{DA})^6\right]\right) dR_{DA} \quad (5.2a)$$

In this work we assumed a Gaussian distribution of the donor-acceptor distances ( $p(R_{DA})$ ) with a mean of  $\langle R_{DA} \rangle$  and a half-width of  $\sigma_{DA}$ :

$$F_D(t) = \sum_i x_D^{(i)} \int_{R_{DA}} \frac{1}{\sqrt{2\pi}\sigma_{DA}} \exp\left(-\frac{(R_{DA} - \langle R_{DA} \rangle)^2}{2\sigma_{DA}^2}\right) \exp\left(-\frac{t}{\tau_{D(0)}^{(i)}} \left[1 + (R_0 / R_{DA})^6\right]\right) dR_{DA} \quad (5.2b)$$

In addition the contributions of a fraction of Donor-only molecules ( $x_{D\text{Only}}$ ) and scattered light was considered (equation (5.3a/5.3b)):

$$F(t) = (1 - x_{D\text{Only}})F_D(t) + x_{D\text{Only}}F_{D(0)}(t) \quad (5.3a)$$

Equation (5.3a) in combination with equation (6.1) was used to fit the polarized components of the donor fluorescence relaxation recovered from smMFD data.

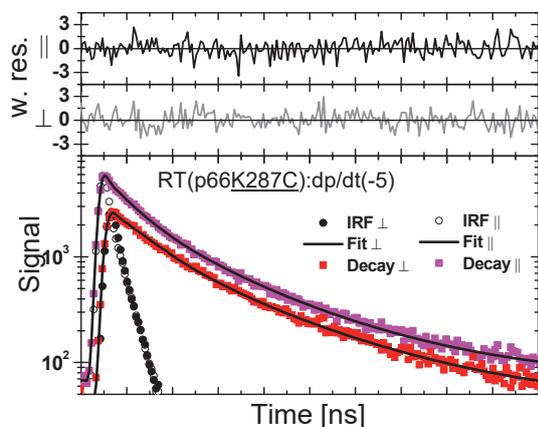
The fraction of scattered light is considered individually as described in equation (5.3b). Whereas  $s_{\parallel/\perp}$  is the fraction of scattered light in the vertical or horizontal channel:

$$F_{\parallel/\perp}^s(t) = (1 - s_{\parallel/\perp})F_{\parallel/\perp}(t) + s_{\parallel/\perp}IRF(t) \quad (5.3b)$$

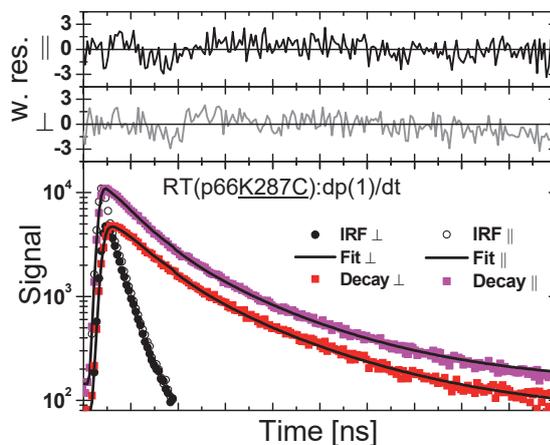
As shown in Figure N5.1 and Supplementary Tables 4 and 5 there are no differences in width between the samples with an acceptor located on the double stranded part of the DNA (dp(1)/dt, dp(10)/dt, dp(19)dt) and the samples with the acceptor fluorophore attached to the template overhang (dp/dt(-6), dp/dt(-15)).

## Supplement A

(A)



(B)



**Figure N5.1:** Decay histograms and fits of a Gaussian model function distributed in  $R_{DA}$ . The fluorescence properties (lifetimes) and anisotropies of the donor were fixed to the values of the donor sample p66K287C:dp/dt, the background fluorescence was fixed to values determined by independent background measurements. Free fitting parameters are the mean DA distance  $\langle R_{DA} \rangle$  and the width of the distance distribution  $\sigma_{DA}$ . The obtained parameters are: (A)  $\langle R_{DA} \rangle = 43 \text{ \AA}$ ,  $\sigma_{DA} = 13 \text{ \AA}$ ,  $\chi^2_r = 1.0$ , for RT(p66K287C):dp/dt(-5), and (B),  $\langle R_{DA} \rangle = 34 \text{ \AA}$ ,  $\sigma_{DA} = 15 \text{ \AA}$ ,  $\chi^2_r = 1.4$  for RT(p66K287C):dp(10)/dt.

## Supplement A

### Supplementary Note 6: Estimation of the $\kappa^2$ related distance uncertainties

#### 6.1. Distributions of possible $\kappa^2$ values derived from anisotropy data

#### 6.2. Estimation of $\langle \kappa^2 \rangle$ and $\kappa^2$ -related errors

#### 6.1. Distributions of possible $\kappa^2$ values derived from anisotropy data

The procedure is described in details in <sup>3,17,18</sup>. Briefly, the knowledge of residual anisotropies ( $r_\infty$ ) of the directly excited D and A, and that of the A excited via FRET ( $r_{\infty,D}$ ,  $r_{\infty,A}$  and  $r_{\infty,A(D)}$ , respectively), allows one to reduce the range of possible  $\kappa^2$ -values. In our case a sub-ensemble analysis of single-molecule data was performed to estimate  $r_{\infty,D}$ . Figure N6.1 shows typical decays of polarized components  $F_{\parallel}(t)$  and  $F_{\perp}(t)$  which are globally fitted with the following model:

$$\begin{aligned} F_{\parallel}(t) &= F(t) \cdot (1 + (2 - 3l_1) \cdot r(t)) / 3 + B_{\parallel} \\ F_{\perp}(t) &= GF(t) \cdot (1 - (1 - 3l_2) \cdot r(t)) / 3 + B_{\perp} \end{aligned} \quad (6.1)$$

In equation (6.1)  $F(t)$  is the fluorescence decay typically modeled by a bi-exponential relaxation,  $G$  is the ratio of detection efficiencies of parallel and perpendicular channels, factors  $l_1$  and  $l_2$  describe polarization mixing in high-NA objectives <sup>19</sup>, and  $B_{\parallel}$  and  $B_{\perp}$  represent background contributions in parallel and perpendicular detection channels, respectively. The anisotropy decay  $r(t)$  reflects local motions of the dye and global rotations of the macromolecule according to equation (6.2a)

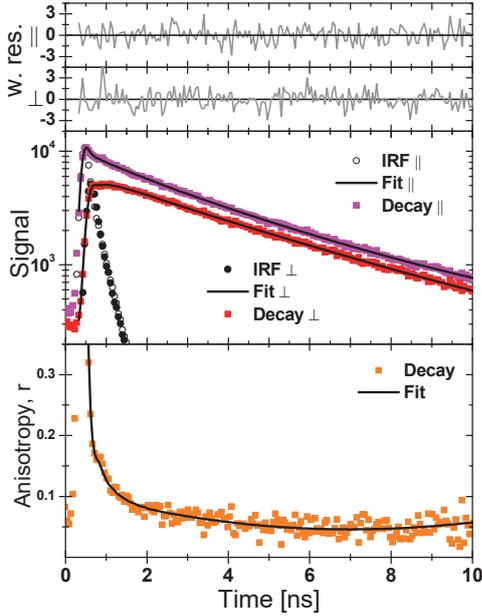
$$r(t) = \left[ (r_0 - r_\infty) e^{-\frac{t}{\rho_{local}}} + r_\infty \right] e^{-\frac{t}{\rho_{global}}} \quad (6.2a)$$

In equation (6.2a),  $r_0$  is the fundamental anisotropy (fixed to  $r_0 = 0.38$ ) and  $r_\infty$  is the residual anisotropy. If  $\rho_{local} \ll \rho_{global}$  equation (6.2a) becomes:

$$r(t) = (r_0 - r_\infty) \exp(-t / \rho_{local}) + r_\infty \exp(-t / \rho_{global}) \quad (6.2b)$$

A fitted anisotropy decay is presented in Figure N6.1. For  $r_{\infty,A}$  we use 0.13 from <sup>3</sup> and for  $r_{\infty,A(D)}$  we used steady state anisotropies  $r_{s,A(D)}$  which were determined via PDA according to <sup>12</sup>. Using the formalism described in <sup>17</sup>, we obtain  $\kappa^2$  uncertainties ( $\Delta R_{DA}(\kappa^2)$ ) corresponding to each FRET distance. A typical  $\kappa^2$ -distribution is shown in Fig. 3b (main text). All values determined for  $\Delta R_{DA}(\kappa^2)$  are listed in Supplementary Tables 2 and 3.

## Supplement A



**Figure N6.1.** (Middle)  $F_{\parallel}(t)$  and  $F_{\perp}(t)$  decays obtained for the D-only population of sample RT(p51E194C):dp/dt. The instrument response functions (IRF) for parallel and perpendicular detection channels are shown as open and filled circles, respectively. The fits to  $F_{\parallel}(t)$  and  $F_{\perp}(t)$  (equation (6.1)) are displayed as black solid lines. Weighed residuals are shown in the upper panel. (Bottom) The experimental anisotropy decay calculated as  $r(t) = [F_{\parallel}(t) - GF_{\perp}(t)]/[F_{\parallel}(t) + 2GF_{\perp}(t)]$  overlaid with that obtained from the fitted model functions (equations (6.1) and (6.2)). The resulting  $r_{\infty}$  is 0.131 (the apparent anisotropy plateau is lower because of background contributions).

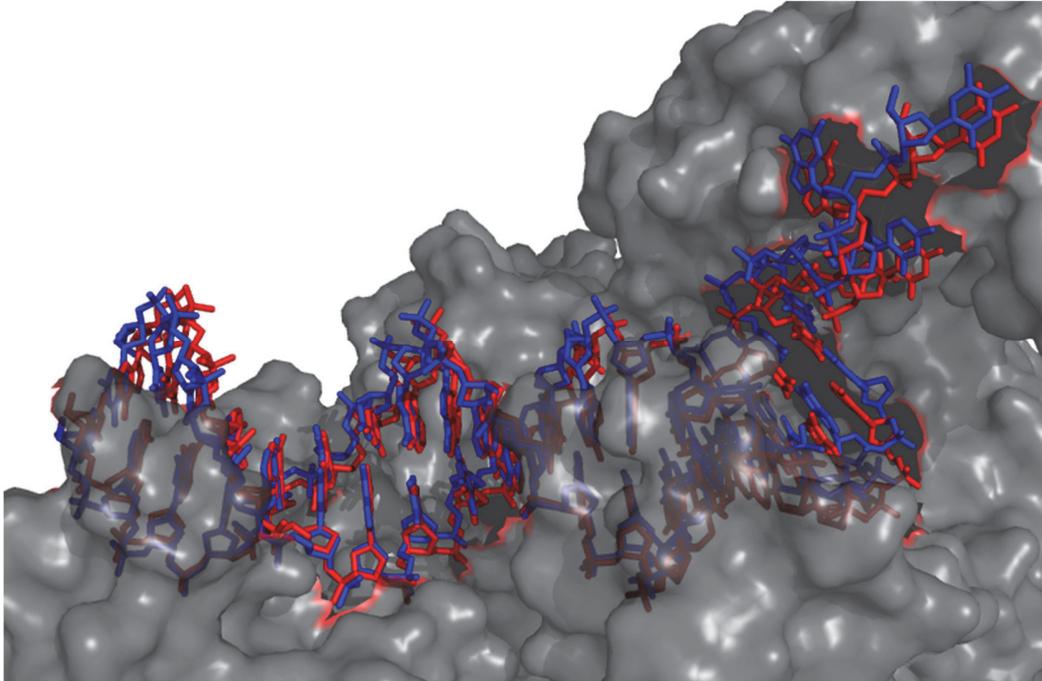
### 6.2. Estimation of $\langle \kappa^2 \rangle$ and $\kappa^2$ -related errors

The Förster radius is typically calculated for  $\langle \kappa^2 \rangle = 2/3$  yielding FRET-averaged distances  $\langle R_{DA} \rangle_E(\kappa^2=2/3)$ . Alternatively, the distributions of possible  $\kappa^2$  values can be accounted for (Fig. 3b, main text). In this case we consider the mean possible  $\langle R_{DA} \rangle_E = \langle (\kappa^2)^{-1/6} \rangle \langle R_{DA} \rangle_E(\kappa^2=2/3)$ . For all cases considered here  $\langle \kappa^2 \rangle < 2/3$  and thus the  $\kappa^2$ -corrected  $\langle R_{DA} \rangle_E$  values are larger than  $\langle R_{DA} \rangle_E(\kappa^2=2/3)$  by 3 to 7 %. The distances that correspond to the maximum of  $(\kappa^2)^{-1/6}$  distributions are similar being larger than the respective  $\langle R_{DA} \rangle_E(\kappa^2=2/3)$  values by 5 to 9 %. It is also important to note that the procedure to obtain distributions of possible  $\kappa^2$  values is based on approximations of Dale et al <sup>17</sup>, i.e. dynamic orientation averaging is assumed and  $r_{\infty}$  is attributed to restricted rotational diffusion. In reality  $r_{\infty}$  could be due to temporarily immobile dye populations, and the assumptions of Dale et al are not necessarily fulfilled. In this case  $\langle \kappa^2 \rangle = 2/3$  is not necessarily less realistic than  $\langle \kappa^2 \rangle$  obtained using the approach of Dale et al.

In addition,  $\kappa^2$ -related uncertainties can be estimated in various ways, for example, by taking  $\pm \text{std}((\kappa^2)^{-1/6})$  as  $1\sigma$  confidence interval, or by disregarding a certain fraction of extreme  $\kappa^2$ -values. Because the exact details of  $\langle \kappa^2 \rangle$  estimation have an effect on the distance restraints and thus on the structural models, we compared docking results obtained for two sets of restraints:  $\langle R_{DA} \rangle_E(\kappa^2=2/3)$  and  $\langle (\kappa^2)^{-1/6} \rangle \langle R_{DA} \rangle_E(\kappa^2=2/3)$ . In the first case, the  $\kappa^2$ -errors were estimated as  $\pm \text{std}((\kappa^2)^{-1/6})$ , and in the second case we disregarded 16.5 % of the lowest and 16.5 % of the highest  $\kappa^2$  values. The final structural models are compared in

## Supplement A

Figure N6.2. The RMSD between these structures is 1.8 Å, which is within the claimed precision of our method (2.9 Å for the best model).



**Figure N6.2.** Overlay of the best FRET model presented in the main text (red, assuming  $\langle \kappa^2 \rangle = 2/3$ ) with the model obtained for  $\kappa^2$ -corrected distance restraints (blue).

## Supplement A

### Supplementary Note 7: Obtaining mean position distances: additional details

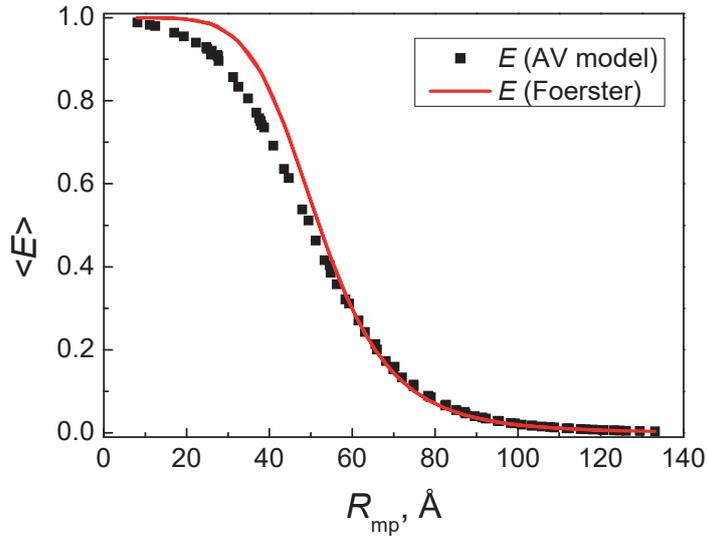
As in this work we focus on FRET between dyes attached to a macromolecule via flexible linkers, we must consider the contribution of linker length to the DA distance. In most of cases it is not negligible and can be even comparable with absolute donor-acceptor distances (Fig. 1 main text). As discussed in detail in <sup>3</sup>, dye linkers are responsible for two major effects: (1) obviously, a considerable displacement of the mean dye position with respect to the attachment point is expected, and (2) averaging of FRET efficiencies over possible DA distances and orientations takes place. The second consideration makes the well-known equation  $E = 1/(1 + R_{DA}^6/R_0^6)$  inapplicable even to calculate the distance between mean dye positions ( $R_{mp}$ ). To illustrate this effect and hence the difference between  $R_{mp}$  and  $\langle R_{DA} \rangle_E$  AVs of Alexa488 and Cy5 dyes attached to DNA using "standard" C6 linkers were simulated. Fig. 1 (main text) schematically shows simulated dye clouds, making it also clear that the C6 linker length ( $\sim 20$  Å between the attachment point and the center of the dye) is comparable with DA distances accessible to FRET ( $\sim 35-70$  Å for the Alexa488-Cy5 pair). By varying the labeling position of the acceptor dye, we obtained a set of mean position distances and corresponding mean FRET efficiencies  $\langle E \rangle$ . In Figure N7.1, the values of  $\langle E \rangle$  are plotted against  $R_{mp}$  values. Significant deviations between the Förster dependence ( $E = 1/(1 + R_{DA}^6/R_0^6)$ ) and the obtained relationship between  $\langle E \rangle$  and  $R_{mp}$  are apparent. Thus, if the average FRET efficiency  $\langle E \rangle$  is directly used to calculate  $R_{mp}$ , errors of up to 10 Å ( $\sim 30\%$ ) are expected (Fig. 1 main text). The deviations are most pronounced for shorter distances, where the size of the dyes' accessible volumes becomes comparable to  $R_{mp}$ .

By interpolation of an  $\langle R_{DA} \rangle_E$ - $R_{mp}$  dependence using a known set of dye clouds, a  $\langle R_{DA} \rangle_E$ - $R_{mp}$  conversion function <sup>3,10</sup> can be generated. In our particular case the conversion function is given by:

$$R_{mp} = 1.109 \times 10^{-5} \langle R_{DA} \rangle_E^3 - 7.286 \times 10^{-3} \langle R_{DA} \rangle_E^2 + 1.979 \langle R_{DA} \rangle_E - 34.345 \text{ [Å]}$$

We should note that  $\langle E \rangle$  depends not only on  $R_{mp}$ , but also on the mutual orientation of the dye clouds (of given size and shape). Fortunately, this effect is relatively weak even for asymmetric dye position distributions such as presented in Fig. 1. Fig. 1 (main text) also shows the conversion function generated for random orientation of the dyes' AVs. The average deviation between the data and the fitted 3rd order polynomial function is only 0.6 Å ( $30 \text{ Å} < \langle R_{DA} \rangle_E < 70 \text{ Å}$ ), which is much smaller than typical uncertainties of FRET-derived distances. Obviously, for more spherical AVs this deviation would be even less pronounced. This justifies the use of a conversion function at least during the initial optimization steps. With some effort the  $\langle E \rangle$ - $R_{mp}$  conversion function can be also obtained empirically <sup>20</sup> by studying a set of molecules with known structure (e.g. dsDNA).

## Supplement A



**Figure N7.1.** The relationship between  $R_{mp}$  and  $\langle E \rangle$  averaged over spatial distributions of donor and acceptor estimated using the AV approach as shown in Fig. 1 in the main text (for simplicity,  $\kappa^2 = 2/3$  is assumed for every DA pair). The red solid line represents the dependence given by  $E = 1/(1 + R_{DA}^6/R_0^6)$  with  $R_{DA} = R_{mp}$ .

## Supplement A

### Supplementary Note 8: Rigid body docking: implementation details

#### 8.1. Mechanical model

#### 8.2. Time evolution of the rigid body system

#### 8.1. Mechanical model

As discussed in the main text, the purpose of rigid body docking is to minimize the weighed deviation between  $n$  experimentally obtained distances  $\{R_{DA}\}$  and corresponding model distances  $\{R_{\text{model}}\}$ , given the uncertainties  $\{\Delta R_{DA}\}$ :

$$\chi_E^2 = \sum_n \frac{(R_{DA} - R_{\text{model}})^2}{(\Delta R_{DA})^2} \quad (8.1)$$

In equation (8.1),  $\{R_{DA}\}$  can represent a set of  $R_{\text{mp}}$  or  $\langle R_{DA} \rangle_E$  values. The first option is easier to implement in combination with rigid body dynamics, whereas directly calculating deviations between experimental and model  $\langle R_{DA} \rangle_E$  is more appropriate for structure screening. Now we notice that the right-hand part of equation (8.1) is equivalent to the energy of a network of mean dye positions (points in space), connected with  $n$  springs with relaxed lengths of  $\{R_{DA}\}$  and corresponding spring constants  $k = 2/(R_{DA})^2$ . The coordinates of mean dye positions are obtained by MD or AV simulations and then fixed with respect to corresponding labeled macromolecules, which are treated as rigid bodies. Minimizing  $\chi_E^2$  in equation (8.1) is equivalent to relaxing this rigid body system. Here and further in this section we use reduced energy which results in unusual units for  $k$  [ $1/\text{\AA}^2$ ] and other familiar quantities. If not stated otherwise, all distances are expressed in  $\text{\AA}$ , and masses are expressed in Da.

In addition to FRET restraints, clashes between different subunits are prevented by considering clash contributions to the total “energy”, equivalently to equation (8.1), by equation (8.2):

$$\chi_{\text{clash}}^2 = \sum_{i,j} \begin{cases} 0 & , r_{ij} \geq r_{wi} + r_{wj} \\ (r_{wi} + r_{wj} - r_{ij})^2 / r_{\text{ctol}}^2 & , r_{ij} < r_{wi} + r_{wj} \end{cases} \quad (8.2)$$

where  $r_{ij}$  is the distance between atoms  $i$  and  $j$  which belong to different subunits,  $r_{wi}$  and  $r_{wj}$  are their van der Waals radii, and  $r_{\text{ctol}}$  is the pre-defined clash tolerance. We typically used 6  $\text{\AA}$  during initial search and 2 to 1  $\text{\AA}$  during refinement (see “step 4” in the main text). More realistic potentials (e.g. the Lennard-Jones potential<sup>21</sup>) can be also used here instead of the harmonic potential in equation (8.2). The simplified approach (equation (8.2)) is justified by low accuracy of FRET data (a few  $\text{\AA}$ ) as compared to possible violations of van der Waals radii ( $r_{wi} + r_{wj} - r_{ij} < 0.1 \text{\AA}$ ). Anyhow, by choosing a sufficiently small  $r_{\text{ctol}}$  the atoms can be made as “hard” as desired, in which case the contribution of  $\chi_{\text{clash}}^2$  becomes negligible. The reduced chi-squared parameter to be minimized is then given by equation (8.3)

## Supplement A

$$\chi_r^2 = (\chi_E^2 + \chi_{\text{clash}}^2)/(n - p) \rightarrow \min \quad (8.3)$$

where  $n$  is the number of distance restraints (in our case  $n = 20$ ) and  $p$  is the number of degrees of freedom, which is equal to  $6 \times (\text{number of bodies} - 1)$ , in our case  $p = 6$ .

### 8.2. Time evolution of the rigid body system

The position and orientation of each subunit at any time  $t$  is described by a coordinate vector of its center of mass  $\mathbf{x}(t)$  and a rotational matrix  $\mathbf{Q}(t)$ . To model translational and rotational motions we used the Verlet algorithm<sup>22</sup> with damping,

$$\mathbf{x}(t + \Delta t) = (2 - \nu)\mathbf{x}(t) - (1 - \nu)\mathbf{x}(t - \Delta t) + \mathbf{F}\Delta t^2/m \quad (8.4)$$

In equation (8.4),  $\Delta t$  is the simulation time step, the factor  $\nu$  accounts for viscosity (see equation (8.9)),  $\mathbf{F}$  is the total force acting on a subunit, and  $m$  stands for its mass. The ‘‘forces’’ are derived from violations of FRET distances, clashes between subunits, and optionally other constraints (e.g. flexible chemical linkages between subunits). Since here we are only interested in finding energy minima (rather than in investigating trajectories), we use a simplified representation of the moment of inertia ( $I$ ) by assuming the same (mean) value of  $I$  for all axes. In analogy with equation (8.4), we obtain for the rotations

$$\mathbf{Q}(t + \Delta t) = \mathbf{Q}\left(\frac{\mathbf{T}}{|\mathbf{T}|}, \theta\right) \mathbf{Q}(\mathbf{w}, -\nu\theta_w) \mathbf{Q}(t) \mathbf{Q}^{-1}(t - \Delta t) \mathbf{Q}(t) \quad (8.5)$$

In equation (8.5),  $\mathbf{T}$  is the total torque vector; the angle  $\theta$  is given by

$$\theta = \Delta t^2 |\mathbf{T}|/I \quad (8.6)$$

the rotational matrices with parameters represent a rotation by angle  $\theta$  about an axis  $\mathbf{u}$ ,

$$\mathbf{Q}(\mathbf{u}, \theta) = \begin{pmatrix} c + u_x^2 d & u_x u_y d - u_z s & u_x u_z d + u_y s \\ u_x u_y d + u_z s & c + u_y^2 d & u_y u_z d - u_x s \\ u_x u_z d - u_y s & u_y u_z d + u_x s & c + u_z^2 d \end{pmatrix} \quad (8.7)$$

with:  $\mathbf{u} = (u_x, u_y, u_z)$ ;  $c = \cos\theta$ ;  $d = 1 - \cos\theta$ ;  $s = \sin\theta$

and  $\mathbf{w}$  and  $\theta_w$  fulfil equation (8.8)

$$\mathbf{Q}(\mathbf{w}, \theta_w) = \mathbf{Q}(t) \mathbf{Q}^{-1}(t - \Delta t) \quad (8.8)$$

To minimize oscillations and to improve convergence, viscosity factor  $\nu$  must be chosen so that the system is close to being critically damped.  $\nu$  is initially estimated by equation (8.9)

$$\nu = 2\Delta t \sqrt{K/M} \quad (8.9)$$

with an option of additional fine-tuning to improve convergence. In equation (8.9)  $K$  is the sum of all spring constants and  $M$  is the total mass of all subunits.

## Supplement A

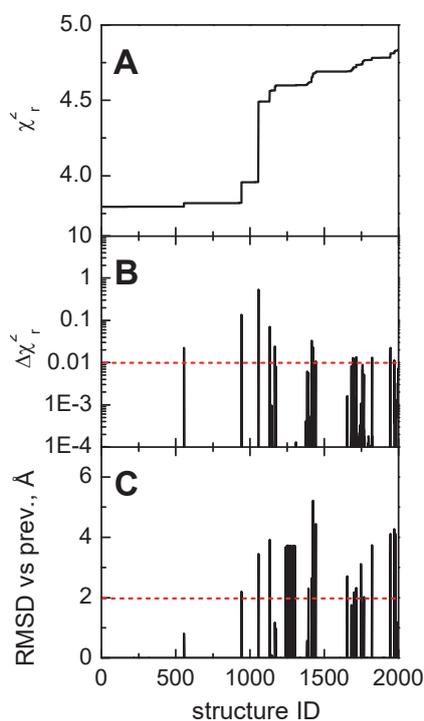
Each rigid body simulation is run until the following criteria are met: kinetic energy  $E_K < 0.001$ ;  $|\mathbf{F}| < 0.001$ ;  $|\mathbf{T}| < 0.02$  (in the units described above). After the initial search procedure, typically more than 99% of structures reach convergence as defined by these criteria. We should also mention that more advanced optimization procedures<sup>23</sup> can be also applied at this step. However, in our experience, even with rigid body dynamics >95% convergence probability is achieved in most of cases.

## Supplement A

### Supplementary Note 9: Cluster analysis on dsDNA

After the *search* phase, the obtained solutions are sorted by “energy” (or equivalently by  $\chi_r^2$ ) and plotted as shown in Figure N9.1A. Obviously, there are groups of very similar solutions (clusters), which are separated by steps in the  $\chi_r^2$  plot and corresponding peaks in the RMSD plot (Figure N9.1).

To generate a cluster plot as shown in Figure 4 (main text), we applied thresholds to  $\Delta\chi_r^2$  and RMSD plots as shown by red lines in Figure N9.1. Solutions for which both  $\Delta\chi_r^2$  and RMSD fell below these thresholds were grouped with previous structures. As a result, several clusters of similar structures were obtained as shown in Figure 4 using different symbol sizes.



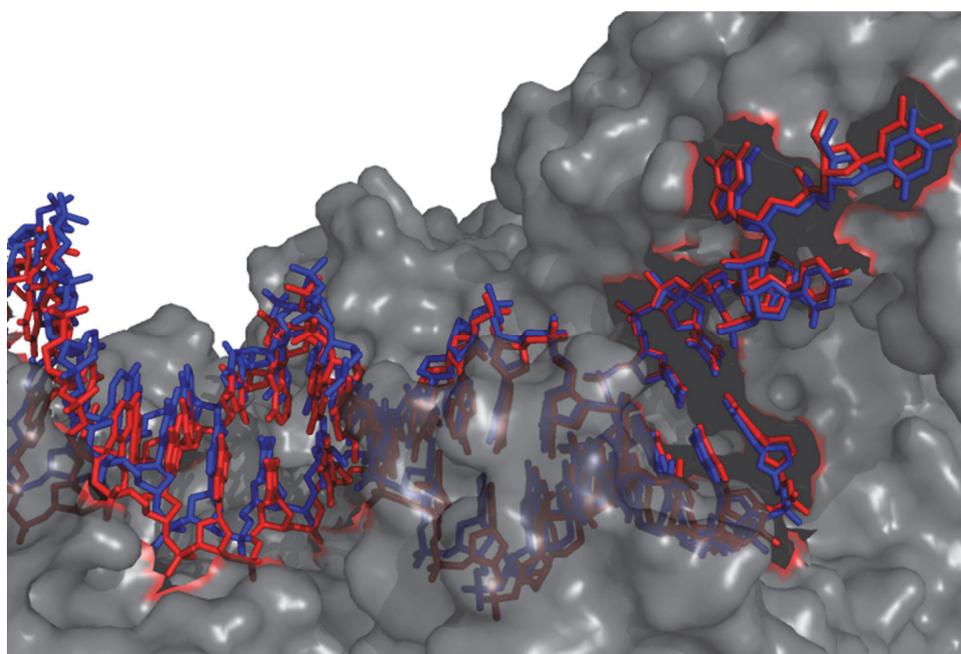
**Figure N9.1.** (A)  $\chi_r^2$  for 2000 best solutions. (B) Increase of  $\chi_r^2$  compared to the previous structure. (C) RMSD between current and previous structures. Best 2000 of 10000 structures generated during the search phase with  $r_{\text{ctol}} = 6 \text{ \AA}$  are shown. Red lines indicate clustering thresholds. Resulting clusters are shown in Figure 4 (main text).

## Supplement A

### Supplementary Note 10: Influence of the clash tolerance

The estimated error in the coordinates of an X-ray structure is inherently related to the resolution, and as a rule of thumb this error amounts to resolution divided by  $6^{24-26}$ . Considering error propagation rules, this then leads to possible clash threshold of  $\sim 1.0$  Å for a structure with a resolution of 2.9 Å.

Further decreasing the clash tolerance does not improve the obtained 3D models. Figure N10.1 compares the best FRET model presented in the main text (1 Å clash tolerance) with the model obtained with a clash tolerance of 0.05 Å. The RMSD between these models is 1.6 Å, and the RMSD between the solution with 0.05 Å clash tolerance and the X-ray structure is 2.1 Å.



**Figure N10.1.** Overlay of the best FRET model presented in the main text (1 Å clash tolerance; red) with the model obtained with a clash tolerance of 0.05 Å (blue).

## Supplement A

### Supplementary Note 11: FRET-guided positioning and screening for ds B-DNA

#### 11.1. Initial docking of B-DNA to HIV-RT

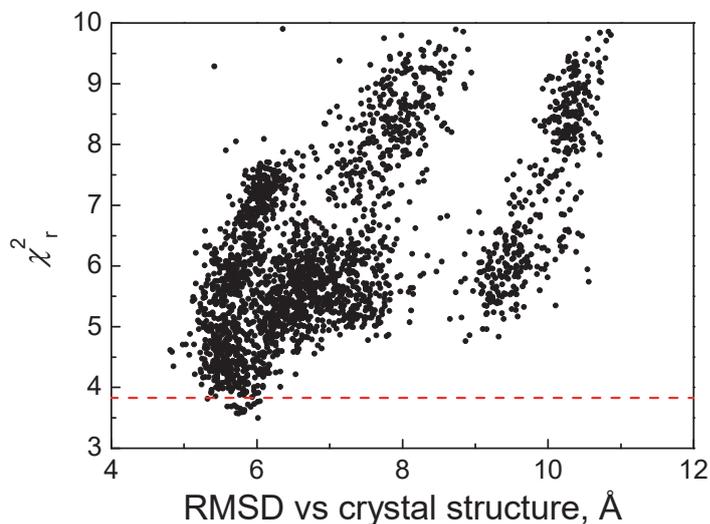
#### 11.2. FRET-guided screening

#### 11.1. Initial docking of B-DNA to HIV-RT

To generate an initial model that is later on relaxed and screened by MD and FPS, respectively, we generated a dsDNA molecule with the same primary and secondary structure as the double-stranded part of dp/dt (see Supplementary Figure 2, positions 1 to 19) and with a canonical B-DNA form using the nucleic acid builder (NAB) which is a part of AmberTools<sup>27</sup>. In the same way as described in “Step 4” in the main text the B-DNA molecule is docked to RT using the same labeling positions and distance restraints. The docked structure with the best agreement to FRET data is shown in Fig. 4c in the main text. This structure is then further relaxed using MD simulations (see Supplementary Note 12).

#### 11.2. FRET-guided screening

The structures generated by MD simulations were screened with respect to agreement with FRET data as described for the overhang part. In Figure N11.1 one can see a correlation between the structural deviation of MD structures from the crystal structure and  $\chi_r^2$ .



**Figure N11.1.** Screening of MD data.  $\chi_r^2$  is calculated using 20 FRET distances measured for the dsDNA part of the complex and is plotted against RMSD with respect to the crystal structure. The dashed red line represents a threshold  $\chi_r^2 < \chi_{r,\min}^2 + 0.378$ .

## Supplement A

### Supplementary Note 12: Structure ensemble generation via MD simulations

#### 12.1. General settings

#### 12.2. MD simulations of the B-DNA:RT complex

#### 12.3. MD simulations of the template 5' overhang in the primer/template DNA:RT complex

#### 12.1. General settings

MD simulations were performed with the AMBER 11 suite of programs<sup>28</sup> together with the force field as described by Cornell et al.<sup>29</sup> using modifications suggested by Simmerling et al.<sup>30</sup> for the protein, and the “bsc0” parameterization for the DNA<sup>31</sup>. The complex structure was placed into an octahedral periodic box of TIP3P water molecules<sup>32</sup> with Na<sup>+</sup> ions added to reach electroneutrality of the system. The distance between the edges of the water box and the closest atom of the complex was at least 10 Å, resulting in a system of ~147000 atoms. The system was minimized by 50 steps of steepest descent minimization followed by 450 steps of conjugate gradient minimization. The particle mesh Ewald (PME) method<sup>33</sup> was used to treat long-range electrostatic interactions, and bond lengths involving bonds to hydrogen atoms were constrained using SHAKE<sup>34</sup>. The time-step for all MD simulations was 2 fs, with a direct-space, non-bonded cutoff of 8 Å. Applying harmonic restraints with force constants of 5 kcal mol<sup>-1</sup> Å<sup>-2</sup> to all solute atoms but the added nucleotides, canonical ensemble (NVT)-MD was carried out for 50 ps, during which the system was heated from 100 K to 300 K. Subsequent isothermal isobaric ensemble (NPT)-MD was used for 50 ps to adjust the solvent density. From there, with the harmonic restraints applied to all solute atoms but the first 15 nucleotides of the template overhang, ten different trajectories were spawned by adjusting the simulation temperature to 300.0 K, 300.1K, ..., 300.9 K, using a time constant of 10 ps for heat-bath coupling. After additional 20 ns of equilibration time for each trajectory, the following 50 to 65 ns of NVT-MD for each trajectory were used for analysis with the program ptraj of the AMBER suite, with conformations saved every 20 ps. This resulted in 571 ns of total simulation time for production.

#### 12.2. MD simulations of the B-DNA:RT complex

In order to relax the initial structure of a canonical B-DNA docked to HIV-1-RT from PDB code 1R0A (see Supplementary Note 2) MD simulations were performed with general settings as described above. After performing the NPT-MD for adjusting the solvent density, harmonic restraints with force constants of 5 kcal mol<sup>-1</sup> Å<sup>-2</sup> were applied to protein atoms only whereas no restraints were applied to the dsDNA anymore. Ten different trajectories were spawned by adjusting the simulation temperature to 300.0 K, 300.1K, ..., 300.9 K, using a time constant of 10 ps for heat-bath coupling. After additional 20 ns of equilibration time for each trajectory, the following 30 ns of NVT-MD for each trajectory were used for analysis with conformations saved every 20 ps. This resulted in 300 ns of total simulation time for production and use in FRET screening. The results are presented in Supplementary Note 11.

## Supplement A

### 12.3. MD simulations of the template 5' overhang in the primer/template DNA:RT complex

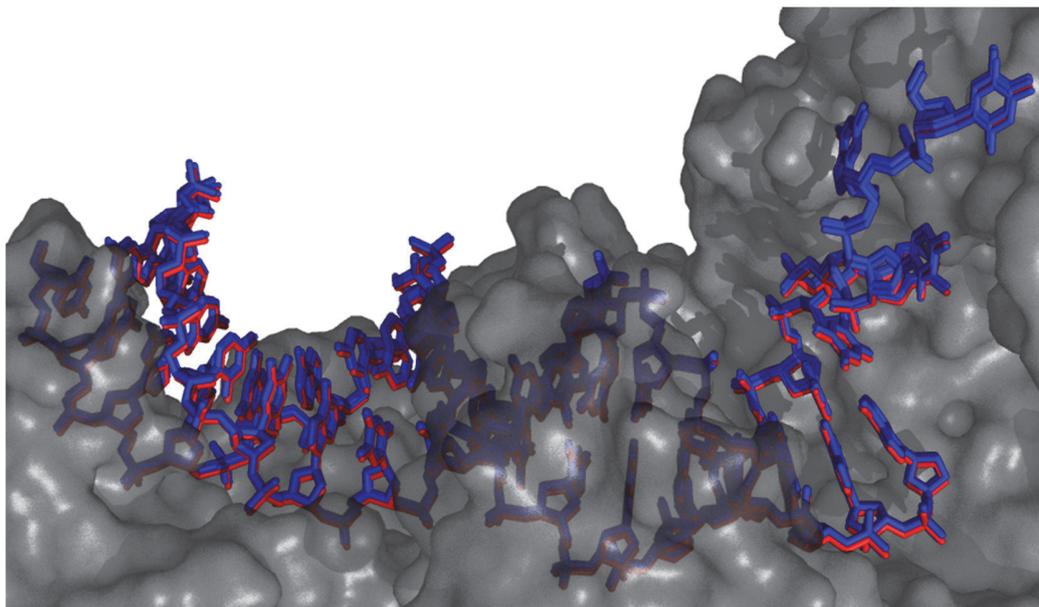
In the complex determined by X-ray crystallography (PDB code: 1R0A) four unpaired nucleotides of the template overhang are present. To generate a starting structure for the template overhang 12 nucleotides were added to the existing strand (sequence: 5'-GGGTTAATCTCT-3') such that the unpaired bases point straight away from the protein, with general settings as described above. After performing the NPT-MD for adjusting the solvent density, with harmonic restraints applied to all solute atoms but the first 15 nucleotides of the template overhang, ten different trajectories were spawned by adjusting the simulation temperature to 300.0 K, 300.1K, ..., 300.9 K, using a time constant of 10 ps for heat-bath coupling. After additional 20 ns of equilibration time for each trajectory, the following 50 to 65 ns of NVT-MD for each trajectory were used for analysis with the program "ptraj" of the AMBER suite, with conformations saved every 20 ps. This resulted in 571 ns of total simulation time for production and use in FRET screening.

For filtering against FRET data (see Fig. 6a, main text), conformations were extracted from each trajectory at intervals of 200 ps and pooled together. Likewise, the trajectories were pooled for generating a 3D histogram of the preferred location of the N1 atoms of the nucleobase of nucleotide dt(-15); the spacing of the cubic grid is 1 Å. Isopleths show preferred regions of occupancy by the N1 atoms at a contour level of 90% with respect to the most highly occupied cube. Figures were generated with PyMOL.

## Supplement A

### Supplementary Note 13: Overfitting tests (dsDNA)

For these tests we randomly discarded 20% of distances in each case and re-optimized the structure using the remaining 80% of FRET data. Figure N15.1 shows the overlay of 5 resulting structures with the best solution (red). Compared to the optimal solution, the mean RMSD of DNA phosphate atoms calculated for these 5 structures is 0.5 Å, the worst-case RMSD is 1.1 Å. This test is similar to commonly used cross-validation tests<sup>35</sup> and clearly demonstrates that our results are not critically dependent on any single distance restraint.

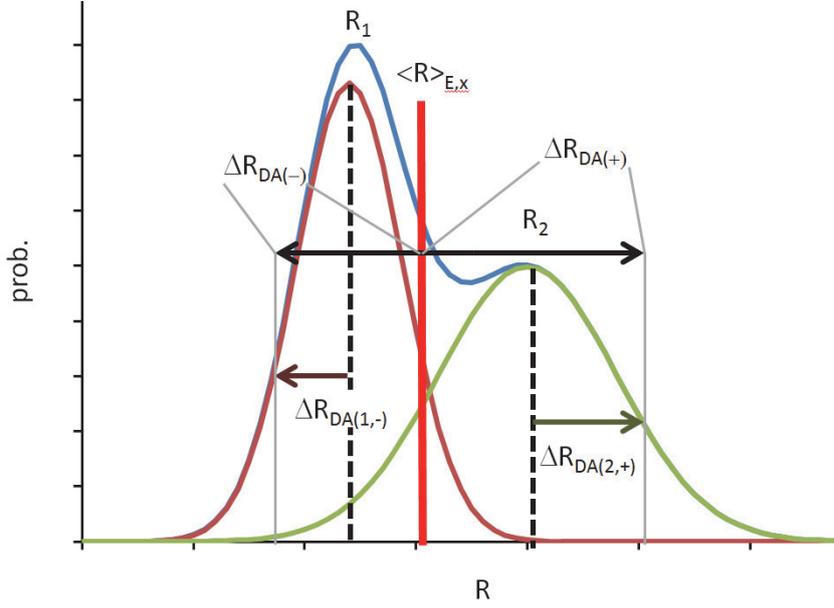


**Figure N15.1.** Overlay of 5 structures obtained using 80% of FRET restraints each (blue) and the best solution described in the main text (red).

### Supplementary Note 14: FRET-guided screening for the template overhang

The width of the states identified by seTCSPC and PDA on the single-stranded DNA is comparable to the donor-acceptor distribution on the double stranded part of the DNA (see Supplementary Tables 4 and 5). Moreover, structures that do not wind around the fingers domain are clearly inconsistent with the FRET data.

Supplementary Note 15: Peak assignment uncertainties



**Figure N15.1.** Error estimation in case of unclear peak assignment. The dark-red and green colored arrows indicate the experimental uncertainties  $\Delta R_{DA(1/2,\pm)}$  in  $R_1$  and  $R_2$ , respectively, the PDA-amplitude weighted distance of the two peaks,  $\langle R \rangle_{E,x}$ , is shown as vertical solid red line, the resulting estimated positive and negative uncertainties ( $\Delta R_{DA(\pm)}$ ) are indicated as black arrows.

In the case of a clear peak assignment to the P-P and P-E states (see Supplementary Note 3) the overall errors for the individual distances  $\Delta R_{DA}$  simply determined as follows with well-known error propagation rules:

$$\Delta R_{DA(\pm)}^2(E, \kappa^2) = \Delta R_{DA(\pm)}^2(E) + \Delta R_{DA}^2(\kappa^2) \quad (15.1)$$

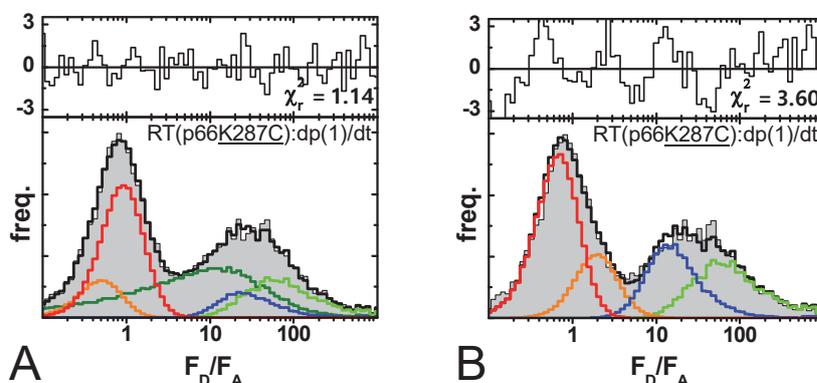
However, in some cases, namely RT(p51Q6C):dp(1)/dt, RT(p51E194C):dp(1)/dt, RT(p66T27C):dp/dt(-6), RT(p66T27C):dp/dt(-15), a clear peak assignment by amplitudes was not possible. In those cases the amplitude weighted distance  $\langle R \rangle_{E,x} = (a_1 R_{DA(1)} + a_2 R_{DA(2)}) / (a_1 + a_2)$  with the corresponding positive and uncertainties  $\text{err}_{+/-}$  was calculated as illustrated in Figure N15.1:

$$\begin{aligned} \Delta R_{DA(+)}(E, \kappa^2, \text{assign}) &= R_{DA(2)} + \Delta R_{DA(2,+)}(E, \kappa^2) - \langle R \rangle_{E,x} \\ \Delta R_{DA(-)}(E, \kappa^2, \text{assign}) &= \langle R \rangle_{E,x} - (R_{DA(1)} - \Delta R_{DA(1,-)}(E, \kappa^2)) \end{aligned} \quad (15.3)$$

## Supplementary Data

### Supplementary Data 1: smFRET data sets

**Assignment of states and color scheme in all MFD plots.** The single molecule data was fitted with PDA as explained in Supplementary Note 4. The different expected states (P-P, P-E, D-E) were assigned and colored as follows. The major FRET-population was assigned to the P-E state and colored red. Thus, it was chosen to be used for structural modeling. For some datasets only one FRET state is visible. In those cases we assumed that the product and educt state are overlapping. If the P-P and P-E-states have similar amplitudes (ratio smaller than 4:3), as for samples RT(p51Q6C):dp(1)/dt, RT(p51E194C):dp(1)/dt, RT(p66T27C):dp/dt(-6) and RT(p66T27C):dp/dt(-15), an assignment of the distances to the corresponding states is not possible and the amplitude weighted distance  $\langle R \rangle_{E,x}$  with the uncertainties  $\text{err}_+$ ,  $\text{err}_-$  has been used for modeling (see Supplementary Tables 2, 3 and Supplementary Note 15). The P-P state is the peak closer in distance to the P-E state. If the P-P state and the D-E state could not be assigned by the proximity to the P-E state, the higher populated state is assumed to be the P-P state. The P-P state is colored orange. The remaining narrow FRET-peak is assigned to the D-E and colored blue. In the PDA-analysis donor only is colored green. In some cases an additional usually broad peak (colored olive) is necessary to fit the data which is most likely present due to impurities and bleached molecules. Excluding this impurity state reduces fit quality, however, it does not significantly shift the position of the major FRET peak (P-E) as shown for sample RT(p66K287C):dp(1)/dt in Figure D1.1. Here the peak shifts from 43.7 to 41.4 Å, which is well within the uncertainty range (10.3%). This sample was chosen as no other sample yields an impurity state with higher amplitude.



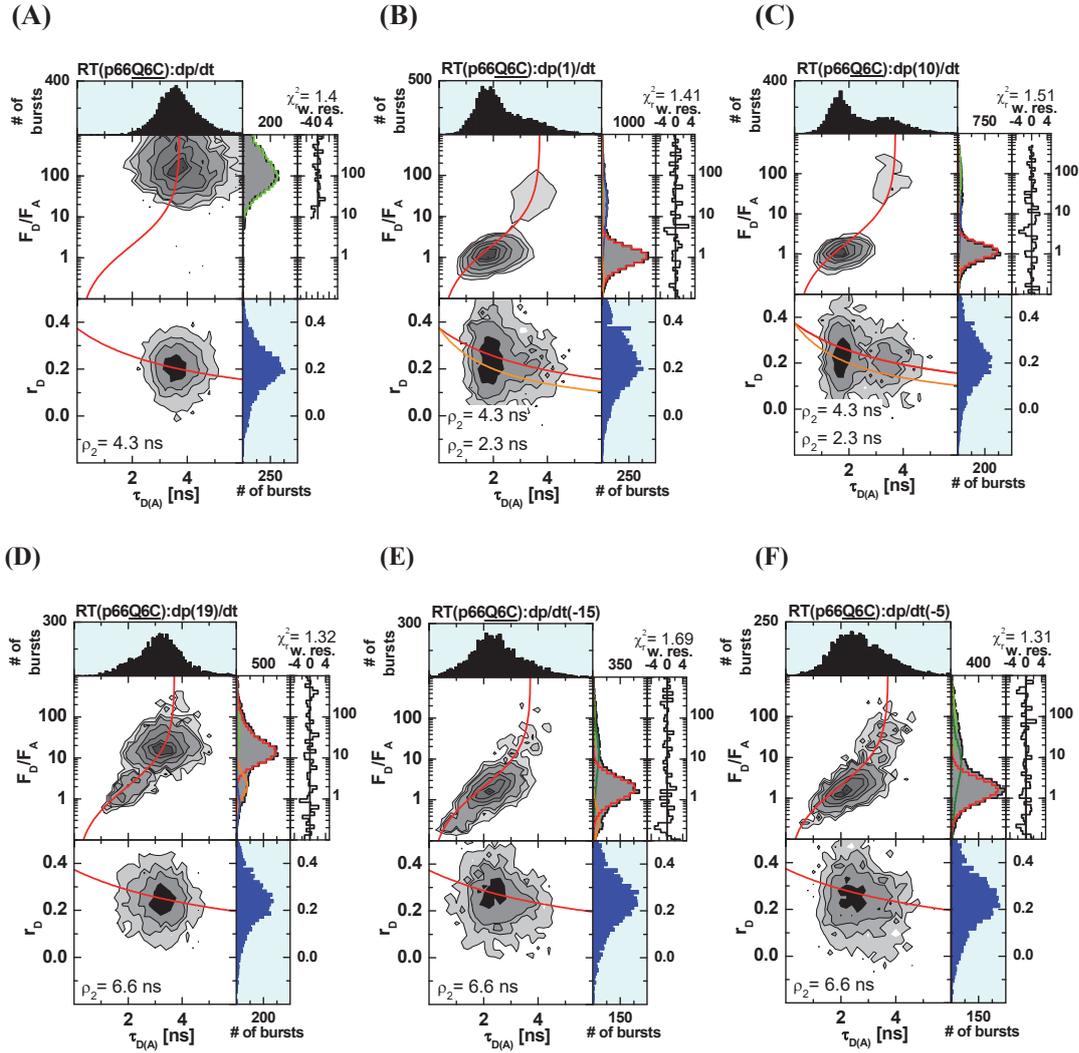
**Figure D1.1:** PDA fits for sample RT(p66K287C):dp(1)/dt (selected bursts) with (A) and without (B) a state accounting for impurities.  $F_D/F_A$  histogram of experimental data (gray area) is fitted (black solid line) using the following parameters: (A) 43.7 Å(26%) (red) 39.3 Å(7.5%) (orange) 81 Å(7.5%) (olive)  $\sigma_{\text{app}}/R=9.0\%$ , 63.1 Å(32.5%)  $\sigma=23.6$  Å (olive),  $x_D=26.5\%$  (green). (B) 41.4 Å(31.9%) (red) 49.8 Å(12.3%) (orange) 71.7 Å(17.8%) (olive)  $\sigma_{\text{app}}/R=9.0\%$ ,  $x_D=38\%$  (green).

**General description of all MFD figures.** Measuring 36 FRET pairs, we present 2D burst frequency histograms of  $F_D/F_A$  versus the donor fluorescence lifetime  $\tau_{D(A)}$  (upper panel) and the donor fluorescence anisotropy  $r_D$  versus  $\tau_{D(A)}$  (lower panel). The number of molecules (fluorescence bursts) in each bin is

## Supplement A

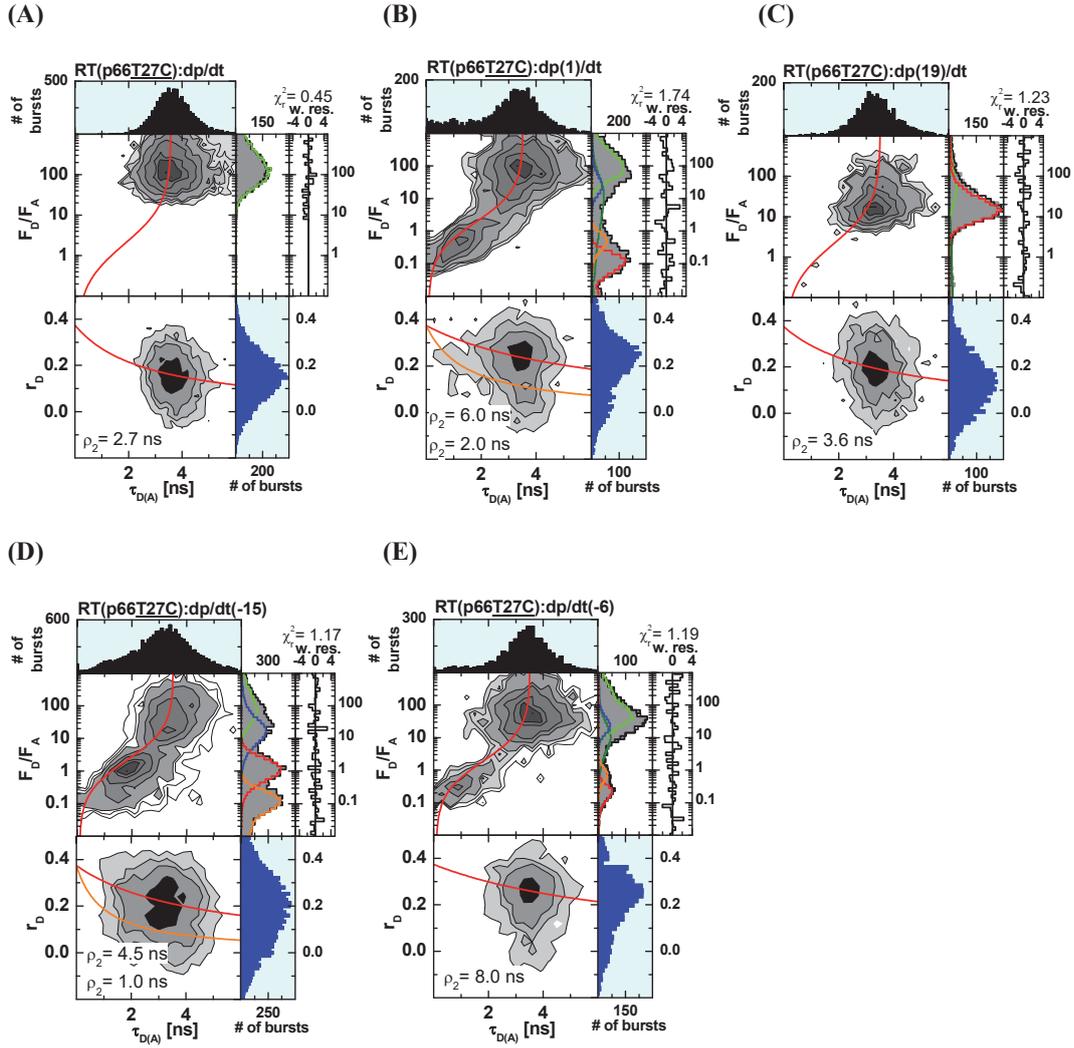
gray scale, shaded from white (lowest) to black (highest). 1D histograms of the lifetime- and anisotropy-distribution are displayed as projections. The theoretical relationship between  $F_D/F_A$  and  $\tau_{D(A)}$  is given by the static FRET line (red line) using equation 4.6 with all parameters compiled in Table N4.1. The solid red and orange lines in the  $r_D$ - $\tau_{D(A)}$  diagram are given by the Perrin equation  $r_D = r_0/(1 + \tau_{D(A)}/\rho)$   $r_0 = 0.374$  whereas  $\rho$  is the rotational correlation time and  $r_0$  the fundamental anisotropy. The PDA analysis of selected bursts,  $\chi_r^2$  and the weighted residuals and of the fits are shown in the upper right panels (The PDA-analysis is not the simple projection of the 2D-histogram. In particular for high-FRET states they species fractions may differ.). The gray area in the PDA- $F_D/F_A$  histogram corresponds to the experimental data and the colored lines to the fitted states. The protein may be in three distinct states (P-P, P-E, and D-E), whereas P-E is generally the major state. Acceptor bleaching and impurities were fitted with Gaussians of free width; the relevant states were fitted with Gaussians with global relative widths of the states  $\sigma/R_i$ . The donor fraction is denoted by  $x_D$ . The relative amplitudes are given in brackets.

## Supplement A



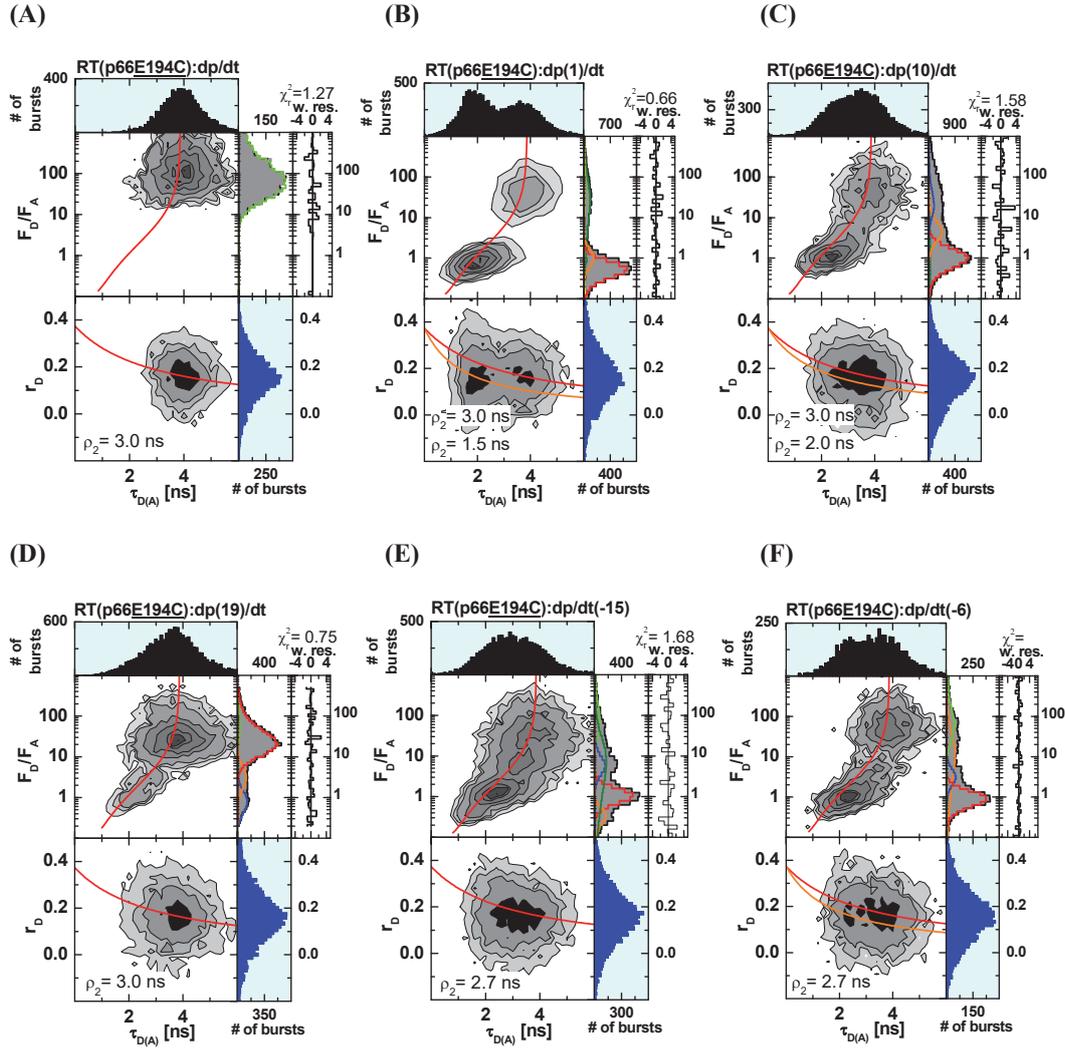
**Figure D1.2:** (A) RT(p66Q6C):p/t donor only sample (B) RT(p66Q6C):dp(1)/dt 46 Å(70%) (red) 37 Å(3%) (orange)  $\sigma_{app}/R=7.8\%$ , 76 Å(23%)  $\sigma=20$  Å (olive),  $x_D=9\%$  (green) (C) RT(p66Q6C):dp(10)/dt 46 Å(58%) (red) 38 Å(3%) (orange)  $\sigma_{app}/R=5.9\%$ , 66 Å(15%)  $\sigma=19$  Å (olive),  $x_D=23\%$  (green) (D) RT(p66Q6C):dp(19)/dt 73 Å(69%) (red) 51 Å(14%) (orange) 42 Å(4%) (blue)  $\sigma_{app}/R=8.5\%$ ,  $x_D=14\%$  (green) (E) RT(p66Q6C):dp/dt(-15) 49 Å(61%) (red) 36 Å(9%) (orange)  $\sigma_{app}/R=10\%$ , 68 Å(19%) (olive),  $x_D=11\%$  (green) (F) RT(p66Q6C):dp/dt(-6) 48 Å(53%) (red)  $\sigma_{app}/R=10\%$ ,  $x_D=20\%$  (green).

## Supplement A



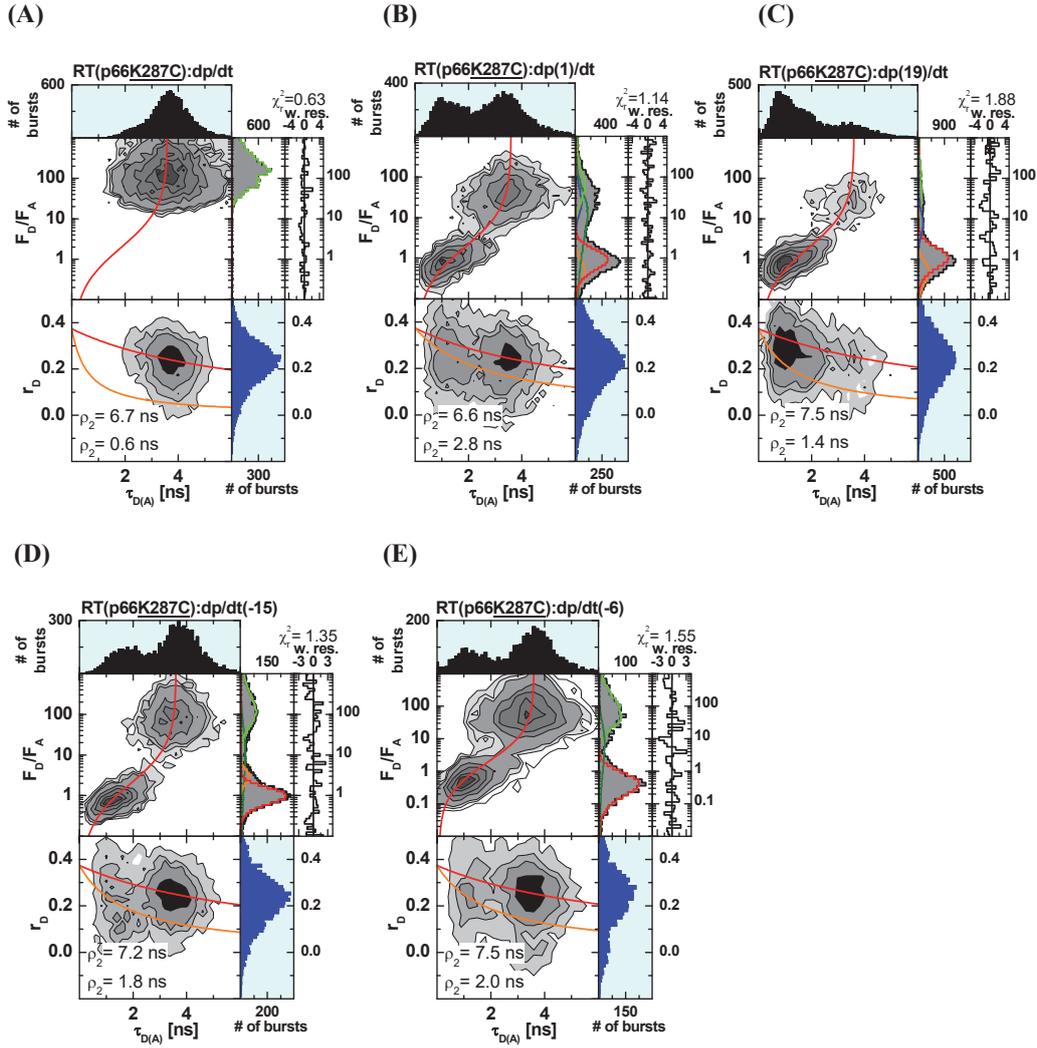
**Figure D1.3:** (A) RT(p66T27C):dp/dt donor only sample (B) RT(p66T27C):dp(1)/dt 30 Å(20%) (red) 39 Å(7%) (orange) 79 Å(6%) (blue)  $\sigma_{app}/R=9.0\%$ , 57 Å(20%)  $\sigma=25$  Å (olive),  $x_D=48\%$  (green) (C) RT(p66T27C):dp(19)/dt 73 Å(59%) (red)  $\sigma_{app}/R=8.0\%$ , 27 Å(16%)  $\sigma=20$  Å  $x_D=25\%$  (green) (D) RT(p66T27C):dp/dt(-15) 29 Å (30%) (red) 45 Å(23%) (orange)  $\sigma_{app}/R=13\%$ , 74 Å(20%)  $\sigma=12$  Å (olive),  $x_D=27\%$  (green) (E) RT(p66T27C):dp/dt(-6) 34 Å(7%) (red) 42 Å(4%) (orange)  $\sigma_{app}/R=7\%$ , 67 Å (14%)  $\sigma=13$  Å (olive), donor only fraction 75% (green).

## Supplement A



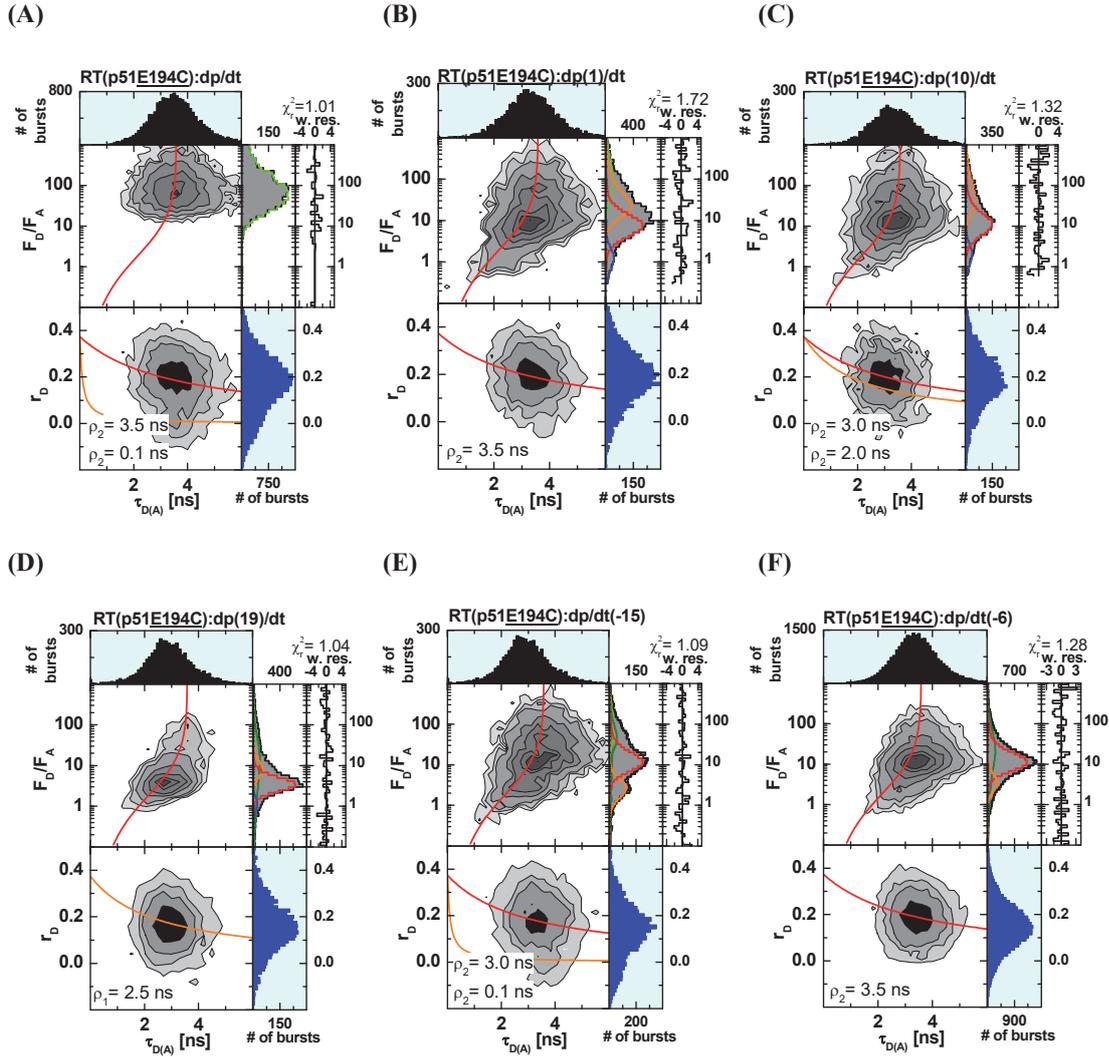
**Figure D1.4:** (A) RT(p66E194C):dp/dt – donor only sample, (B) RT(p66E194C):dp(1)/dt 41 Å(52%) (red) 46 Å(6%) (orange)  $\sigma_{app}/R=7.2\%$ , 84 Å  $\sigma=39$  Å (olive),  $x_D=7.7\%$  (green), red dirt 30 Å (3%) (C) RT(p66E194C):dp(10)/dt 44 Å(35%) (red) 58 Å(13%) (orange) 79 Å (9%)  $\sigma_{app}/R=7.9\%$ , 35 Å(7%)  $\sigma=22$  Å (olive),  $x_D=36\%$  (D) RT(p66E194C):dp(19)/dt 83 Å(58%) (red) 55 Å(11%) (orange) 43 Å(12%) (blue)  $\sigma_{app}/R=9.2\%$ ,  $x_D=20\%$  (E) RT(p66E194C):dp/dt(-15) 46 Å(28%) (red) 61 Å(12%) (blue) 38 Å(8%) (orange)  $\sigma/R=6\%$ , 60 Å(35%)  $\sigma=24$  Å,  $x_D=17\%$  (F) RT(p66E194C):dp/dt(-6) 44 Å(34%) (red) 55 Å(8%) (orange)  $\sigma_{app}/R=7\%$ , 64 Å(28%)  $\sigma=21$  Å,  $x_D=30\%$

## Supplement A



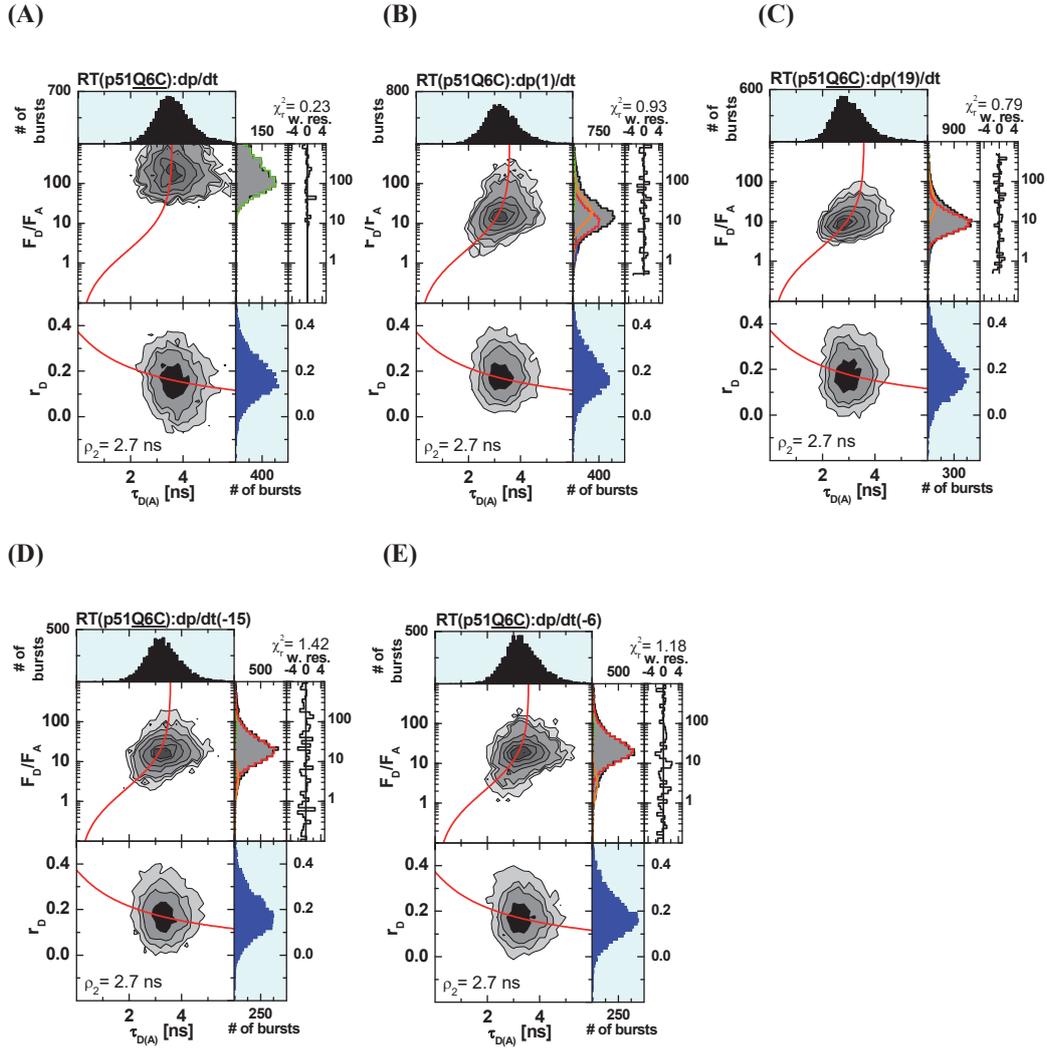
**Figure D1.5:** (A) RT(p66K287C):dp/dt donor only sample (B) RT(p66K287C):dp(1)/dt 43.7 Å(26%) (red) 39.3 Å(7.5%) (orange) 81 Å(7.5%) (olive)  $\sigma_{app}/R=9.0\%$ , 63.1 Å(32.5%)  $\sigma=23.6$  Å (olive),  $x_D=26.5\%$  (green) (C) RT(p66K287C):dp(19)/dt 45 Å(50%) (red) 39 Å(18%) (orange) 64 Å(9%) (blue)  $\sigma_{app}/R=9.1\%$ ,  $x_D=23\%$  (green) (D) RT(p66K287C):dp/dt(-15) 43 Å(32%) (red) 51 Å(5%) (orange)  $\sigma_{app}/R=7\%$ , 57 Å(15%)  $\sigma=23$  Å (olive),  $x_D=48\%$  (green) (E) RT(p66K287C):dp/dt(-6) 36 Å(34%) (red)  $\sigma_{app}/R=13\%$ , 51 Å(15%)  $\sigma=23$  Å (olive),  $x_D=51\%$  (green)

## Supplement A



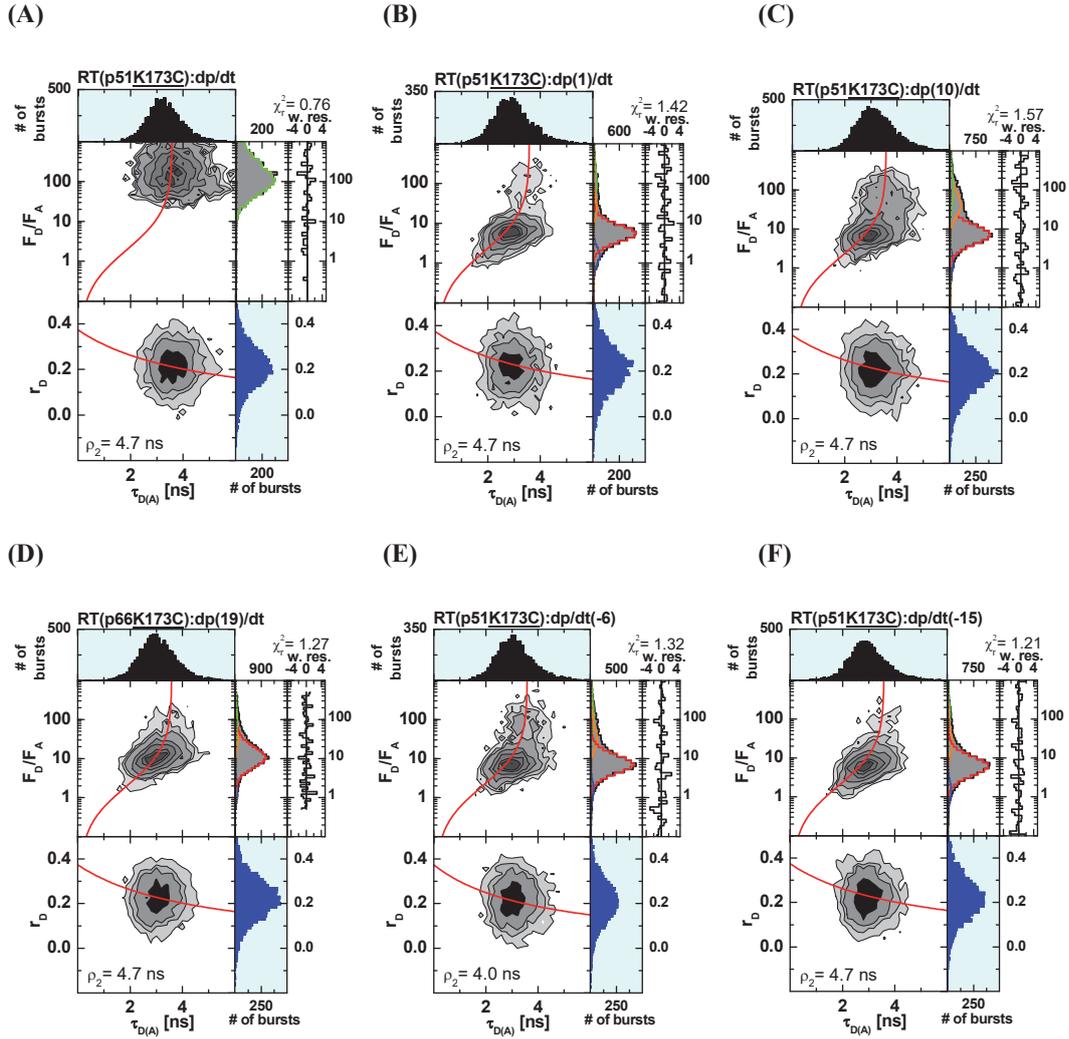
**Figure D1.6:** (A) RT(p51E194C):dp/dt donor only sample (B) RT(p51E194C):dp(1)/dt 63 Å(37%) (red) 79 Å(32%) (orange) 46 Å(8%) (blue)  $\sigma_{app}/R=9.0\%$ ,  $x_D=24\%$  (green) (C) RT(p51E194C):dp(10)/dt 68 Å(53%) (red) 98 Å(36%) (orange) 50 Å (11%) (blue)  $\sigma_{app}/R=9.0\%$  (D) RT(p51E194C):dp(19)/dt 55 Å(40%) (red) 64 Å(11%) (orange) 45 Å(4%) (blue)  $\sigma/R=5.0\%$ , 64 Å(32%)  $\sigma=24$  Å (olive),  $x_D=12\%$  (green) (E) RT(p51E194C):dp/dt(-15) 68 Å(38%) 52 Å (16%)  $\sigma_{app}/R=8\%$ , 65 Å(15%),  $x_D=31\%$  (F) RT(p51E194C):dp/dt(-6) 68 Å(53%) (red) 55 Å(11%) (orange) 45 Å(2%) (blue)  $\sigma_{app}/R=7\%$ , 75 Å (29%)  $\sigma=21$  Å (olive),  $x_D=4\%$

## Supplement A



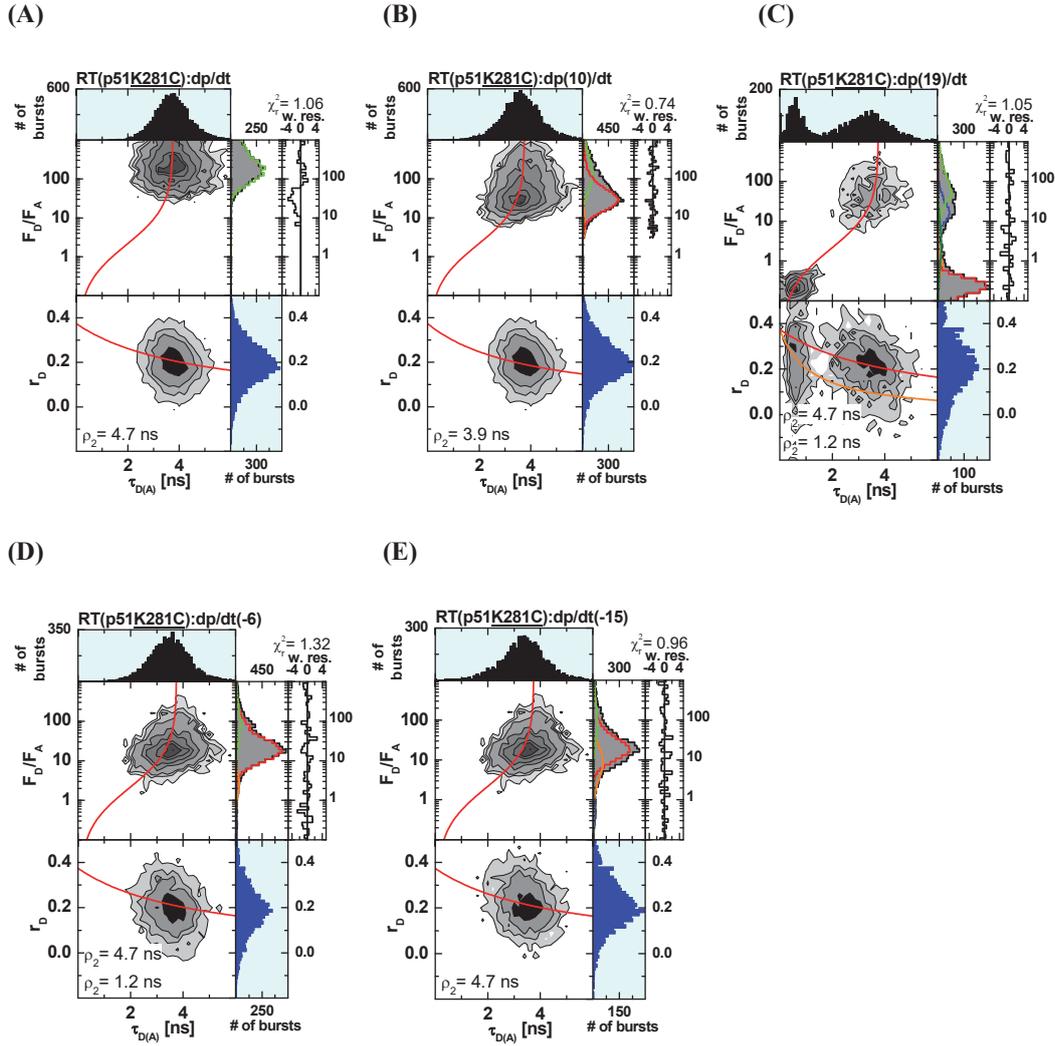
**Figure D1.7:** (A) RT(p51Q6C):dp/dt donor only sample (B) RT(p51Q6C):dp(1)/dt 68 Å(45%) (red) 76 Å(36%) (orange) 51 Å(6%) (blue)  $\sigma_{app}/R=9.0\%$  65 Å(26%)  $\sigma=11$  Å (olive),  $x_D=15\%$  (C) RT(p51Q6C):dp(19)/dt 65 Å(80%) (red) 83 Å(16%) (orange) 49 Å(3%) (blue)  $\sigma_{app}/R=7.3\%$ ,  $x_D=3\%$  (green) (D) RT(p51Q6C):dp/dt(-15) 74Å(83%) (red) 53 Å(5%) (orange)  $\sigma_{app}/R=11\%$ ,  $x_D=12\%$  (E) RT(p51Q6C):dp/dt(-6) 76Å(82%) 54 Å(9%)  $\sigma_{app}/R=11\%$ ,  $x_D=9\%$

## Supplement A



**Figure D1.8:** (A) RT(p51K173C):dp/dt donor only sample (B) RT(p51K173C):dp(1)/dt 60 Å(61%) (red) 79 Å (7%) (orange) 48 Å(9%) (blue)  $\sigma_{app}/R = 7.1\%$ ,  $x_D = 23\%$  (C) RT(p51K173C):dp(10)/dt 62 Å(55%) (red) 88 Å(23%) (orange) 47 Å(4%) (blue)  $\sigma_{app}/R = 8.5\%$ ,  $x_D = 18\%$  (D) RT(p51K173C):dp(19)/dt 67 Å(71%) (red) 86 Å(10%) (orange) 48 Å(5%) (blue)  $\sigma_{app}/R = 9\%$ ,  $x_D = 14\%$  (E) RT(p51K173C):dp/dt(-6) 62 Å(63%) (red) 83 Å(14%) (orange) 48 Å(4%) (blue)  $\sigma_{app}/R = 8\%$ ,  $x_D = 19\%$  (green) (F) RT(p51K173C):dp/dt(-15) 61 Å(69%) 78 Å(11%) 47 Å(6%)  $\sigma_{app}/R = 9\%$ ,  $x_D = 15\%$

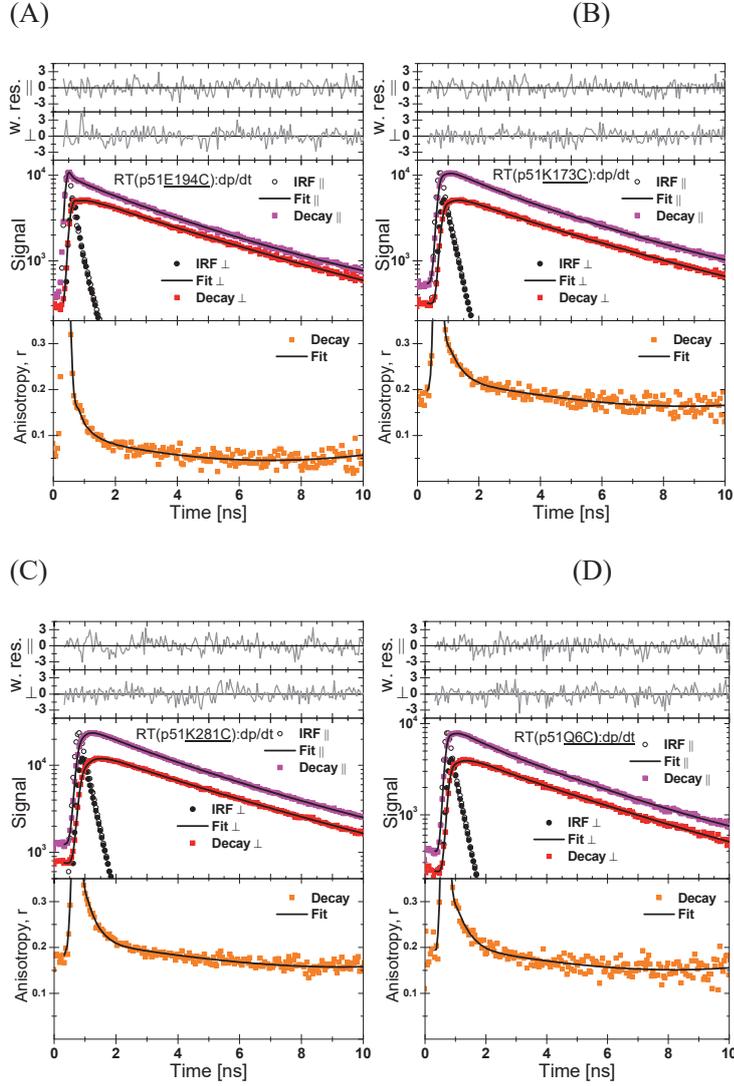
## Supplement A



**Figure D1.9:** (A) RT(p51K281C):dp/dt donor only sample (B) RT(p51K281C):dp(10)/dt 81 Å(54%) (red) 62 Å(3%) (orange)  $\sigma_{app}/R=9.0\%$ ,  $x_D=43\%$  (green) (C) RT(p51K281C):dp(19)/dt 19bp 35 Å(37%) (red) 43 Å (3%) (orange) 74 Å (12%) (blue)  $\sigma_{app}/R=6\%$ , 51 Å(36%)  $\sigma=13$  Å (olive) ,  $x_D=22\%$  (green) (D) RT(p51K281C):dp/dt(-6) 76 Å (76%) (red) 52 Å (4%) (orange) 34 Å (1%) (blue)  $\sigma_{app}/R=10\%$ ,  $x_D=20\%$  (green) (E) RT(p51K281C):dp/dt(-15) 77 Å(59%) (red) Å 64 Å(11%) (orange) 39 Å(3%) (blue)  $\sigma_{app}/R=9\%$ ,  $x_D=28\%$  (green)

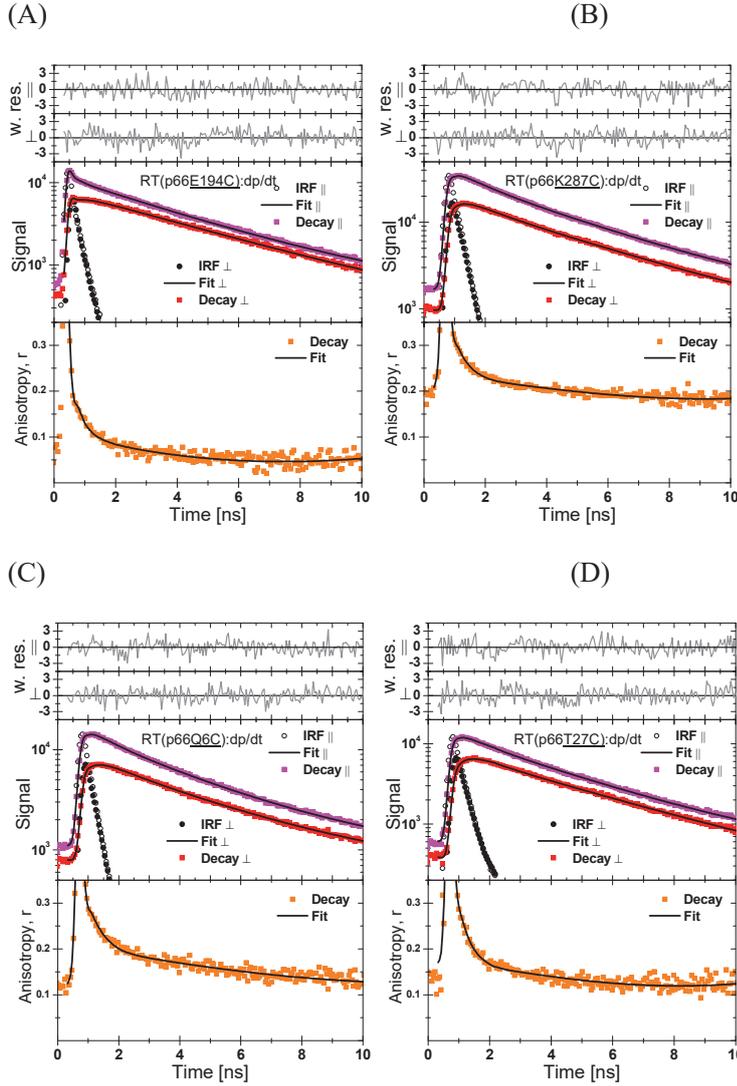
## Supplement A

### Supplementary Data 2: seTCSPC data sets



**Figure D2.1:** Donor only decays,  $r_0=0.38$  (fixed): **(A)** RT(p51E194C):dp/dt -  $s_{||}=7\%$ ,  $bg_{||}=248$ ,  $s_{\perp}=9\%$ ,  $bg_{\perp}=126$ , lifetimes=3.7 ns (88%), 0.9 ns (12%), anisotropy ( $r_0-r_{\infty}$ )=0.25  $r_{\infty}=0.13$ ;  $\rho_{local}=0.10$  ns;  $\rho_{global}=5.45$  ns; **(B)** RT(p51K173C):dp/dt  $s_{||}=1\%$ ,  $bg_{||}=307$ ,  $s_{\perp}=1\%$ ,  $bg_{\perp}=130$ , lifetimes=3.7 ns (83%), 1.0 ns (17%), anisotropy ( $r_0-r_{\infty}$ )=0.12;  $r_{\infty}=0.26$ ;  $\rho_{local}=0.31$  ns;  $\rho_{global}=15.3$  ns; **(C)** RT(p51K281C):dp/dt  $s_{||}=1\%$ ,  $bg_{||}=567$ ,  $s_{\perp}=1.3\%$ ,  $bg_{\perp}=232$ , lifetimes = 3.9 ns (84%), 1.3 ns (16%), anisotropy ( $r_0-r_{\infty}$ )=0.13  $r_{\infty}=0.25$   $\rho_{local}=0.28$  ns;  $\rho_{global}=16.5$  ns; **(D)** RT(p51Q6C):dp/dt  $s_{||}=0.5\%$ ,  $bg_{||}=229$ ,  $s_{\perp}=0.8\%$ ,  $bg_{\perp}=98$ , lifetimes=3.7 ns (84%), 1.0 ns (16%), anisotropy ( $r_0-r_{\infty}$ )=0.14  $r_{\infty}=0.24$  ;  $\rho_{local}=0.34$  ns;  $\rho_{global}=14.2$  ns.

## Supplement A



**Figure D2.2:** Donor only decays,  $r_0=0.38$  (fixed): **(A)** RT(p66E194C):dp/dt -  $s_{\parallel}=4.6\%$ ,  $bg_{\parallel}=352$ ,  $s_{\perp}=6.2\%$   $bg_{\perp}=197$ , lifetime = 3.9 ns (100%), anisotropy ( $r_0-r_{\infty}$ )=0.25;  $r_{\infty}=0.13$ ;  $\rho_{local}=0.14$  ns;  $\rho_{global}=5.53$  ns; **(B)** RT(p66K287C):dp/dt  $s_{\parallel}=2.3\%$ ,  $bg_{\parallel}=888$ ,  $s_{\perp}=3.2\%$   $bg_{\perp}=384$ , lifetimes = 3.75 ns (80%), 0.97 ns (20%), anisotropy ( $r_0-r_{\infty}$ )=0.11;  $r_{\infty}=0.27$ ;  $\rho_{local}=0.30$  ns;  $\rho_{global}=19.6$  ns; **(C)** RT(p66Q6C):dp/dt  $s_{\parallel}=1.0\%$ ,  $bg_{\parallel}=799$ ,  $s_{\perp}=1.4\%$   $bg_{\perp}=557$ , lifetimes=3.72 ns (82%), 1.13 ns (18%), anisotropy ( $r_0-r_{\infty}$ )=0.14  $r_{\infty}=0.24$ ;  $\rho_{local}=0.36$  ns;  $\rho_{global}=17.9$  ns; **(D)** RT(p66T27C):dp/dt  $s_{\parallel}=0.7\%$ ,  $bg_{\parallel}=349$ ,  $s_{\perp}=1.0\%$   $bg_{\perp}=164$ , lifetimes = 3.63 ns (82%), 0.68 ns (18%), anisotropy ( $r_0-r_{\infty}$ )=0.18  $r_{\infty}=0.20$ ;  $\rho_{local}=0.21$  ns;  $\rho_{global}=11.4$  ns.

## Supplement A

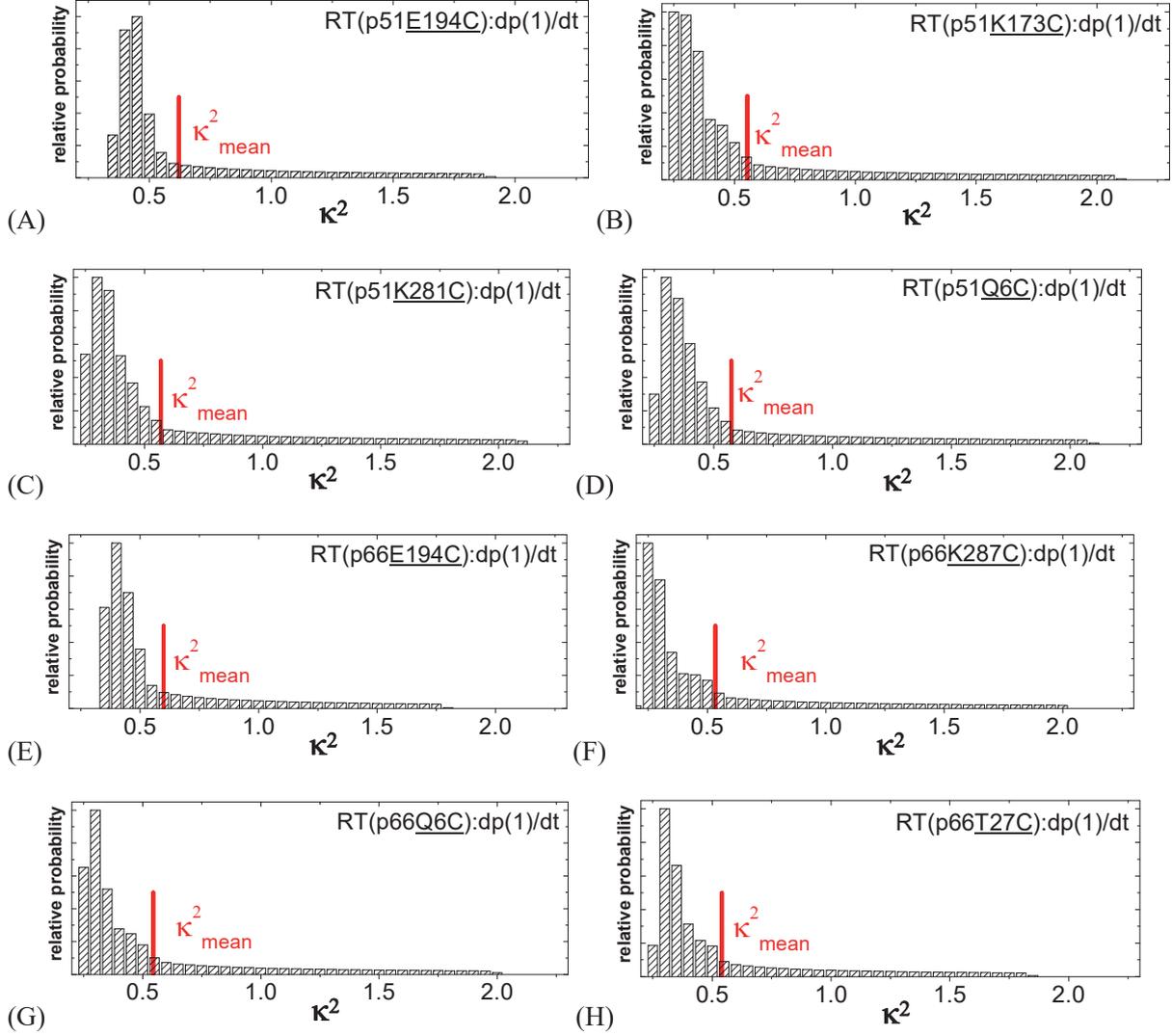
**Table D2.1.** Fitting parameters of the time resolved donor-decays - lifetimes  $\tau_{1,2}$  species amplitudes  $x_{1,2}$  (see equation (5.1)) anisotropies ( $r_1, r_\infty$ ), rotational correlation times  $\rho_1, \rho_2$ .

Sample	$x_1$	$\tau_1$ [ns]	$x_2$	$\tau_2$ [ns]	$(r_0 - r_\infty)$	$\rho_{local}$ [ns]	$r_\infty$	$\rho_{global}$ [ns]
RT(p66Q6C):dp/dt	0.82	3.72	0.18	1.13	0.14	0.36	0.24	17.9
RT(p66T27C):dp/dt	0.82	3.63	0.18	0.68	0.18	0.21	0.20	11.4
RT(p66E194C):dp/dt	1.00	3.87			0.25	0.14	0.13	5.5
RT(p66K287C):dp/dt	0.80	3.75	0.20	0.97	0.11	0.30	0.27	19.6
RT(p51Q6C):dp/dt	0.84	3.71	0.16	0.95	0.14	0.34	0.24	14.2
RT(p51K173C):dp/dt	0.83	3.73	0.17	1.03	0.12	0.31	0.26	15.3
RT(p51E194C):dp/dt	0.88	3.68	0.12	0.92	0.25	0.10	0.13	5.4
RT(p51K281C):dp/dt	0.84	3.91	0.16	1.32	0.13	0.28	0.25	16.5

<sup>[a]</sup>The fundamental anisotropy  $r_0$  was fixed to 0.38.

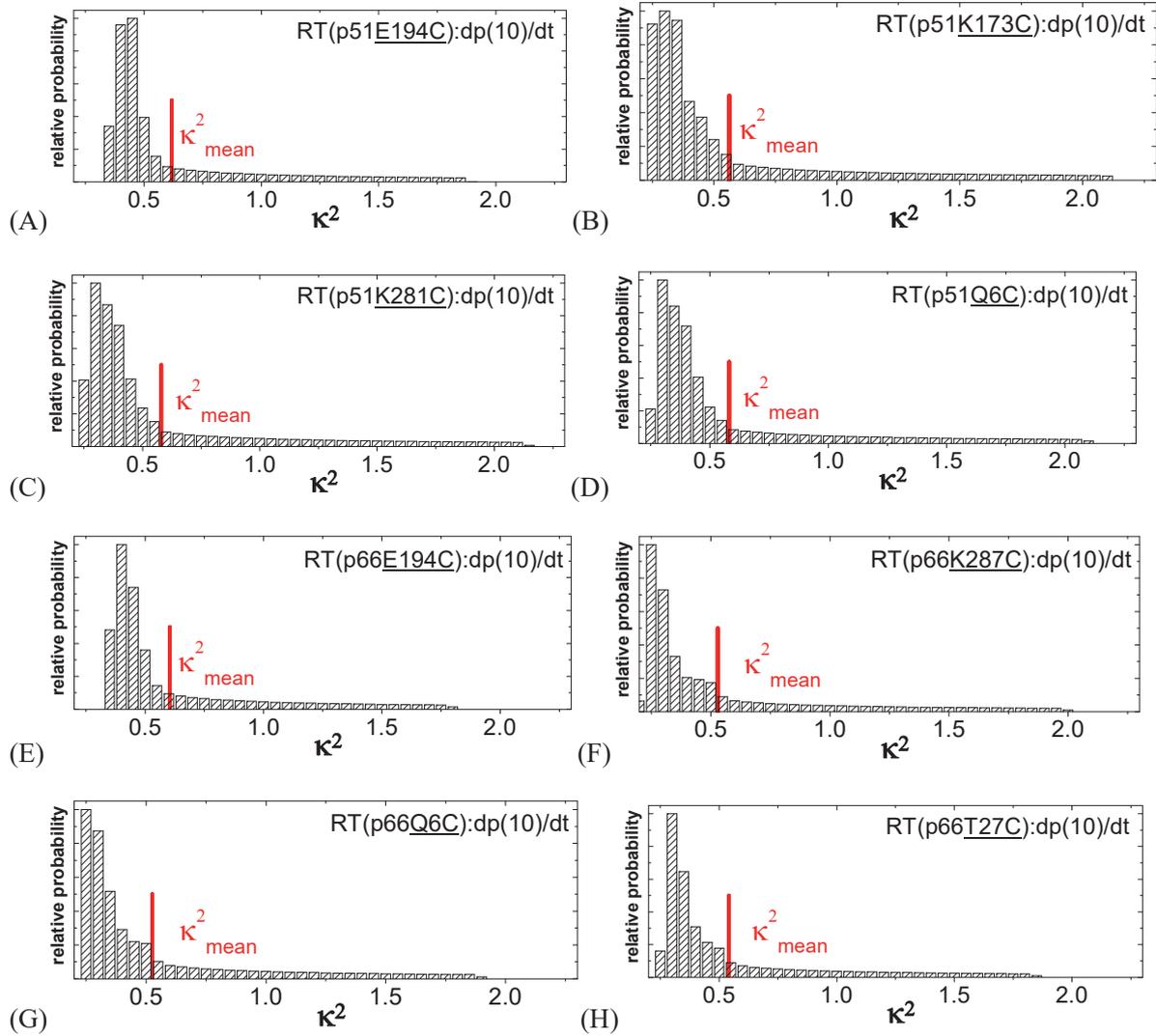
## Supplement A

### Supplementary Data 3: Obtained $\kappa^2$ distributions



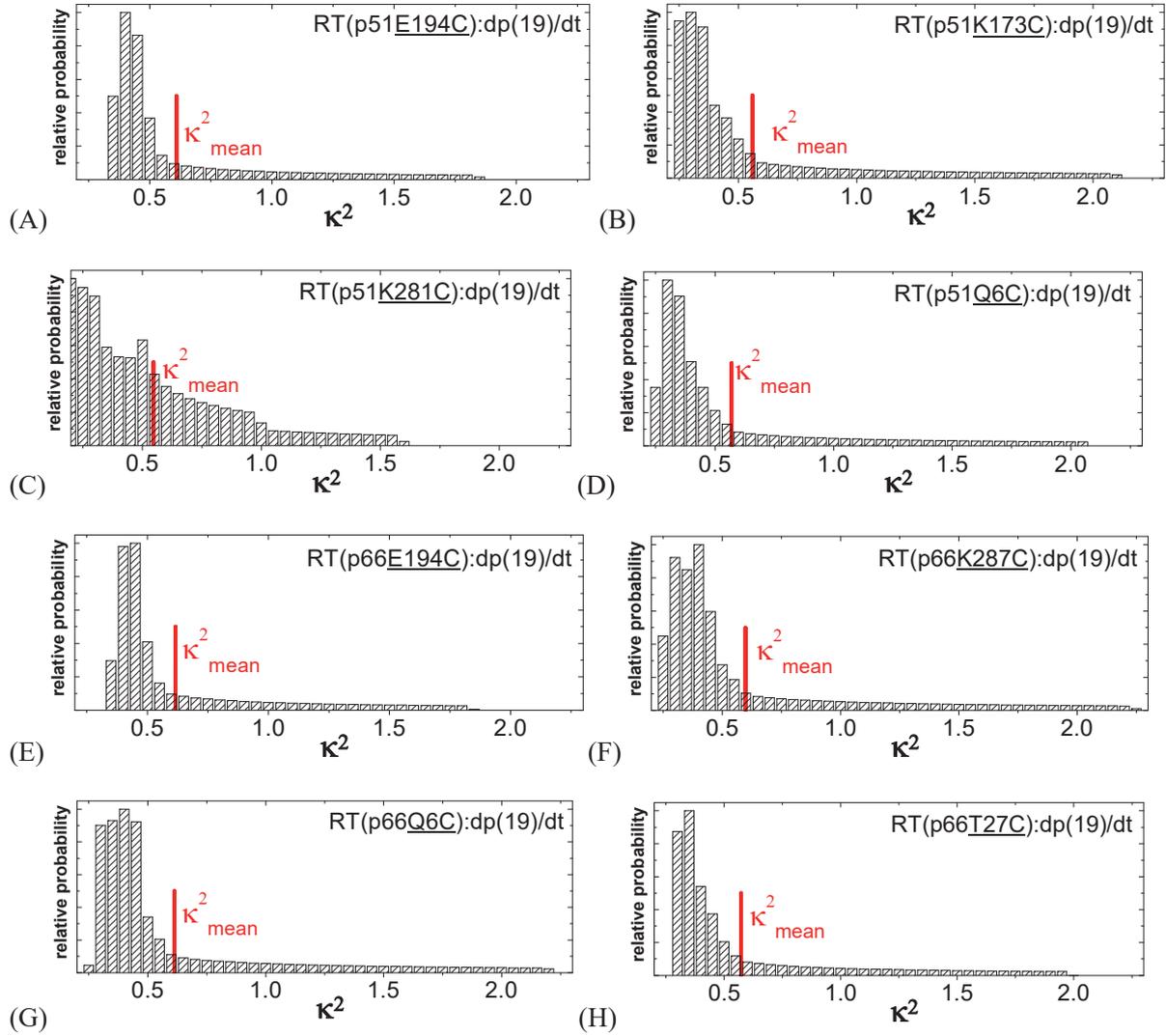
**Figure D3.1:** Possible  $\kappa^2$  values for sample: **(A)** RT(p51E194C):dp(1)/dt  $\langle \kappa^2 \rangle = 0.62$  uncertainty in  $\langle R_{DA} \rangle_E$  7.2% **(B)** RT(p51K173C):dp(1)/dt  $\langle \kappa^2 \rangle = 0.55$  uncertainty in  $\langle R_{DA} \rangle_E$  9.7% **(C)** RT(p51K281C):dp(1)/dt  $\langle \kappa^2 \rangle = 0.57$  uncertainty in  $\langle R_{DA} \rangle_E$  9.7% **(D)** RT(p51Q6C):dp(1)/dt  $\langle \kappa^2 \rangle = 0.57$  uncertainty in  $\langle R_{DA} \rangle_E$  9.4% **(E)** RT(p66E194C):dp(1)/dt  $\langle \kappa^2 \rangle = 0.60$  uncertainty in  $\langle R_{DA} \rangle_E$  7.3% **(F)** RT(p66K287C):dp(1)/dt  $\langle \kappa^2 \rangle = 0.53$  uncertainty in  $\langle R_{DA} \rangle_E$  10.3% **(G)** RT(p66Q6C):dp(1)/dt  $\langle \kappa^2 \rangle = 0.55$  uncertainty in  $\langle R_{DA} \rangle_E$  9.8% **(H)** RT(p66T27C):dp(1)/dt  $\langle \kappa^2 \rangle = 0.54$  uncertainty in  $\langle R_{DA} \rangle_E$  9.0%

## Supplement A



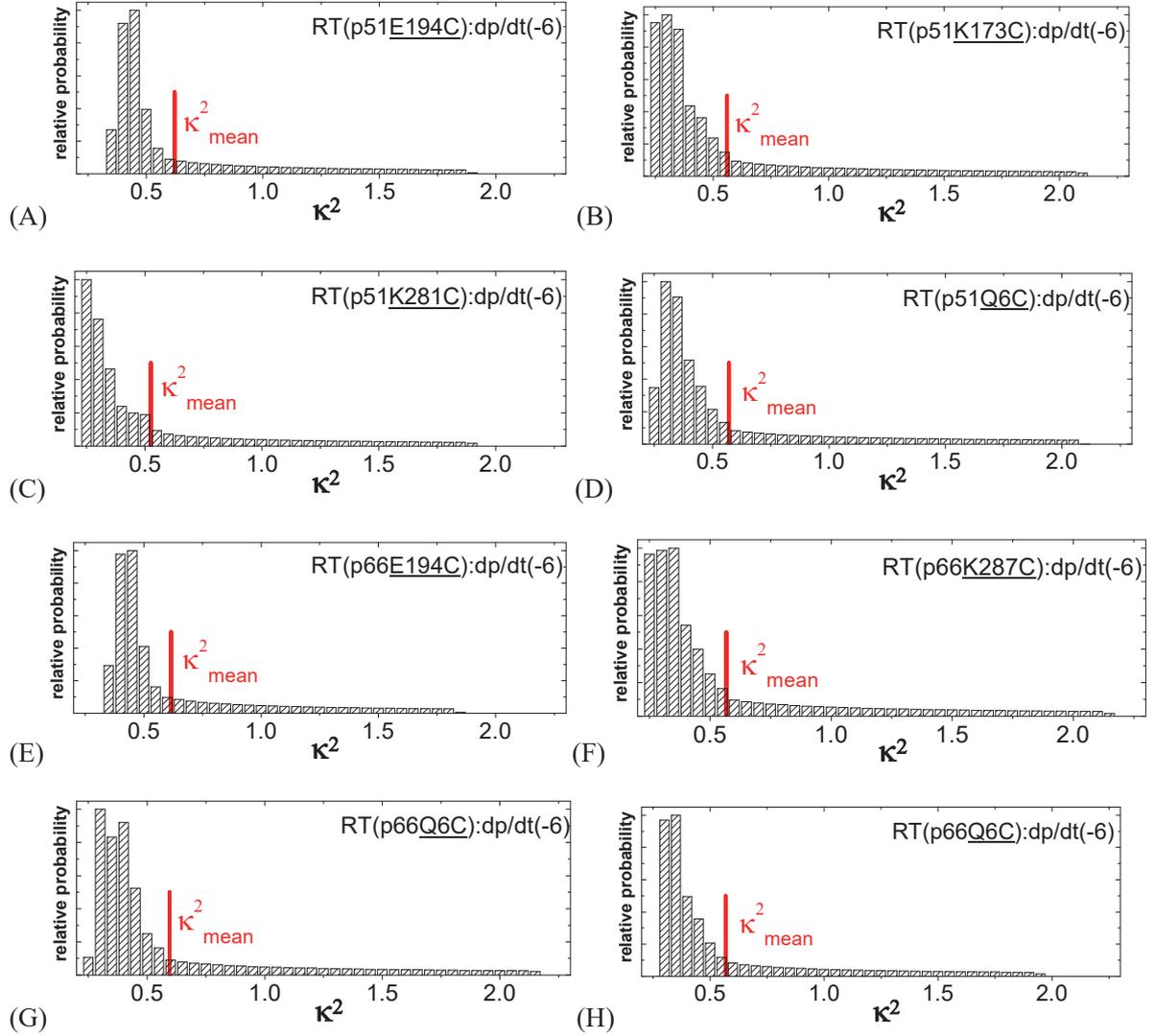
**Figure D3.2:** Possible  $\kappa^2$  values for sample: **(A)** RT(p51E194C):dp(10)/dt  $\langle \kappa^2 \rangle = 0.62$  uncertainty in  $\langle R_{DA} \rangle_E$  7.2% **(B)** RT(p51K173C):dp(10)/dt  $\langle \kappa^2 \rangle = 0.56$  uncertainty in  $\langle R_{DA} \rangle_E$  9.9% **(C)** RT(p51K281C):dp(10)/dt  $\langle \kappa^2 \rangle = 0.58$  uncertainty in  $\langle R_{DA} \rangle_E$  9.5% **(D)** RT(p51Q6C):dp(10)/dt  $\langle \kappa^2 \rangle = 0.58$  uncertainty in  $\langle R_{DA} \rangle_E$  9.3% **(E)** RT(p66E194C):dp(10)/dt  $\langle \kappa^2 \rangle = 0.60$  uncertainty in  $\langle R_{DA} \rangle_E$  7.3% **(F)** RT(p66K287C):dp(10)/dt  $\langle \kappa^2 \rangle = 0.53$  uncertainty in  $\langle R_{DA} \rangle_E$  10.3% **(G)** RT(p66Q6C):dp(10)/dt  $\langle \kappa^2 \rangle = 0.53$  uncertainty in  $\langle R_{DA} \rangle_E$  9.8% **(H)** RT(p66T27C):dp(10)/dt  $\langle \kappa^2 \rangle = 0.54$  uncertainty in  $\langle R_{DA} \rangle_E$  9.0%

## Supplement A



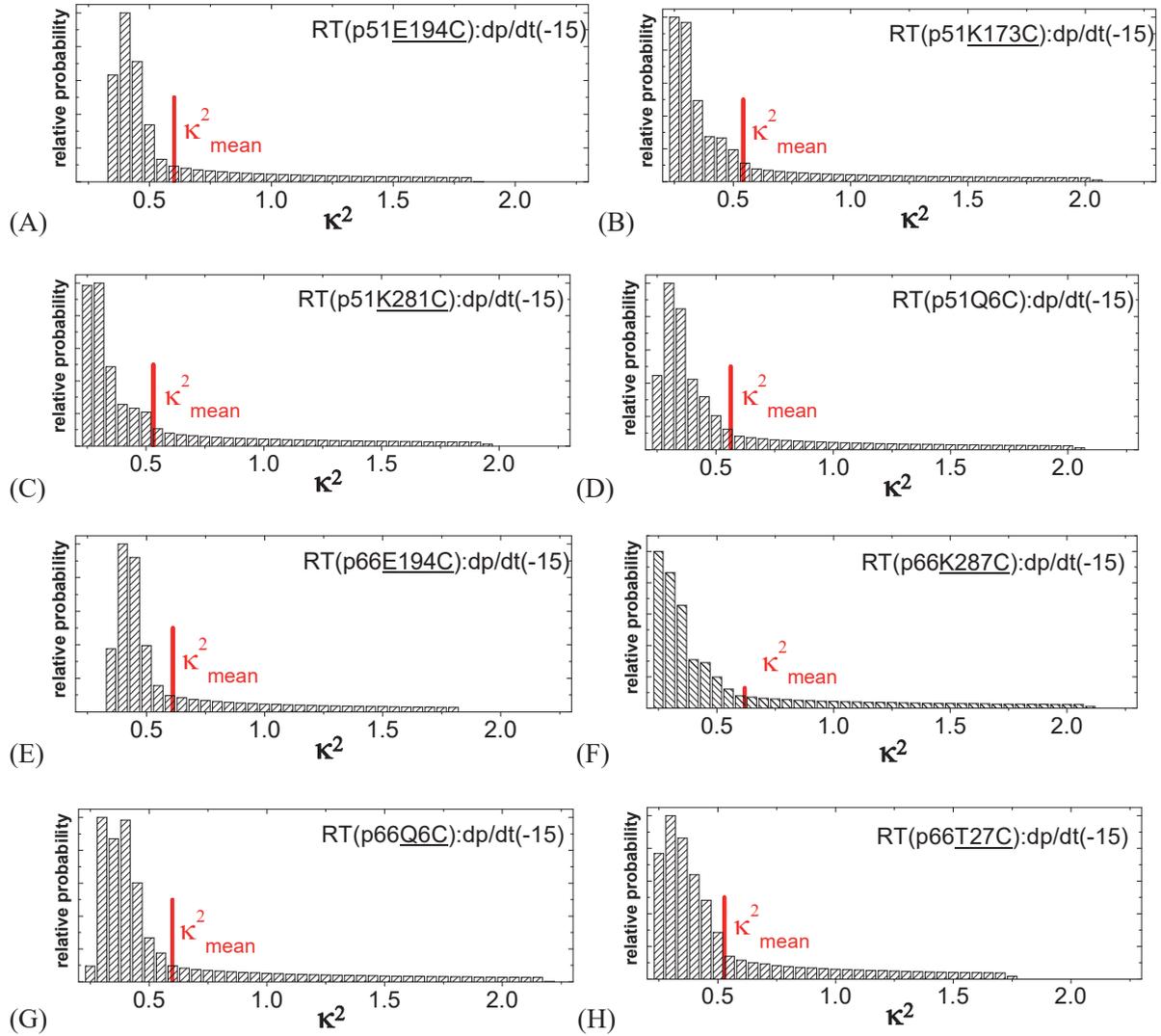
**Figure D3.3:** Possible  $\kappa^2$  values for sample (A) RT(p51E194C):dp(19)/dt  $\langle \kappa^2 \rangle = 0.61$  uncertainty in  $\langle R_{DA} \rangle_E$  7.2% (B) RT(p51K173C):dp(19)/dt  $\langle \kappa^2 \rangle = 0.56$  uncertainty in  $\langle R_{DA} \rangle_E$  10.0% (C) RT(p51K281C):dp(19)/dt  $\langle \kappa^2 \rangle = 0.55$  uncertainty in  $\langle R_{DA} \rangle_E$  10.0% (D) RT(p51Q6C):dp(19)/dt  $\langle \kappa^2 \rangle = 0.57$  uncertainty in  $\langle R_{DA} \rangle_E$  9.4% (E) RT(p66E194C):dp(19)/dt  $\langle \kappa^2 \rangle = 0.62$  uncertainty in  $\langle R_{DA} \rangle_E$  7.2% (F) RT(p66K287C):dp(19)/dt  $\langle \kappa^2 \rangle = 0.60$  uncertainty in  $\langle R_{DA} \rangle_E$  9.6% (G) RT(p66Q6C):dp(19)/dt  $\langle \kappa^2 \rangle = 0.61$  uncertainty in  $\langle R_{DA} \rangle_E$  8.9% (H) RT(p66T27C):dp(19)/dt  $\langle \kappa^2 \rangle = 0.57$  uncertainty in  $\langle R_{DA} \rangle_E$  8.8%

## Supplement A



**Figure D3.4:** Possible  $\kappa^2$  values of sample (A) RT(p51E194C):p/t(-6)  $\langle \kappa^2 \rangle = 0.62$  uncertainty in  $\langle R_{DA} \rangle_E$  7.2% (B) RT(p51K173C):dp/dt(-6)  $\langle \kappa^2 \rangle = 0.56$  uncertainty in  $\langle R_{DA} \rangle_E$  10.0% (C) RT(p51K281C):dp/dt(-6)  $\langle \kappa^2 \rangle = 0.52$  uncertainty in  $\langle R_{DA} \rangle_E$  10.0% (D) RT(p51Q6C):p/t(-6)  $\langle \kappa^2 \rangle = 0.57$  uncertainty in  $\langle R_{DA} \rangle_E$  9.4% (E) RT(p66E194C):p/t(-6)  $\langle \kappa^2 \rangle = 0.62$  uncertainty in  $\langle R_{DA} \rangle_E$  7.2% (F) RT(p66K287C):p/t(-6)  $\langle \kappa^2 \rangle = 0.57$  uncertainty in  $\langle R_{DA} \rangle_E$  10.0% (G) RT(p66Q6C):dp/dt(-6)  $\langle \kappa^2 \rangle = 0.59$  uncertainty in  $\langle R_{DA} \rangle_E$  9.1% (H) RT(p66T27C):dp(19)/dt  $\langle \kappa^2 \rangle = 0.57$  uncertainty in  $\langle R_{DA} \rangle_E$  8.8%

## Supplement A



**Figure D3.5:** Possible  $\kappa^2$  values of sample (A) RT(p51E194C):dp/dt(-15)  $\langle \kappa^2 \rangle = 0.60$  uncertainty in  $\langle R_{DA} \rangle_E$  7.4% (B) RT(p51K173C):dp/dt(-15)  $\langle \kappa^2 \rangle = 0.54$  uncertainty in  $\langle R_{DA} \rangle_E$  10.0% (C) RT(p51K281C):dp/dt(-15)  $\langle \kappa^2 \rangle = 0.53$  uncertainty in  $\langle R_{DA} \rangle_E$  9.9% (D) RT(p51Q6C):dp/dt(-15)  $\langle \kappa^2 \rangle = 0.56$  uncertainty in  $\langle R_{DA} \rangle_E$  9.5% (E) RT(p66E194C):dp/dt(-15)  $\langle \kappa^2 \rangle = 0.61$  uncertainty in  $\langle R_{DA} \rangle_E$  7.2% (F) RT(p66K287C):p/t(-15)  $\langle \kappa^2 \rangle = 0.60$  uncertainty in  $\langle R_{DA} \rangle_E$  10.1% (G) RT(p66Q6C):dp/dt(-15)  $\langle \kappa^2 \rangle = 0.60$  uncertainty in  $\langle R_{DA} \rangle_E$  9.1% (H) RT(p66T27C):dp/dt(-15)  $\langle \kappa^2 \rangle = 0.52$  uncertainty in  $\langle R_{DA} \rangle_E$  9.0%

## Supplement A

### Supplementary Data 4: PDB file for the FRET-restrained structural model of the RT:dp/dt complex including the single strand template overhang

The PDB file contains the positions of dsDNA obtained by rigid body docking (Fig. 6a) and the full conformational ensemble (145 structures) of the template overhang (Fig. 6b) satisfying FRET data. The conformational ensemble of the template overhang is represented by three major configurations depicted in magenta, orange and yellow. Conformer ensemble 1 (magenta: structures 1-23, representative in Fig 6b: conformer 17. Conformer ensemble 2 (orange: structures 24-79, representative in Fig 6b: conformer 31. Conformer ensemble 3 (yellow: structures 80-145, representative in Fig 6b: conformer 84).

Supplementary Data 4 is available on the Nature Methods website at <http://www.nature.com/doifinder/10.1038/nmeth.2222>.

#### Header of the PDB file

```
TITLE   STRUCTURE OF REVERSE TRANSCRIPTASE FROM THE HUMAN
TITLE 2 IMMUNODEFICIENCY VIRUS TYPE 1 (FRET-PDA SCREENED
TITLE 3 TRAJECTORY)
REMARK 1
REMARK 1 REFERENCE 1
REMARK 1 AUTH  S.KALININ,T.O.PEULEN,S.SINDBERT,P.J.ROTHWELL
REMARK 1 AUTH 2 S.BERGER,T.RESTLE,R.S.GOODY,H.GOHLKE,C.A.M.SEIDEL
REMARK 1 TITL  A TOOLKIT AND BENCHMARK STUDY FOR FRET-RESTRAINED
REMARK 1 TITL 2 HIGH-PRECISION STRUCTURAL MODELING
REMARK 1 REF   NAT.METH.   2012
REMARK 2
REMARK 2 MODEL-GENERATION
REMARK 2 SOFTWARE   : AMBER11 (DETAILS SEE PAPER SUPPLEMENT)
REMARK 2 SIMULATION TIME: 571NS
REMARK 2
REMARK 3 DISTANCE-DETERMINATION   : GAUSSIAN FITS OF SUBENSEMBLE TCSPC
REMARK 3 ERROR-ESTIMATION         : WIDTH OF SUBENSEMBLE TCSPC
REMARK 3 FRET-PAIR DISTANCE/ANG, POS.NEG. ERR./ANG :
REMARK 3   P66Q6C :DP/DT(-6) : 56 15.4 15.4
REMARK 3   P66Q6C :DP/DT(-15): 49 19.7 19.7
REMARK 3   P66T27C :DP/DT(-6) : 43 17.6 17.6
REMARK 3   P66T27C :DP/DT(-15): 46 17.0 17.0
REMARK 3   P66E194C:DP/DT(-6) : 58 11.0 11.0
REMARK 3   P66E194C:DP/DT(-15): 54 11.6 11.6
REMARK 3   P66K287C:DP/DT(-6) : 38 11.0 11.0
REMARK 3   P66K287C:DP/DT(-15): 42 11.0 11.0
REMARK 3   P51Q6C  :DP/DT(-6) : 76 20.2 20.2
REMARK 3   P51Q6C  :DP/DT(-15): 74 15.0 15.0
REMARK 3   P51K173C:DP/DT(-6) : 63 20.6 20.6
```

## Supplement A

REMARK 3 P51K173C:DP/DT(-15): 63 15.0 15.0  
REMARK 3 P51E194C:DP/DT(-6) : 69 13.8 13.8  
REMARK 3 P51E193C:DP/DT(-15): 63 15.0 15.0  
REMARK 3 P51K281C:DP/DT(-6) : 70 15.0 15.0  
REMARK 3 P51K281C:DP/DT(-15): 72 15.0 15.0  
REMARK 4 STRUCTURE CLASS COLORS IN MAIN TEXT (FIG 6C)  
REMARK 4 REPRESENTATIVES FOR CLASSES IN BRACKETS  
REMARK 4 1-23 : MAGENTA (17)  
REMARK 4 24-79 : ORANGE (31)  
REMARK 4 80-145: YELLOW (84)  
NUMMDL 145

## Supplement A

### Supplementary Methods

#### 1. HIV-RT:dp/dt complex

Sample preparation and labeling was done according to <sup>9</sup> (<http://hdl.handle.net/2003/5546>). We used RT mutants containing a single accessible cysteine at positions 6, 27, 194 and 287 on the p66 subunit (RT(p66Q6C), RT(p66T27C), RT(p66E194C) and RT(p66K287C), respectively) and at positions 6, 173, 194 and 281 of the p51 subunit (RT(p51Q6C), RT(p51K173C), RT(p66E194C) and RT(p66K281C), respectively) <sup>36</sup> (see Fig. 2b in the main text). These cysteines were labeled with the green donor fluorophore Alexa488-C5 maleimide <sup>37</sup>. All labeled proteins were tested using an RNA-dependent DNA polymerase activity assay to determine whether the labeling procedure impaired enzyme activity <sup>9</sup>. All proteins maintained similar to wild type activity after labeling: RT(wt) 100%, RT(p66Q6C) 182%, RT(p66T27C) 74%, RT(p66E194C) 123%, RT(p66K287C) 81%, RT(p51Q6C) 206%, RT(p51K173C) 70%, RT(p51E194C) 132%, RT(p51K281C) 68%.

The red acceptor dye Cy5 was attached to the primer strand at positions 1, 10 and 19 (dp(1)/dt, dp(10)/dt and dp(19)/dt, respectively) or the template strand at positions -15 and -6 (dp/dt(-15) and dp/dt(-6), respectively) of a 19/35 DNA/DNA primer/template (see Supplementary Figure 2 and Supplementary Table 1 for sequences, labeling positions, secondary structure and numbering of dp/dt) either by internal labeling with a C6-aminolink with the NHS-ester of Cy5 (for dp(10)/dt, dp(19)/dt, dp/dt(-6) and dp/dt(-15)) or to the 3' end of the primer with the phosphoramidite derivative of Cy5 (for dp(1)/dt) (Supplementary Figure 1). The sequence of the primer/template is based on the HIV-1 viral primer binding site.

## Supplement A

### 2. Measurement conditions

The aqueous measurement buffer contained 10 mM KCl, 50 mM Tris·HCl, pH 7.8, 6 mM MgCl<sub>2</sub>, and 400 μM ascorbic acid. Because of the extremely low concentrations used in single-molecule studies (50 pM RT and 200 pM dp/dt), high binding affinities are required to ensure stable complex formation. By reducing the KCl concentration from the 50 mM used in previous studies on this system<sup>38</sup> to 10 mM, we were able to reduce the dissociation constant,  $K_d$ , dramatically. Data from classical ensemble titrations indicated an upper limit for the  $K_d$  value at this salt concentration of <100 pM. The actual single-molecule measurements confirmed that the affinity was high enough to ensure complex formation, because most RT molecules sampled contained bound (labeled) substrate. Ensemble measurements on single nucleotide incorporation under these salt conditions showed similar kinetic behavior to that seen at higher salt, with three kinetic phases being observed<sup>9,38</sup>. Before measurements, a “stock” complex solution was made at a concentration of 250 nM protein and 1 μM dp/dt. This solution was equilibrated on ice for at least 30 min. Immediately before a measurement, the stock solution was diluted rapidly into the measurement buffer 1:10, followed by a further 1:10 dilution before a final 1:50 dilution to a final concentration of 50 pM protein and 200 pM dp/dt. A droplet of the sample solution (50 μl) was applied to a coverslip which formed the bottom of a closed chamber with a water-saturated atmosphere. Coating the coverslip with κ-casein led to complete suppression of adsorption of the sample molecules during the measurement time of up to an hour.

### 3. Multiparameter Fluorescence Detection (MFD)

MFD is performed using a confocal (pinhole  $\varnothing=100\mu\text{m}$ ) epi-illuminated microscope with excitation by a pulsed linearly polarized argon-ion laser at 496 nm. The laser is focused into the solution with a 60x 1.2 water immersion objective. Fluorescence detection is performed with the same objective, but with an additional pinhole that results in a detection volume of 2 fl, as estimated from fluorescence correlation spectroscopy (FCS) measurements. Further setup parameters obtained via FCS are the ratio  $\omega_0/z_0$  of 4, and the characteristic diffusion times  $t_D$  for Rhodamine 110 and RT:dp/dt complex single-labeled at position p66K287C of 0.3 and 1.3 ms, respectively. Here,  $\omega_0$  and  $z_0$  are the distances from the center of the laser beam focus in the radial and axial directions, respectively, at which the collected fluorescence intensity has dropped by a factor of  $e^2$ , compared with its peak value. Sample molecules diffusing freely through the solution occasionally pass through the detection volume, resulting in a brief (~ 1ms) burst of fluorescence. Dilute solutions of molecules (~ 50 pM) ensure that only single molecules are detected, each molecule producing a single burst. The fluorescence signal is divided into parallel and perpendicular components in wavelength ranges below and above 620 nm (green and red, respectively; filters: HQ535/50, HQ730/140). The photons are detected by 4 avalanche photodiodes (APD) coupled to a counting board (SPC 431, Becker & Hickel, Berlin) and a personal computer. For each photon the arrival time after the laser pulse, the time since the last photon, the polarization, and the wavelength is recorded. Fluorescence bursts are distinguished from the background signal of 1 to 2 kHz by applying certain threshold intensity criteria<sup>39</sup>.

## Supplement A

### Supplementary Software

Software for FRET restrained position and screening (FPS) is freely available on the Nature Methods website (<http://www.nature.com/doi/10.1038/nmeth.2222>) and the homepage of the Seidel group (<http://www.mpc.hhu.de/software>). Regular updates and additional tools can be found on our webpage <http://www.mpc.hhu.de/software>.

The file contains the software for FRET-restrained positioning and screening (FPS) with a help file and test data.

At the time of publication the FPS software could be run under Windows, MacOS and Linux.

## Supplement A

### References

1. Cai, Q. *et al.* Nanometer distance measurements in RNA using site-directed spin Labeling. *Biophys. J.* **93**, 2110-2117 (2007).
2. Muschielok, A. *et al.* A nano-positioning system for macromolecular structural analysis. *Nat. Meth.* **5**, 965-971 (2008).
3. Sindbert, S. *et al.* 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 (2011).
4. HyperChem (TM) Professional v. 7.51 (1115 NW 4th Street, Gainesville, Florida 32601, USA).
5. Ding, J. P. *et al.* Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 angstrom resolution. *J. Mol. Biol.* **284**, 1095-1111 (1998).
6. Peletskaya, E. N., Kogon, A. A., Tuske, S., Arnold, E. & Hughes, S. H. Nonnucleoside inhibitor binding affects the interactions of the fingers subdomain of human immunodeficiency virus type 1 reverse transcriptase with DNA. *J. Virol.* **78**, 3387-3397 (2004).
7. Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282**, 1669-1675 (1998).
8. Rothwell, P. J. *et al.* Multi-parameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 Reverse Transcriptase:primer/template complexes. *Proc. Natl. Acad. Sci. USA* **100**, 1655-1660 (2003).
9. Rothwell, J. P. *Structural Investigation of the HIV-1 RT using single pair Fluorescence Energy Transfer*, University of Dortmund, (2002).
10. Woźniak, A. K., Schröder, G., Grubmüller, H., Seidel, C. A. M. & Oesterhelt, F. Single molecule FRET measures bends and kinks in DNA. *Proc. Natl. Acad. Sci. USA.* **105**, 18337-18342 (2008).
11. Kalinin, S., Felekyan, S., Valeri, A. & Seidel, C. A. M. Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis. *J. Phys. Chem. B* **112**, 8361-8374 (2008).
12. Kalinin, S., Felekyan, S., Antonik, M. & Seidel, C. A. M. Probability distribution analysis of single-molecule fluorescence anisotropy and resonance energy transfer. *J. Phys. Chem. B* **111**, 10253-10262 (2007).
13. Antonik, M., Felekyan, S., Gaiduk, A. & Seidel, C. A. M. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J. Phys. Chem. B* **110**, 6970-6978 (2006).
14. Kalinin, S., Valeri, A., Antonik, M., Felekyan, S. & Seidel, C. A. M. 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 (2010).
15. Kalinin, S., Sisamakias, E., Magennis, S. W., Felekyan, S. & Seidel, C. A. M. On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits. *J. Phys. Chem. B* **114**, 6197-6206 (2010).
16. Soong, T. T. *Fundamentals of probability and statistics for engineers.* (John Wiley & Sons, 2004).
17. Dale, R. E., Eisinger, J. & Blumberg, W. E. Orientational freedom of molecular probes - Orientation factor in intra-molecular energy transfer. *Biophys. J.* **26**, 161-193 (1979).
18. Ivanov, V., Li, M. & Mizuuchi, K. Impact of Emission Anisotropy on Fluorescence Spectroscopy and FRET Distance Measurements. *Biophys. J.* **97**, 922-929 (2009).
19. Koshioka, M., Sasaki, K. & Masuhara, H. Time-Dependent Fluorescence Depolarization Analysis in Three Dimensional Microscopy. *Appl. Spectrosc.* **49**, 224-228 (1995).

## Supplement A

20. Choi, U. B. *et al.* Single-molecule FRET-derived model of the synaptotagmin 1-SNARE fusion complex. *Nat. Struct. Mol. Biol.* **17**, 318-U384 (2010).
21. Atkins, P. W. & de Paula, J. *Atkins' Physical Chemistry*. (Oxford University Press, 2001).
22. Martyna, G. J., Tobias, D.J., Kleinm M.L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177-4189 (1994).
23. Marushchak, D., Grenklo, S., Johansson, T., Karlsson, R. & Johansson, L. B. A. Fluorescence depolarization studies of filamentous actin analyzed with a genetic algorithm. *Biophys. J.* **93**, 3291-3299 (2007).
24. Kossiakoff, A. A., Randal, M., Guenot, J. & Eigenbrot, C. Variability of Conformations at Crystal Contacts in Bpti Represent True Low-Energy Structures - Correspondence among Lattice Packing and Molecular-Dynamics Structures. *Proteins-Structure Function and Genetics* **14**, 65-74 (1992).
25. Wlodawer, A., Nachman, J., Gilliland, G. L., Gallagher, W. & Woodward, C. Structure of Form-Iii Crystals of Bovine Pancreatic Trypsin-Inhibitor. *J. Mol. Biol.* **198**, 469-480 (1987).
26. Dauberosguthorpe, P. *et al.* Structure and Energetics of Ligand-Binding to Proteins - Escherichia-Coli Dihydrofolate Reductase Trimethoprim, a Drug-Receptor System. *Proteins-Structure Function and Genetics* **4**, 31-47 (1988).
27. Amber v. Version 10 (University of California, San Francisco, 2008).
28. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668-1688 (2005).
29. Cornell, W. D. *et al.* A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* **117**, 5179-5197 (1995).
30. Simmerling, C., Strockbine, B. & Roitberg, A. E. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* **124**, 11258-11259 (2002).
31. Perez, A. *et al.* Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys. J.* **92**, 3817-3829 (2007).
32. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **79**, 926-935 (1983).
33. Darden, T., York, D. & Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **98**, 10089-10092 (1993).
34. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **23**, 327-341 (1977).
35. Brunger, A. T. Free R-Value - a Novel Statistical Quantity for Assessing the Accuracy of Crystal-Structures. *Nature* **355**, 472-475 (1992).
36. Kensch, O., Restle, T., Wohrl, B. M., Goody, R. S. & Steinhoff, H. J. Temperature-dependent Equilibrium between the Open and Closed Conformation of the p66-Subunit of HIV-1 Reverse Transcriptase Revealed by Site-directed Spin Labelling. *J. Mol. Biol.* **301**, 1029-1039 (2000).
37. Kensch, O. *et al.* HIV-1 Reverse Transcriptase-Pseudoknot RNA Aptamer Interaction Has a Binding Affinity in the Low Picomolar Range Coupled with High Specificity. *J. Biol. Chem.* **275**, 18271-18278 (2000).
38. Wohrl, B. M., Krebs, R., Goody, R. S. & Restle, T. Refined model for primer/template binding by HIV-1 reverse transcriptase: Pre-steady-state kinetic analyses of primer/template binding and nucleotide incorporation events distinguish between different binding modes depending on the nature of the nucleic acid substrate. *J. Mol. Biol.*, 333-344 (1999).
39. Eggeling, C. *et al.* Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *J. Biotechnol.* **86**, 163-180 (2001).

## **Supplement B**

**Triphosphate Induced Dimerization of Human Guanylate Binding Protein 1 Involves Association of the C-Terminal Helices: A Joint Double Electron-Electron Resonance and FRET Study**

# Triphosphate Induced Dimerization of Human Guanylate Binding Protein 1 Involves Association of the C-Terminal Helices: A Joint Double Electron–Electron Resonance and FRET Study

Tobias Vöpel,<sup>†,§,#</sup> Carola S. Hengstenberg,<sup>†,§</sup> Thomas-Otavio Peulen,<sup>||,§</sup> Yathrib Ajaj,<sup>||</sup> Claus A. M. Seidel,<sup>\*,||</sup> Christian Herrmann,<sup>\*,†</sup> and Johann P. Klare<sup>\*,‡</sup>

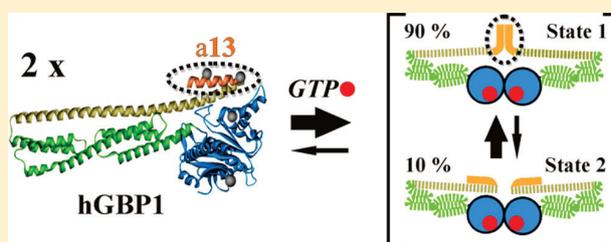
<sup>†</sup>Physical Chemistry I, Faculty of Chemistry and Biochemistry, Ruhr-University Bochum, Universitätsstrasse 150, 44780 Bochum, Germany

<sup>||</sup>Chair for Molecular Physical Chemistry, Heinrich-Heine University, Universitätsstrasse 1, 40225 Düsseldorf, Germany

<sup>\*</sup>Macromolecular Structure Group, Department of Physics, University of Osnabrück, Barbarastrasse 7, 49076 Osnabrück, Germany

## Supporting Information

**ABSTRACT:** Human guanylate binding protein 1 (hGBP1) is a member of the dynamin superfamily of large GTPases. During GTP hydrolysis, the protein undergoes structural changes leading to self-assembly. Previous studies have suggested dimerization of the protein by means of its large GTPase (LG) domain and significant conformational changes in helical regions near the LG domain and at its C-terminus. We used site-directed labeling and a combination of pulsed electron paramagnetic resonance and time-resolved fluorescence spectroscopy for structural investigations on hGBP1 dimerization and conformational changes of its C-terminal helix  $\alpha 13$ . Consistent distance measurements by double electron–electron resonance (DEER, also named pulse double electron resonance = PELDOR) spectroscopy and Förster resonance energy transfer (FRET) measurements using model-free analysis approaches revealed a close interaction of the two  $\alpha 13$  helices in the hGBP1 dimer formed upon binding of the nonhydrolyzable nucleoside triphosphate derivate GppNHp. In molecular dynamics (MD) simulations, these two helices form a stable dimer in solution. Our data show that dimer formation of hGBP1 involves multiple spatially distant regions of the protein, namely, the N-terminal LG domain and the C-terminal helices  $\alpha 13$ . The contacts formed between the two  $\alpha 13$  helices and the resulting juxtaposition are expected to be a key step for the physiological membrane localization of hGBP1 through the farnesyl groups attached to the end of  $\alpha 13$ .



Human guanylate binding protein 1 (hGBP1) belongs to the dynamin superfamily of large GTPases.<sup>1</sup> A common feature of this family of proteins is the nucleotide-dependent assembly to homotypic oligomers, leading to stimulation of the GTPase activity, which in the case of dynamin accounts for scission of vesicles from membranes.<sup>1</sup> In contrast to dynamin, the cellular function of the hGBPs belonging to the same superfamily of large GTPases is only partially understood. Expression of hGBP1 is mainly activated by type II IFN (interferon  $\gamma$ ).<sup>2</sup> It participates in immune responses against viral and bacterial targets such as hepatitis C virus, influenza A virus, and bacterial meningitis,<sup>3–8</sup> and exhibits antiangiogenic<sup>9–11</sup> and antitumoral<sup>12–16</sup> activities. The antiangiogenic effects of inflammatory cytokines in cultivated endothelial cells *in vitro*<sup>9,10</sup> and in tumor vessel endothelial cells of colorectal carcinoma patients *in vivo*<sup>11</sup> have been shown to be mediated by hGBP1. Furthermore, hGBP1 overexpression is associated with different tumor types, such as glioblastoma,<sup>12</sup> oral cancer,<sup>13</sup> and mammary cancer,<sup>14</sup> and was also observed to be associated with paclitaxel drug resistance in ovarian cancer cells and with docetaxel resistance in prostate cancer cells.<sup>15,16</sup>

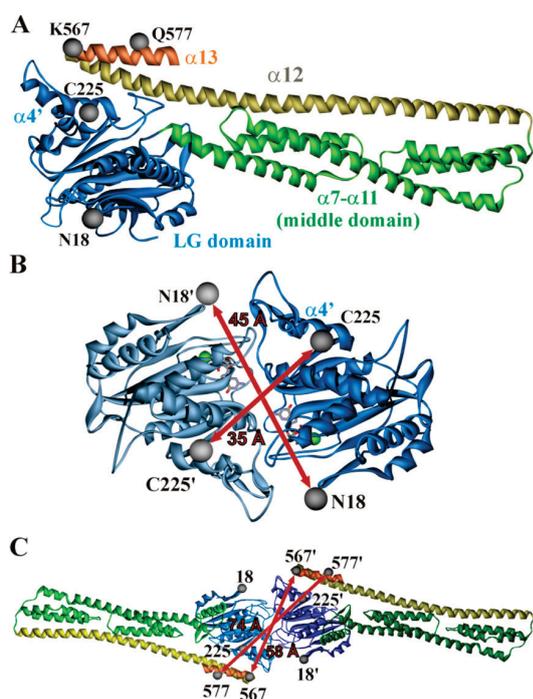
The elongated 67 kDa protein is generally assumed to consist of three domains (Figure 1A). The LG domain (blue) resembling the canonical GTPase domain of Ras with insertions and extensions is followed by the purely  $\alpha$ -helical middle domain (green) and the  $\alpha 12/13$  domain (dark yellow/orange) which flanks the full length of the protein. At the C-terminal end the  $\alpha 12/13$  domain has contacts with the LG domain, thereby stabilizing its position.<sup>17</sup> hGBP1 can bind all three guanine nucleotides with similar affinities in the low micromolar range.<sup>18–20</sup> It hydrolyzes GTP to GDP and GMP in two successive steps utilizing the same catalytic center.<sup>19,21–24</sup>

Using size exclusion chromatography and dynamic light scattering, it was shown that binding of the nonhydrolyzable GTP analogue GppNHp results in the formation of dimers, and binding of the GTP hydrolysis transition mimic GDP·AlF<sub>x</sub> leads to the formation of tetramers.<sup>17</sup> This oligomer formation

**Received:** April 30, 2014

**Revised:** July 3, 2014

**Published:** July 3, 2014



**Figure 1.** Crystal structures and a model for the dimer of hGBP1. (A) Apo state (pdb: 1DG3).<sup>17</sup> Spin labeled residues are marked by gray spheres at the positions of their C $\alpha$  atoms. (B) LG domain dimer structure in the presence of GppNHp (pdb: 2BC9).<sup>23</sup> (C) Putative head-to-head dimer model obtained by superimposing the LG domains of the full-length crystal structure in the GppNHp bound state (pdb: 1F5N)<sup>24</sup> onto the LG domains in the dimer structure obtained in the presence of GppNHp (pdb: 2BC9); the expected distances based on this structure are summarized in Table 1.<sup>23</sup> Dark-red arrows and numbers in panels B and C indicate the expected distances (C $\alpha$ –C $\alpha$ ) based on the crystal structure and the dimer model, respectively.

of hGBP1 triggered by GTP binding leads to a self-stimulation of its GTPase reaction.<sup>25</sup> Therefore, hGBP1 can be categorized as a G protein activated by nucleotide-dependent dimerization.<sup>26</sup> The binding of GppNHp leads to the formation of a putative head-to-head dimer (Figure 1B,C), and structural rearrangements within the LG domain that are believed to be transmitted to  $\alpha$ 12/13 located at the C-terminus. This in turn is believed to control the oligomerization behavior and the second step of nucleotide hydrolysis.<sup>27–29</sup> However, no detailed structural information on the full length protein in the dimeric or tetrameric state could be obtained so far.

To evaluate the hGBP1 dimer arrangement and possible conformational changes upon dimerization in the GppNHp-bound state, we applied two labeling based techniques for several reasons: (i) to check the consistency between the observations made by electron paramagnetic resonance (EPR) spectroscopy on frozen samples (50 K) and by FRET at physiological temperatures (298 K), (ii) to exclude label-specific effects on the measurements, (iii) to increase the range of accessible distances, and (iv) to take advantage of the method-specific sensitivity ranges for detecting also minor populations that exhibit different structures. Quantitative distance measurements were performed by DEER (PELDOR) spectroscopy<sup>30–34</sup> using site-directed spin labeling (SDSL),<sup>32</sup> and by Förster resonance energy transfer (FRET) using

ensemble time-correlated single photon counting (eTCSPC)<sup>35–37</sup> on samples carrying fluorescence labels.

## EXPERIMENTAL PROCEDURES

**Mutagenesis.** All cysteine mutants were constructed based on a cysteine-free hGBP1 variant (C12A/C82A/C225S/C235A/C270A/C311S/C396A/C407S/C589S). All mutants used were generated by QuikChange site directed mutagenesis (Agilent Technologies Sales & Services GmbH & Co.KG, Germany) using templates of hGBP1 in the pQE80L vector (Qiagen GmbH, Germany) as described earlier.<sup>28</sup> All products were verified by DNA sequencing.

**Protein Preparation, Spin Labeling, and FRET Labeling.** All proteins were expressed from a pQE80L vector (Qiagen GmbH, Germany) in the *Escherichia coli* strain BL21 (DE3). Protein preparations were performed as described.<sup>19</sup> The buffers did not contain any DTE as it would interfere with the following labeling reaction. Concentrations were determined by the protein absorbance at 276 nm ( $\epsilon_{276}$  (hGBP1) = 45400 M<sup>-1</sup> cm<sup>-1</sup>) according to the method described by Gill and von Hippel.<sup>38</sup>

For the spin labeling reaction, all proteins were incubated with an 8-fold excess of (1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate spin label MTSSL (Enzo Life Sciences GmbH, Germany) for 3 h at 4 °C. The reaction was performed in 50 mM Tris, 5 mM MgCl<sub>2</sub> solved in D<sub>2</sub>O at pH 7.4. Unbound spin labels were removed with Zeba Spin Desalting Columns (Thermo Fisher Scientific GmbH, Germany) equilibrated with 50 mM Tris, 5 mM MgCl<sub>2</sub> solved in D<sub>2</sub>O at pH 7.4. Concentrations were determined as described above. Labeling efficiencies have been determined by double integration of cw room temperature (RT) EPR spectra in comparison with EPR samples of known concentrations and were ~90–100% in all cases. In all EPR experiments, the protein concentration was 100  $\mu$ M.

For labeling with fluorescent dyes proteins were incubated with a 4-fold excess of either Alexa488-C5-maleimide or Alexa647-C2-maleimide (Life Technologies GmbH, Germany) for 1.5 h at 4 °C. The fluorescence labeling buffer had the same composition (solved in H<sub>2</sub>O) as described for the spin labeling reaction. Unbound dye was removed with Zeba Spin Desalting Columns (Thermo Fisher Scientific GmbH, Germany). The labeling degree was determined by UV absorption measurements to 70%, 93%, 63%, and 53% for Q577C<sup>D</sup>, Q577C<sup>A</sup>, N18C<sup>D</sup>, and N18C<sup>A</sup>, respectively.

**GTPase Activity.** The hydrolytic activity of the different protein mutants was measured by high performance liquid chromatography using a Chromolith Performance RP-18 end-capped column (Merck KGaA, Germany) as described earlier.<sup>27</sup> 2  $\mu$ M protein was incubated with 350  $\mu$ M GTP at 25 °C. Samples were analyzed after different reaction periods. The time dependence of the substrate concentration was used to calculate the specific activities of the different protein mutants. The obtained values are given in Supplementary Table S2, Supporting Information.

**cw EPR Measurements.** Room temperature continuous wave (cw) EPR spectra were recorded on a Miniscope X-band benchtop EPR spectrometer MS200 (Magnetech GmbH, Germany) equipped with a rectangular TE102 resonator. The microwave power was set to 10 mW and the B-field modulation to 0.15 mT. Twenty microliters of sample volume containing 100  $\mu$ M protein was filled in EPR glass capillaries (0.9 mm inner diameter).

Cw EPR spectra for interspin distance determination in the range from ~0.8 to 1.7 nm were obtained on a homemade cw X-band EPR spectrometer equipped with a Super High Sensitivity Probehead (BrukerBiospin GmbH, Germany). The magnetic field was measured with a RMN-2 B-field meter (Drusch GmbH, Germany). A continuous flow cryostat Oxford ESR900 (Oxford Instruments, UK) was used in combination with an Intelligent Temperature Controller ITC 4 (Oxford Instruments, UK) allowing the stabilization of the sample temperature to 160 K. The microwave power was set to 0.2 mW and the B-field modulation amplitude was set to 0.25 mT. EPR quartz capillaries (3 mm inner diameter) were filled with sample volumes of 40  $\mu$ L. Fitting of simulated dipolar broadened EPR powder spectra to the experimental ones detected at 160 K was carried out using the program ShortDistances by Chr. Altenbach.<sup>39</sup>

**Pulse EPR Measurements.** Pulse EPR experiments (DEER) were done at X-band frequencies (9.3–9.5 GHz) with a Bruker Elexsys 580 spectrometer equipped with a Bruker Flexline split-ring resonator ER 4118X-MS3 (Bruker Biospin GmbH, Germany) and a continuous flow helium cryostat CF935 (Oxford Instruments, UK) controlled by an Oxford Intelligent Temperature Controller ITC 503S. Sample conditions for the EPR experiments were 100  $\mu$ M protein in 100 mM NaCl, 50 mM Tris-HCl, 5 mM MgCl<sub>2</sub>, pH 7.4 dissolved in D<sub>2</sub>O with 12.5% (v/v) glycerol-*d*<sub>8</sub>, and 1 mM GDP or 1 mM GppNHp, respectively.

All measurements were performed using the four-pulse DEER sequence:  $\pi/2(v_{\text{obs}}) - \tau_1 - \pi(v_{\text{obs}}) - t' - \pi(v_{\text{pump}}) - (\tau_1 + \tau_2 - t') - \pi(v_{\text{obs}}) - \tau_2 - \text{echo}$ .<sup>30,31</sup> A two-step phase cycling (+(*x*), -(*x*)) was performed on  $\pi/2(v_{\text{obs}})$ . Time *t'* is varied, whereas  $\tau_1$  and  $\tau_2$  are kept constant, and the dipolar evolution time is given by  $t = t' - \tau_1$ . Data were analyzed only for  $t > 0$ . The resonator was overcoupled to  $Q \approx 100$ ; the pump frequency  $\nu_{\text{pump}}$  was set to the center of the resonator dip and coincided with the maximum of the nitroxide EPR spectrum, whereas the observer frequency  $\nu_{\text{obs}}$  was ~65 MHz higher, coinciding with the low field local maximum of the spectrum. All measurements were performed at a temperature of 50 K with observer pulse lengths of 16 ns for  $\pi/2$  and 32 ns for  $\pi$  pulses and a pump pulse length of 12 ns. Deuterium modulation was averaged by adding traces at eight different  $\tau_1$  values, starting at  $\tau_{1,0} = 400$  ns and incrementing by  $\Delta\tau_1 = 56$  ns. Data points were collected in 8 ns time steps or, if the absence of fractions in the distance distribution below an appropriate threshold was checked experimentally, in 16 ns time steps. The total measurement time for each sample was 4–24 h. Analysis of the data was performed with DeerAnalysis 2011<sup>40</sup> using a Tikhonov regularization approach.<sup>41</sup> For details about DEER data analysis, see Supporting Information.

**Rotamer Library Analysis.** Inter-spin label distance distributions were simulated using a rotamer library of spin labeled residues as described earlier.<sup>42</sup> The rotamer library implemented in the software package MMM2011<sup>41</sup> consisted of 210 rotamers of MTSSL bound to cysteine, which have been used to replace the native residues at the positions of interest in the respective hGBP1 structural models. Energies and resulting populations for individual rotamers were calculated by means of a Lennard–Jones potential at 175 K (the glass transition temperature for a water–glycerol mixture) and have been used as weights in the simulation of the distance distributions. For more details about the RLA, see Supporting Information.

**FRET Measurements.** Ensemble time-correlated single-photon-counting (eTCSPC) measurements were performed on an IBH-5000U (HORIBA Jobin Yvon IBH Ltd., UK) system. The excitation source was a 470 nm diode laser LDH-P-C 470 (Picoquant GmbH, Germany) operating at 8 MHz for donor excitation. The emission wavelength was set to 520 nm for donor emission. The corresponding monochromator slits were set to 2 nm (excitation path) and 16 nm (emission path) resolution. An additional 500 nm cutoff filter was used to reduce the contribution of the scattered light. All measurements were performed at room temperature under magic-angle with a total protein concentration of approximately 9  $\mu$ M. The concentration of the donor-labeled protein was 0.5  $\mu$ M. The actual donor-, acceptor-, and unlabeled-protein concentrations were calculated based on degree of labeling of the individual samples determined by fluorescence and UV-spectroscopy and are summarized in Supplementary Figure S4, Supporting Information. The fluorescence intensity decay curves were fitted using the iterative deconvolution approach.<sup>43</sup> To correct instrumental nonlinearities, the response to uncorrelated light was recorded and considered in the fitting procedure by multiplying the model-function with the normalized/smoothed uncorrelated instrumental response. The fits cover 99.95% of the total fluorescence intensity and start about 500 channels before the laser-pulse recorded in the instrument response function (IRF).

**FRET-Accessible Volume Calculations.** To describe the behavior of the fluorophore labels on the macromolecules accessible volume (AV), simulations were performed as described previously.<sup>44,37</sup> The accessible volume algorithm implemented in FPSv1.2 calculates all sterically accessible positions given the spatial extension of the dyes (all parameters used are compiled in Supplementary Table S1, Supporting Information). As previously shown, this approach is useful in comparing given model structures with experimental data and can also be used for the generation of new structural models.<sup>45</sup>

**Quantitative FRET Analysis: Ensemble TCSPC-Fitting and Determination of the Precision for the Obtained Distances.** Under our measurement conditions, we have a mixture of monomeric and dimeric hGBP1. Thus, all donor fluorescence decays  $F_D(t)$  were fitted with the decays of at least two molecular species,  $F_{D(0)}(t)$  for the donor-/unlabeled protein complexes (DOnly) and  $F_{D(A)}(t)$  for the donor-/acceptor-labeled protein complex which is assumed to be dimeric:

$$F(t) = (1 - x_{\text{DOnly}})F_{D(A)}(t) + x_{\text{DOnly}}F_{D(0)}(t) + c \quad (1)$$

where  $x_{\text{DOnly}}$  corresponds to the fraction of DOnly molecules and *c* is a constant offset. Due to local quenching, the fluorescence decay of the donor is triexponential in the absence of FRET with the individual species fractions  $x_D^{(i)}$  and fluorescence lifetimes  $\tau_{D(0)}^{(i)}$  (see Table S3, Supporting Information):

$$F_{D(0)}(t) = \sum_i x_D^{(i)} \exp(-t/\tau_{D(0)}^{(i)}) \quad (2)$$

Thus, the time-resolved fluorescence intensity decays of donor-/acceptor-labeled protein-complex (FRET sample) were fitted globally with the decays of the donor-/unlabeled protein-complexes (DOnly sample). Generally, it is reasonable to assume that the radiative lifetime of the donor is not affected by quenching. Hence, the FRET-rate constant ( $k_{\text{FRET}}$ ) is

actually only determined by the donor–acceptor distance and their relative orientation.<sup>46</sup> Expressing the FRET rate constant in terms of distances the donor–fluorescence in the presence of acceptor is given by

$$F_{D(A)}(t) = F_{D(0)} \int_{R_{DA}} p(R_{DA}) \exp(-tk_0(R_0/R_{DA})^6) dR_{DA} \quad (3)$$

Therein  $p(R_{DA})$  is a FRET-rate distribution expressed as distance and  $R_0$  is the Förster radius (in this case  $R_0 = 52 \text{ \AA}$ ) and  $k_0 = 1/\tau_0$  is the radiative rate of the unquenched dye.

The fluorophores are attached to the biomolecule by long flexible linkers. Hence, a donor–acceptor distance distribution is expected which is not averaged during the fluorescence lifetime of the dyes,<sup>37</sup> and the fluorescence decay  $F_{D(A)}$  has to be expressed as a donor–acceptor distance distribution  $p(R_{DA})$  with a nonzero width. Here, the experimental time-resolved fluorescence intensities were either fitted by a Gaussian distribution of donor–acceptor distances ( $p(R_{DA})$ ) with a mean interdy distance  $\langle R_{DA} \rangle$  and a width  $w_{DA}$  (eq 4) or, analog to the Tikhonov regularization,<sup>40</sup>  $p(R_{DA})$  was determined model free by deconvolution of the fluorescence intensity decays by using the maximum-entropy method (MEM).<sup>47,48</sup>

$$F_{D(A)}(t) = F_{D(0)} \int_{R_{DA}} \frac{1}{w_{DA} \sqrt{\pi/2}} \exp\left(-2 \left[ \frac{R_{DA} - \langle R_{DA} \rangle}{w_{DA}} \right]^2\right) \exp(-tk_0[1 + (R_0/R_{DA})^6]) dR_{DA} \quad (4)$$

The width of the Gaussian donor–acceptor distance distribution  $w_{DA}$  should not be misinterpreted as the experimental/statistical-error but it describes a real physical property of the donor–acceptor pair. The experimental fluorescence decays presented below are described by combining the above formulas and were fitted by custom software written in Python.

The fluorescence fractions  $f_1$  and  $f_2$  of the states described by  $p_{1,2}(R_{DA})$  are calculated by eq 5.

$$f_{1,2} = \frac{1}{\langle F \rangle} \int p_{1,2}(R_{DA}) \frac{1}{k_0} \left( 1 - \left( 1 + \left( \frac{R_{DA}}{R_0} \right)^6 \right)^{-1} \right) dR_{DA} \quad (5)$$

with  $1 = f_1 + f_2$

where  $\langle F \rangle$  is used as normalization factor for the total steady state fluorescence intensity.

The parameters and their uncertainties were determined by Markov chain Monte Carlo sampling using the Metropolis–Hasting algorithm.<sup>49,50</sup> All free fitting parameters were sampled using at least 30 individual Markov chains with 150 000 steps each. Each Markov chain started at  $\vec{w}_{\min}$ , whereas  $\vec{w}_{\min}$  was previously determined by fitting the data with a model function using a conventional Levenberg–Marquardt algorithm. In the case of Gaussian distributed noise on the signal (counts in the TAC-channels) the probability density  $P(\vec{m}|\vec{w}_i)$  of observing the measurement result  $\vec{m}$ , given a set of model parameters  $\vec{w}_i$ , is proportional to  $P(\vec{m}|\vec{w}_i) \propto \exp(-\chi_i^2(\vec{m}|\vec{w}_i)/2)$ . Therefore, the probability in the Metropolis algorithm to move from a set of model parameters  $\vec{w}_i$  to a new set of model parameters  $\vec{w}_{i+1}$  is given by

$$P(\vec{m}, \vec{w}_{i \rightarrow i+1}) = \exp(-1/2(\chi_{i+1}^2 - \chi_i^2)) \quad (6)$$

with  $\vec{w}_{i+1} = \vec{w}_i + \vec{\delta}$ . Here  $\vec{\delta}$  are Gaussian distributed random values. To reduce the parameter space the Metropolis sampling was restricted to parameters  $\vec{w}$  fulfilling  $\chi_r^2(\vec{w}) < \chi_{r,\max}^2$  with

$$\chi_{r,\max}^2(P, \vec{w}_{\min}) = \chi_{r,\min}^2 [1 + n/\nu \cdot \text{cdf}^{-1}(F(n, \nu, P))] \quad (7)$$

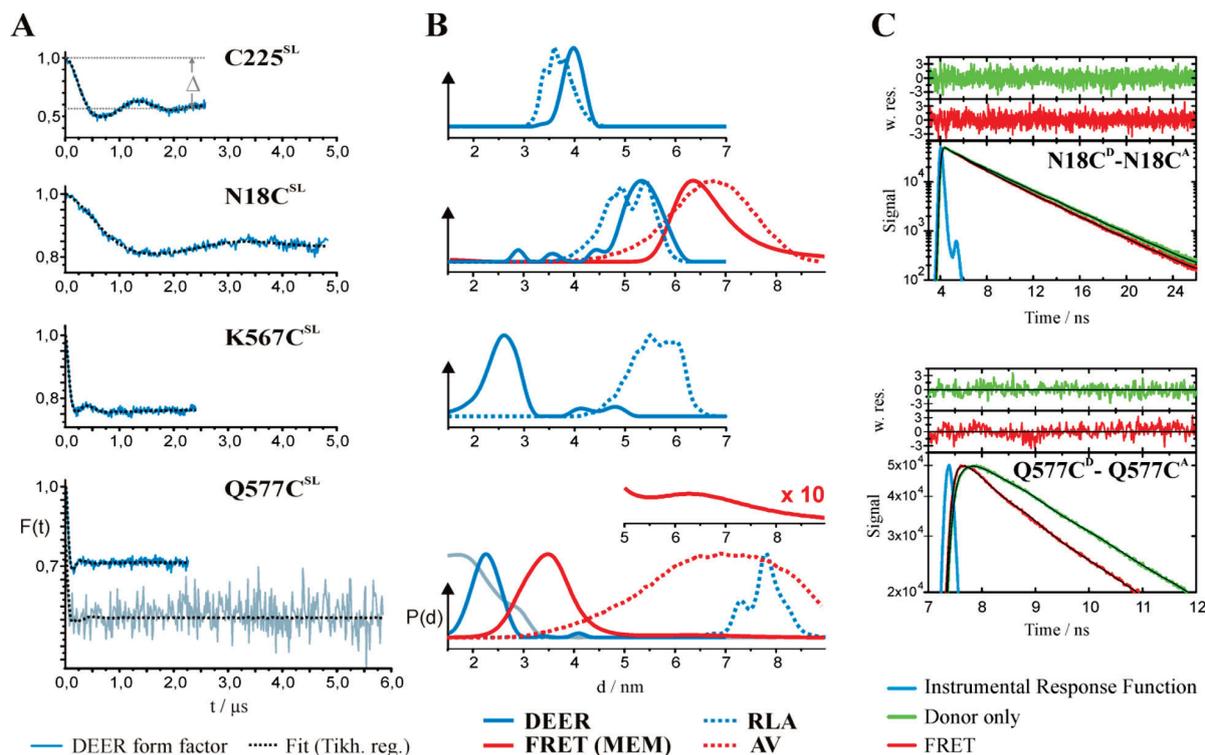
where  $\text{cdf}^{-1}(F(n, \nu, P))$  is the inverse of the cumulative distribution function of the  $F$ -distribution for  $n$  free parameters determined by the dimension of the parameter space, and with  $\nu$  degrees of freedom given by the number of fitting channels,  $\chi_{r,\min}^2$  is the minimum determined  $\chi_r^2$  (here usually  $\chi_{r,\min}^2 = \chi_{r,\min}^2(\vec{w}_{\min})$ ).<sup>51</sup> If a trial  $\vec{w}_{i+1}$  left the region of interest determined by  $\chi_{r,\max}^2$  the step was neglected and a new trial attempt starting at  $\vec{w}_i$  was performed. The presented analysis was at least performed up to the maximum confidence-level of  $P_{\text{sample,max}} = 1-10^{-5}$ . The step-size  $\vec{\delta}$  of the Metropolis-Hastings algorithm was adjusted in shorter preruns to obtain an acceptance rate of approximately 60%. After sampling the parameter space samples  $\vec{w}_i$  were selected according to their  $\chi_r^2(\vec{w}_i)$ . Samples with a  $\chi_r^2(\vec{w}_i)$  above a maximum chi-square  $\chi_{r,\max}^2(P_{\text{select,max}})$  can be excluded with a confidence level of  $P_{\text{select,max}}$ . The samples with a  $\chi_r^2(\vec{w}_i)$  below  $\chi_{r,\max}^2(P_{\text{select,max}})$  define a distribution in the parameter space. The individual fitting parameters are obtained by calculating the mean of their individual marginal distributions, whereas the marginal distributions define the parameter uncertainty.<sup>52</sup>

**Construction of the  $\alpha 13$  Dimer and Molecular Dynamics (MD) Simulations.** The atomic coordinates for helix  $\alpha 13$  (residues F565–M583) were extracted from the crystal structure of hGBP1 (pdb: 1F5N).<sup>24</sup> Two such isolated helices were manually oriented in a way that interaction between the hydrophobic faces of the two helices (comprised by residues M572, I576, L579, and M583) was possible (see Figure 4A). The helix dimer was immersed in a water box, at least 12  $\text{\AA}$  larger than the dimer in any direction, filled with TIP3P water and  $\sim 150$  mM sodium and chloride ions, neutralizing the system's net charge. Periodic boundary conditions have been applied. Energy minimization and removal of initial atomic clashes in the starting structure were removed by energy minimization (steepest descent) with the software package Yasara Structure (<http://www.yasara.org>).<sup>53</sup> After the initial energy minimization, a 100 ns MD simulation was carried out in Yasara, utilizing the Amber03 force field, using Particle Mesh Ewald (PME) summation for long-range electrostatic interactions with a cutoff at 7.86  $\text{\AA}$ . The time step for the calculation of intramolecular forces was 1.25 fs (simulation substep), intermolecular forces have been calculated every two simulation substeps (2.5 fs). The simulation temperature was 298.0 K. Temperature control was carried out by rescaling atom velocities. Pressure control was achieved by keeping the solvent ( $\text{H}_2\text{O}$ ) density at 0.997 g/mL and rescaling the simulation cell along all three axes. Simulation snapshots have been taken each 25 ps.

A second simulation has been carried out using the same starting structure at a temperature of 298.1 K, resulting in different starting velocities for the atoms.

The structure ensemble shown in Figure 4B was prepared by aligning the structures from simulation snapshots taken at 0, 5, 10, ... 100 ns of the 298.0 K trajectory with the MUSTANG algorithm<sup>54</sup> implemented in Yasara Structure.

Analysis of the MD trajectories has been carried out with Yasara. Energies and RMSD values for  $C\alpha$  atoms for the two



**Figure 2.** DEER and FRET data. (A) Background corrected DEER traces (form factors, blue) and fits (black) for C225<sup>SL</sup>, N18C<sup>SL</sup>, K567C<sup>SL</sup>, and Q577C<sup>SL</sup> with GppNHp. In the panel for hGBP1-C225<sup>SL</sup> the modulation depth  $\Delta$  is indicated (gray). (B) Distance distributions obtained by Tikhonov regularization (blue) of the DEER data and maximum-entropy deconvolution of the fluorescence decay (red). The distance distribution for Q577C shows enlarged the second population at 67 Å revealed from the fluorescence data (see text). Dashed lines represent distance distributions obtained by a RLA (blue) and accessible volume (AV) calculations<sup>32</sup> (red) on the model of the full-length dimer in Figure 1C. Abbreviations in graphical legend: MEM, maximum entropy method. (C) Experimental time-resolved donor (D) fluorescence decays (Alexa488) for mixtures with acceptor (A) labeled samples (Alexa647 FRET decay, red) or unlabeled proteins (U) (DOnly decay, green). The DOnly decays are fitted formally by either two or three lifetimes (eq 3, see Table S3, Supporting Information), the FRET decays are fitted by MEM. The weighted residuals of the donor decay and the FRET decay are displayed in the upper panel in green and red, respectively.

simulations are shown in Supplementary Figure S7, Supporting Information.

## RESULTS

Cysteine mutants, based on a cysteine-free hGBP1 variant, were labeled with the MTS spin label (MTSSL) and fluorescence labels, respectively. Four single labeled (SL) hGBP1 constructs were subjected to DEER experiments: hGBP1-N18C<sup>SL</sup>, -C225<sup>SL</sup>, -K567C<sup>SL</sup>, and -Q577C<sup>SL</sup> (Figure 1A,C). Functionality of the cysteine-free variant Cys-9 and the labeled constructs was verified by their specific GTPase activities, determined as described earlier.<sup>27</sup> All cysteine mutants and labeled proteins exhibit either similar or increased (factor 2–3) GTPase activities compared to the hGBP1 wildtype protein (hGBP1 wt, see Supplementary Table S2, Supporting Information).

First, we confirmed that dimerization of hGBP1 upon binding of GppNHp takes place via the LG domains as suggested from the crystal structure of the isolated LG domains with GppNHp<sup>22</sup> (see Figure 1B). We chose positions 225 on helix  $\alpha 4'$ , and 18 on the opposite side of the LG domain. The DEER experiments reveal no detectable interspin distances for singly labeled hGBP1 without any nucleotide or with GDP (see Supplementary Figure S1, Supporting Information). This in accordance with the protein being monomeric, whereas in the presence of GppNHp well-defined interspin distances are

obtained, indicating that the protein forms dimers. Figure 2 shows the results of the DEER experiments for hGBP1-C225<sup>SL</sup> and hGBP1-N18C<sup>SL</sup> with bound GppNHp. In Figure 2A the background corrected DEER time traces are shown. Figure 2B depicts the corresponding distance distributions obtained by Tikhonov regularization (for details see Supporting Information and Methods and Figure S1). The distance distribution for position 225 exhibits a single peak at 40 Å. Labels at position 18 on the opposite side of the LG domain show an interspin distance of 54 Å. Small peaks at shorter distances in the latter case represent noise artifacts and are not reproducible. The observed distances differ by  $\sim 5$  Å (pos. 225) and  $\sim 9$  Å (pos. 18) from the  $C\alpha$ – $C\alpha$  distances calculated from the crystal structure (Figure 1B). Nevertheless, when comparing interspin distances with structural models the length and flexibility of the spin label side chain (Supplementary Figure S2, Supporting Information) has to be taken into account. For this purpose, we performed a rotamer library analysis (RLA)<sup>42</sup> on the crystal structure of the isolated LG domain dimer with GppNHp (Figure 1B). In both cases, the calculated distance distributions (Figure 2B; blue, dotted lines) are in good agreement with the experimental data, indicating that the LG domains exhibit an orientation in the hGBP1 dimer resembling the crystal structure of the isolated LG domain dimer (Figure 1B), which has also recently been corroborated by mutational studies.<sup>54</sup> Furthermore, also the widths of the experimental and

RLA calculated distance distributions agree reasonably well. The quality of the data sets (signal-to-noise) and application of the L-curve criterion in the regularization procedure (see Supplementary Experimental Procedures, Supporting Information) give confidence to assume sufficiently high accuracy in prediction of the experimental distance widths for this comparison. This indicates that mainly the mobility of the spin label side chain contributes to the distribution of inter spin distances, as the RLA does not account for protein backbone flexibility. Consequently, the secondary structure elements the labels are attached to are relatively rigid, underlining the stability of the LG domain dimer.

With the given arrangement of the LG domains in the hGBP1 dimer, it is straightforward to construct a model of the full-length dimer by using rigid monomers and superimposing the LG domains of the full-length crystal structure in the GppNHp bound state<sup>24</sup> onto the LG domain dimer structure<sup>23</sup> (Figure 1C). Performing the RLA on this model for positions 567 and 577 reveals calculated distance distributions centered at  $\sim 57$  Å and  $\sim 78$  Å, respectively (Figure 2B; blue, dotted lines). Surprisingly, the experimental distance distributions (Figure 2B; blue, solid lines) reveal significantly shorter distances, 26 Å for position 567 and 22 Å for position 577. This can only be achieved if  $\alpha 12$  and  $\alpha 13$  detach from the LG domain and the two  $\alpha 13$  helices in the dimer come into close vicinity or associate. Such kind of detachment has already been proposed, first based on the observation that in the presence of GDP·AlF<sub>x</sub> (i.e., in the course of GTP hydrolysis) helix  $\alpha 4'$  in the LG domain undergoes a conformational change leading to a steric clash with  $\alpha 12/13$ , and second based on point mutations distorting the contact site between the LG domain and  $\alpha 12/13$ .<sup>23,27</sup> Our results indicate that the uncoupling between  $\alpha 4'$  and  $\alpha 12/13$  occurs already upon dimer formation in the presence of GppNHp.

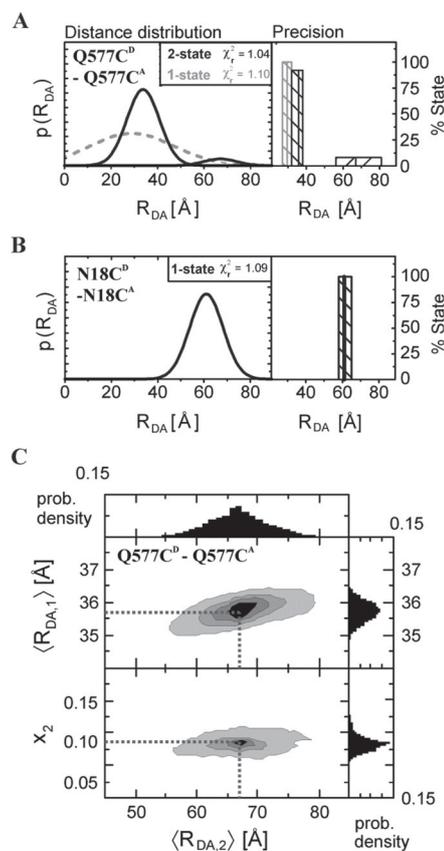
The DEER experiment not only yields a distance distribution but also information about the number of interacting spins,<sup>40</sup> reflected in the modulation depth  $\Delta$  indicated (gray) for the DEER trace for hGBP1-C225<sup>SL</sup>. The modulation depth for position 225 corresponds to two interacting nitroxides and consequently >90% dimerization. For hGBP1 spin labeled at positions 18, 567, and 577, respectively, 50%, 64%, and 75% of the proteins show dipolar interaction between the spin labels (the absence of distances <17 Å has been confirmed by cw EPR; see Supplementary Figure S1B, Supporting Information). From protein concentration-dependent GTP hydrolysis activities, an apparent dissociation constant for the hGBP1 dimer is found to be sub-micromolar.<sup>55</sup> Yet, in the case of GppNHp-bound hGBP1 our unpublished data indicate a  $K_d$  value around 10  $\mu$ M. Thus, small differences in the dimer affinities of the four hGBP1 variants together with experimental uncertainties on, e.g., protein concentrations and modulation depths, can explain the variation in the observed dimer populations.

DEER experiments are carried out in frozen solution (50 K), raising the question whether the observed association of the  $\alpha 13$  helices takes also place at more physiological temperatures. Quantitative FRET distance measurements by eTCSPC of hGBP1 labeled with fluorescent dyes at positions 18 or 577 at room temperature and comparison of the experimental results to the expectancies given by the structural models corroborated the observation made by DEER.

To study intermolecular FRET, we mixed protein singly labeled with the donor dye Alexa488 (0.5  $\mu$ M) and the acceptor

dye Alexa647 (7.1  $\mu$ M, 5.1  $\mu$ M for 577 and 18 respectively). The actual protein concentrations were calculated based on the degree of labeling and the protein concentration determined by UV–vis spectroscopy. The quantitative analysis was performed in two steps: (a) Data analysis using model free and model based approaches combined with accessible volume calculations to determine absolute donor–acceptor distances, (b) determination of the precision and accuracy of the obtained inter dye distances.

To obtain distance distributions in the first step, the fluorescence decay data were analyzed by the model-free method MEM (see Figures 2 and 3), which is equivalent to the Tikhonov regularization in EPR analysis. In the N18C<sup>D</sup>-N18C<sup>A</sup> sample, an asymmetric peak at position  $\sim 63$  Å tailing toward longer distances is observed. The tailing is most likely caused by



**Figure 3.** FRET-parameter and error-estimation using Gaussian donor–acceptor distance distributions. (A, B) donor–acceptor distance distributions  $p(R_{DA})$  in dependency of the applied fitting model (eq 4). The experimental data was either fitted with a single-state model (gray) or a two-state model (black). By fitting the data with a single-state model a mean-donor–acceptor distance  $\langle R_{DA} \rangle$  of  $30 \pm 3$  Å is obtained with a width of  $w_{DA} = 36.2 \pm 6$  Å ( $\chi^2_r = 1.10$ , confidence-level of 95%). The fit of the data by a two-state model yields the distances  $\langle R_{DA,1} \rangle = 33 \pm 3$  Å and  $\langle R_{DA,2} \rangle = 69 \pm 3$  Å with a global donor–acceptor distribution width  $w_{DA} = 15.4$  Å ( $\chi^2_r = 1.04$ , confidence-level of 95%). The precision of the fits demonstrates that the results are independent of the applied fitting model and that both distances are well resolved. The fraction of the  $\langle R_{DA,2} \rangle$  population is  $8 \pm 3\%$ . (C) Projections of the probability density of the 2-state model (marginal distributions) of the model parameters  $\langle R_{DA,1} \rangle$ ,  $\langle R_{DA,2} \rangle$ , and  $x_2$  for a confidence-level 95%. All fit results are summarized in Supplementary Table S3, Supporting Information.

**Table 1. Experimental DEER and FRET Mean Distances Compared to Mean Distances Calculated by the Rotamer Library Approach (RLA) for DEER Data and Accessible Volume (AV) Calculations for FRET Data from the Initial Model ( $2 \times 1F5N$ , Head-to-Head) and from the  $\alpha 13$  Dimer (Helix–Helix) Model<sup>a</sup>**

technique	sample	experiment		model			
		distance		1F5N		helix–helix	
		mean	$w$	mean	$w$	mean	$w$
EPR <sup>b</sup>	C225C <sup>SL</sup>	40 ± 1	4 ± 2	37	8		
	N18C <sup>SL</sup>	54 ± 1	10 ± 2	50	12		
	K567C <sup>SL</sup>	26 ± 1	8 ± 2	57	12	20	9
	Q577C <sup>SL</sup>	22 ± 1	7 ± 2	78	10	20	12
FRET <sup>c</sup>	N18C <sup>D</sup> -N18C <sup>A</sup>	63 ± 9	14 ± 4	67	16		
	Q577C <sup>D</sup> -Q577C <sup>A</sup> state 1	35 ± 2	16 ± 4	68	26	32	16
	Q577C <sup>D</sup> -Q577C <sup>A</sup> state 2	67 ± 12	18 ± 4				

<sup>a</sup>All distances are given in Å. <sup>b</sup>The EPR-model distances are based on the RLA approach. <sup>c</sup>FRET-model distances are based on the AV-approach.

the presence of donor-dyes in absence of acceptor (monomeric proteins). In the Q577C<sup>D</sup>-Q577C<sup>A</sup> sample, two distinct well separated peaks at ~35 Å (peak 1) and ~67 Å (peak 2) are observed. The widths of all peaks determined by MEM are comparable and range from 14 to 18 Å. These widths are within the expectance values for flexible coupled dyes.<sup>37</sup> Compared to model-based approaches MEM has some drawbacks. For instance, the peak width is to some extent determined by the choice of the regularization parameter and depends on the noise level of the data.

To ensure that the second low amplitude peak (~10%) is justified (signal-to-noise) the L-curve criterion was applied to determine the regularization parameter (see Supplementary Experimental Procedures and Figure S5A–D). Additionally the data were analyzed by model functions where the donor–acceptor distances are Gaussian distributed (see eq 4). The use of simple Gaussians to describe the donor–acceptor distance distributions is justified by the comparison of Gaussian with identical mean and width with distributions calculated by the AV approach and is supported by the maximum entropy method (see Figure 2 and S6 in Supporting Information). The fit of the Q577C<sup>D</sup>-Q577C<sup>A</sup> sample using a one-state/distance model results in a mean inter dye distance  $\langle R_{DA} \rangle = 29$  Å and an unusual broad distance distribution with a width of  $w_{DA} = 36.2$  Å (see Figure 3A). This width cannot be explained by dye-linker distributions (see Supplemental Table S1, Supporting Information). The broadening of the fluorescence signal by dye-linker distributions not averaging out during the fluorescence lifetime of the dye usually only contributes up to a width from 12 to 22 Å.<sup>37</sup> Besides the unusual broad width, a strong correlation of the fitted width and the donor–acceptor distance was observed (see Supplementary Figure S8B, Supporting Information). Therefore, the experimental data (Q577C<sup>D</sup>-Q577C<sup>A</sup>) was fitted using two Gaussians with a global width  $w_{DA}$  and the species fractions  $x_i$ . Two populations with mean inter dye distances  $\langle R_{DA} \rangle$  of 36 Å ( $x_1 = 92\%$ ) and 67 Å ( $x_2 = 8\%$ ) and a global width of  $w_{DA} = 15.4$  Å were obtained (see Figure 3A). This high sensitivity of fluorescence spectroscopy to minor populations with a high fluorescence quantum yield is given by the fact that the quality of the fit is not weighted by the species fraction of each state  $x_i$  but by its fluorescence fraction  $f_i$  (see eq 5). For Q577C<sup>D</sup>-Q577C<sup>A</sup>, the fluorescence fraction of the population with the long inter dye distance is significant ( $f_2 = 0.32$ ) so that it is clearly detectable.

The same analysis was performed also on the data of the N18C<sup>D</sup>-N18C<sup>A</sup> sample (Figure 3B). In the sample N18C<sup>D</sup>-

N18C<sup>A</sup>, no additional broadening beyond the expected dye-linker distribution was observed. Only one Gaussian distributed FRET population ( $\langle R_{DA} \rangle = 63 \pm 9$  Å,  $w_{DA} = 14 \pm 4$  Å) is needed to describe the decay satisfactorily which agrees well with the MEM analysis in Figure 2.

In the second step of the data analysis, we estimated the precision of the parameter by a Markov chain Monte Carlo sampling as described in the method section. The obtained results are presented in Figure 3C in form of two two-dimensional histograms. These histograms represent possible realizations of the model parameters given a confidence level of  $P_{select,max} = 95$ , where the uncertainties correspond to the width of the bar plots in Figure 3A,B.

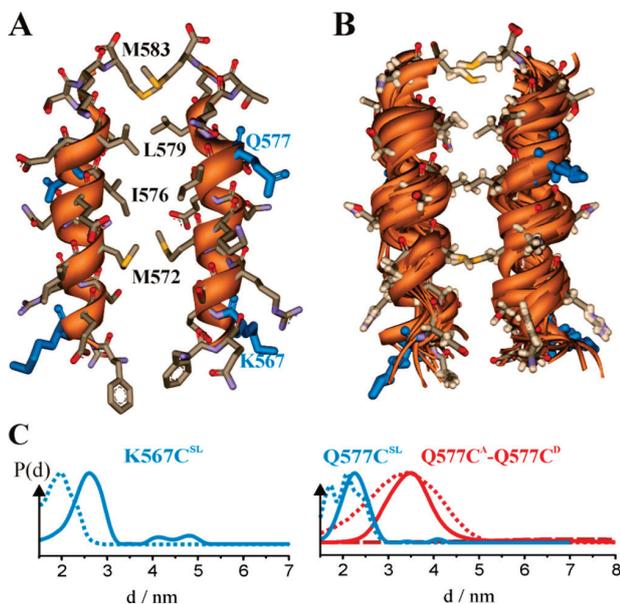
The accuracy of the inter dye distances is mainly governed by uncertainties of the dye orientation factor  $\kappa^2$ . Independent single-molecule measurements of donor and the acceptor anisotropy (data not shown) show that  $\kappa^2$  only contributes only by 6% percent to the total uncertainty.<sup>55</sup> Even though the geometric accessible volume approach is a rough approximation of the dye probability distribution, it proved to be a very good estimate in previous studies.<sup>37,44</sup> The results of the accessible volume (AV) simulations<sup>37</sup> (Figure 2B, Supplementary Figure S6, Supporting Information) reveal good agreement for position 18, corroborating the relative LG/LG domain orientation in the dimer, but a strong deviation for the major population in position 577 with the experimental distance being 33 Å shorter compared to the simulation results and well above the estimated errors (Figure 2). To conclude, the precision of the fits demonstrates that the results are independent of the applied fitting model (MEM or Gaussian distance distributions) and both the distances are well resolved.

Consequently, the results of the FRET and DEER measurements which are summarized in Table 1 indicate, within the given experimental errors, the same type of structural alterations. Hence, we conclude that the major mode of dimerization via the LG domains and the conformational change leading to association of the  $\alpha 13$  helices prevails also at physiological temperatures.

Being confident that the observed close association of the  $\alpha 13$  helices in the hGBP1 dimer is of physiological relevance, we constructed a dimer of two such helices and confirmed by molecular dynamics (MD) simulations that they can form a stable dimer.

Inspection of the primary sequence reveals that one surface of  $\alpha 13$  comprises hydrophobic residues (M572, I576, L579, and M583) that could stabilize a parallel helix dimer by van der

Waals interactions. We manually constructed a model with two  $\alpha 13$  helices (F565-M583) being in contact via the above-mentioned residues in a parallel orientation (Figure 4A), and performed, after an initial energy minimization, MD simulations in explicit water at 298 K (see Supporting Information).



**Figure 4.**  $\alpha 13$  dimer MD simulation. (A) Initial dimer structure after energy minimization. (B) Ensemble of structures from the MD trajectory, taken at simulation times of 0, 5, 10, ..., 100 ns (ribbons) and average structure (ribbon + side chains) calculated from the MD trajectory. (C) Experimental DEER (blue, solid) vs calculated (RLA, blue, dotted) distance distributions from the average dimer structure for  $K567C^{SL}$  and  $Q577C^{SL}$  ( $K567$  and  $Q577$  are shown in blue in panels a and b) and corresponding FRET data for  $Q577$  (red, solid: experiment (MEM), dotted: calculated (AV)).

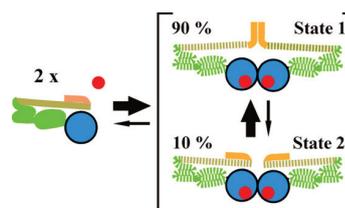
The two  $\alpha 13$  helices remained reproducibly associated over the whole trajectory length of 100 ns, exhibiting only small fluctuations of the overall dimer arrangement, although the simulations were carried out without any constraints on the atom positions. The persistence of the isolated  $\alpha 13$  dimer during the simulations indicates that interaction via this helix could significantly contribute to stabilization of the hGBP1 dimer. Closer inspection of the MD results revealed that interaction between the two helices is mainly conferred by the four hydrophobic side chains mentioned above and depicted in Figure 4A and Supplementary Figure S7, Supporting Information.

Figure 4B shows an ensemble of 21 structures from the MD simulation and an average structure calculated from the MD trajectory. A RLA performed on the latter structure, shown in Figure 4C, yields a calculated distance for  $K567C^{SL}$  being  $\sim 7$  Å shorter compared to the experimentally determined one. A more distant location of the N-terminal ends of the helices than in our  $\alpha 13$  dimer model might be caused by their connection to  $\alpha 12$ , thereby creating additional constraints on the N-termini of the  $\alpha 13$  helices that prevent their close interaction. Nevertheless, an almost perfect match of the experimental and the calculated inter spin distances is observed for  $Q577C^{SL}$ . Remarkably, here not only the experimental mean distances

but also the shape of the distance distribution agrees almost perfectly with the calculations for our  $\alpha 13$  dimer model. The calculated distance widths are largely comparable with the experimental distance distributions being broader toward shorter distances. The DEER and FRET results are also consistent for position 577 (Figure 4C, red). Analogous to our interpretation in the case of the LG domain label positions, we suggest the  $\alpha 13$  dimer to be rather stable and rigid, corroborating the observed stability of the model in the MD simulations.

## DISCUSSION

Dimerization of hGBP1 is triggered by GTP binding and small conformational changes in the LG domain interface. GTP binding alters the conformation of the guanine cap, exposing residues R240 and R244 in a way that a LG domain dimer is formed.<sup>22,54</sup> In another work a buried, hydrophobic helix ( $\alpha 6$ , P291–S306) in the connecting region between the LG and the middle domain is reported to become exposed upon GTP binding and suggested to mediate dimerization.<sup>56</sup> Using a combination of pulsed EPR and fluorescence spectroscopy based inter label distance measurements with model free analysis approaches for both data sets, we find actually two conformers of the hGBP1 dimer (see Figure 5). For the majority



**Figure 5.** Mechanistic insights—the oligomerization mechanism of the hGBP1 dimer—dashed parts are currently unknown. Blue, LG domain; green, middle-domain; yellow, helix  $\alpha 12$ ; orange, helix  $\alpha 13$ ; red spheres, nucleotide (GTP/GppNHp). The structure of the protein in the apo-form is known and was already solved by X-ray crystallography. Binding of GTP/GppNHp leads to the formation of a putative head-to-head dimer. The protein structure of the nucleotide bound state prior oligomerization is unknown. Experimental data indicate the presence of an equilibrium between dimers with helix  $\alpha 13$  contacts ( $\sim 90\%$ ) and the simple putative head-to-head dimers ( $\sim 10\%$ ). Our experimental data do not allow assumptions on the relative orientation of the middle domain and  $\alpha 12$ ; therefore, these protein regions are depicted with higher transparency.

( $\sim 90\%$ ) populated state 1, we identify a new conformer with another dimerization interface which is formed by the C-terminal helix  $\alpha 13$ . Moreover, we find a low ( $\sim 10\%$ ) populated state 2 characterized by long distances (67 Å for position 577 from the FRET measurements,  $\sim 40$ –50 Å for  $K567C^{SL}$  from DEER) that is well described by the known dimer structure (helix  $\alpha 13$  at large distances). We propose that  $\alpha 13$  detaches from  $\alpha 12$  enabling it to establish additional contacts for dimerization together with  $\alpha 13$  from the other hGBP1 molecule. Using a flexible assembly of four rigid domains (LG, middle,  $\alpha 12$  and  $\alpha 13$ ) for each molecule, we propose a model for the dimer as illustrated in Figure 5 taking into account all experimental observations of this study. Investigations to clarify the dimerization process as well as conformational changes of the middle domain and helix  $\alpha 12$  in the hGBP1 dimer are subject of ongoing research in our groups.

Our measurements and simulations suggest that the binding of GppNHp leads to a dimerization of helices  $\alpha 13$  (Figure 4). This association brings two lipid modifications into close vicinity as hGBP1 possesses a “CaaX” motif at the end of helix  $\alpha 13$  that is farnesylated *in vivo*.<sup>57,58</sup> In addition, a polybasic sequence directly adjacent to the CaaX box (<sup>582</sup>KMRRRK<sup>587</sup>) might further increase membrane affinity.<sup>59,60</sup> Therefore, membrane association, that has already been proposed to be regulated by dimerization,<sup>27,59</sup> might be directly controlled by association of the  $\alpha 13$  helices. Recent studies revealed that also hGBP2 and hGBP5 are isoprenylated *in vivo* and that they can form not only homo- but also heterodimers,<sup>59</sup> suggesting that association of the C-terminal domains might be a general feature of this subclass of GBPs with substantial importance for membrane localization and physiological function.

Finally, we would like to reflect the mechanism of dimer or oligomer formation for members within the dynamin superfamily. It was shown earlier that the LG domain of hGBP1 is essential to form homo dimers and to increase its GTPase activity.<sup>23</sup> Together with our observation of the additional interaction of the  $\alpha$ -helical C terminus as illustrated in Figure 4, we find a striking similarity to the dimer structure of bacterial dynamin-like protein (BDLP).<sup>61</sup> It also forms contacts between the G domains, and in addition, between the  $\alpha$ -helical, so-called paddle regions at the C-terminus which can be regarded to be homologous to  $\alpha 13$  of hGBP1. This close relationship is also reflected in the phylogenetic analysis of the dynamin family members carried out by Low and Löwe on the basis of the LG domain sequences.<sup>61</sup> In contrast, other members of the family of large GTPases form homo-oligomers by a different mechanism. For example, the closely related dynamin and MxA are known to be associated through the helical part of the enzyme in the first place, and LG domain contacts come into play only later.<sup>62–64</sup> Future studies will have to answer the question if and how the two mechanisms of oligomer formation described here are governed by the evolutionary origin of the LG domain.

## ■ ASSOCIATED CONTENT

### ● Supporting Information

Detailed results of the DEER and FRET data analysis, RLA, AV, and MD simulations. This material is available free of charge via the Internet at <http://pubs.acs.org>

## ■ AUTHOR INFORMATION

### Corresponding Authors

\*(C.A.M.S.) E-mail: [cseidel@hhu.de](mailto:cseidel@hhu.de); tel: +49 211 81 15881; fax: +49 211 81 12803.

\*(C.H.) E-mail: [chr.herrmann@rub.de](mailto:chr.herrmann@rub.de); tel: +49 234 32 24173; fax: +49 234 32 14785.

\*(J.P.K.) E-mail: [jkclare@uos.de](mailto:jkclare@uos.de); tel: +49 541 969 2664; fax: +49 541 969 2656.

### Present Address

# (T.V.) Physical Chemistry II, Faculty of Chemistry and Biochemistry, Ruhr-University Bochum, Universitätsstr. 150, 44780 Bochum, Germany.

### Author Contributions

§T.V., C.S.H., and T.O.P. contributed equally.

### Funding

Parts of this work were supported by DFG Grant KL2077/1-1 and HE2679/6-1.

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank H.-J. Steinhoff for providing the EPR equipment, fruitful discussions, and reading of the manuscript. Parts of this work were supported by DFG Grant KL2077/1-1 and HE2679/6-1.

## ■ ABBREVIATIONS

DEER, double electron–electron resonance; EPR, electron paramagnetic resonance; eTCSPC, ensemble time-correlated single-photon counting; FRET, Förster resonance energy transfer; GppNHp, guanosine 5'-imidotriphosphate; hGBP1, human guanylate binding protein 1; MD, molecular dynamics; MTSSL, (1-oxyl-2,2,5,5-tetramethyl-pyrrolinyl-3-methyl) methanethiosulfonate spin label; PELDOR, pulse electron double resonance; RLA, rotamer library analysis; SDSL, site-directed spin labeling

## ■ REFERENCES

- (1) Praefcke, G. J. K., and McMahon, H. T. (2004) The dynamin superfamily: universal membrane tubulation and fission molecules? *Nat. Rev. Mol. Cell. Biol.* 5, 133–147.
- (2) Cheng, Y.-S., Colonna, R. J., and Yin, F. H. (1983) Interferon Induction of Fibroblast Proteins with Guanylate Binding Activity. *J. Biol. Chem.* 258, 7746–7750.
- (3) Anderson, S. L., Carton, J. M., Lou, J., Xing, L., and Rubin, B. Y. (1999) Interferon-Induced Guanylate Binding Protein-1 (GBP-1) Mediates an Antiviral Effect against Vesicular Stomatitis Virus and Encephalomyocarditis Virus. *Virology* 256, 8–14.
- (4) Itsui, Y., Sakamoto, N., Kanikuma, S., Nakagawa, M., Sekine-Osajima, Y., Tasaka-Fujita, M., Nishimura-Sakurai, Y., Suda, G., Karakama, Y., Mishima, K., Yamamoto, M., Watanabe, T., Ueyama, M., Funaoka, Y., Sazuma, S., and Watanabe, M. (2009) Antiviral effects of the interferon-induced protein guanylate binding protein 1 and its interaction with the hepatitis C virus NSSB protein. *Hepatology* 50, 1727–1737.
- (5) Naschberger, E., Lubeseder-Martellato, C., Meyer, N., Gessner, R., Kremmer, E., Gessner, A., and Stürzl, M. (2006) Human Guanylate Binding Protein-1 Is a Secreted GTPase Present in Increased Concentrations in the Cerebrospinal Fluid of Patients with Bacterial Meningitis. *Am. J. Pathol.* 169, 1088–1099.
- (6) Tietzel, I., El-Haibi, C., and Carabeo, R. A. (2009) Human Guanylate Binding Proteins Potentiate the Anti-Chlamydia Effects of Interferon- $\gamma$ . *PLoS One* 4, e6499.
- (7) Zhu, Z., Shi, Z., Yan, W., Wei, J., Shao, D., Deng, X., Wang, S., Li, B., Tong, G., and Ma, Z. (2013) Nonstructural Protein 1 of Influenza A Virus Interacts with Human Guanylate-Binding Protein 1 to Antagonize Antiviral Activity. *PLoS One* 8, e55920.
- (8) Selleck, E., Fentress, Sarah J., Beatty, W. L., Degrandi, D., Pfeffer, K., Virgin, H. W., MacMicking, J. D., and Sibley, L. D. (2013) Guanylate-binding Protein 1 (Gbp1) Contributes to Cell-autonomous Immunity against *Toxoplasma gondii*. *PLoS Pathogens* 9, e1003320.
- (9) Guenzi, E., Töpolt, K., Lubeseder-Martellato, C., Naschberger, E., Benilli, R., Albini, A., and Stürzl, M. (2003) The guanylate binding protein-1 GTPase controls the invasive and angiogenic capability of endothelial cells through inhibition of MMP-1 expression. *EMBO J.* 22, 3772–3782.
- (10) Weinländer, K., Naschberger, E., Lehmann, M. H., Tripal, P., Paster, W., Stockinger, H., Hohenadl, C., and Stürzl, M. (2008) Guanylate binding protein-1 inhibits spreading and migration of endothelial cells through induction of integrin  $\alpha 4$  expression. *FASEB J.* 22, 4168–4178.
- (11) Naschberger, E., Croner, R. S., Merkel, S., Dimmler, A., Tripal, P., Amann, K. U., Kremmer, E., Brueckl, W. M., Papadopoulos, T., Hohenadl, C., Hohenberger, W., and Stürzl, M. (2008) Angiostatic

immune reaction in colorectal carcinoma: Impact on survival and perspectives for antiangiogenic therapy. *Int. J. Cancer* 123, 2120–2129.

(12) Li, M., Mukasa, A., del-Mar Inda, M., Zhang, J., Chin, L., Cavenee, W., and Furnari, F. (2011) Guanylate binding protein 1 is a novel effector of EGFR-driven invasion in glioblastoma. *J. Exp. Med.* 208, 2657–2673.

(13) Yu, C. J., Chang, K. P., Chang, Y. J., Hsu, C. W., Liang, Y., Yu, J. S., Chi, L. M., Chang, Y. S., and Wu, C. C. (2011) Identification of Guanylate-Binding Protein 1 as a Potential Oral Cancer Marker Involved in Cell Invasion Using Omics-Based Analysis. *J. Proteome Res.* 10, 3778–3788.

(14) Lipnik, K., Naschberger, E., Gonin-Laurent, N., Kodajova, P., Petznek, H., Rungaldier, S., Asigiano, S., Ferrini, S., Stürzl, M., and Hohenadl, C. (2010) Interferon  $\gamma$ -Induced Human Guanylate Binding Protein 1 Inhibits Mammary Tumor Growth in Mice. *Mol. Med.* 16, 177–187.

(15) Desarnaud, F., Geck, P., Parkin, C., Carpinito, G., and Makarovskiy, A. N. (2011) Gene expression profiling of the androgen independent prostate cancer cells demonstrates complex mechanisms mediating resistance to docetaxel. *Cancer Biol. Ther.* 11, 204–212.

(16) Duan, Z., Foster, R., Brakora, K. A., Yusuf, R. Z., and Seiden, M. V. (2006) GBP1 overexpression is associated with a paclitaxel resistance phenotype. *Cancer Chemother. Pharmacol.* 57, 25–33.

(17) Prakash, B., Praefcke, G. J. K., Renault, L., Wittinghofer, A., and Herrmann, C. (2000) Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins. *Nature* 403, 567–571.

(18) Cheng, Y.-S., Becker-Manley, M. F., Chow, T. P., and Horan, D. C. J. (1985) Affinity purification of an interferon-induced human guanylate-binding protein and its characterization. *Biol. Chem.* 260, 15834–15839.

(19) Praefcke, G. J. K., Geyer, M., Schwemmler, M., Kalbitzer, H. R., and Herrmann, C. (1999) Nucleotide-binding characteristics of human guanylate-binding protein 1 (hGBP1) and identification of the third GTP-binding motif. *J. Mol. Biol.* 292, 321–332.

(20) Praefcke, G. J. K., Kloep, S., Benschaid, U., Lilie, H., Prakash, B., and Herrmann, C. (2004) Identification of Residues in the Human Guanylate-binding Protein 1 Critical for Nucleotide Binding and Cooperative GTP Hydrolysis. *J. Mol. Biol.* 344, 257–269.

(21) Cheng, Y.-S., Patterson, C. E., and Staeheli, P. (1991) Interferon-induced guanylate-binding proteins lack an N(T)KXD consensus motif and bind GMP in addition to GDP and GTP. *Mol. Cell. Biol.* 11, 4717–4725.

(22) Kunzelmann, S., Praefcke, G. J. K., and Herrmann, C. (2006) Transient Kinetic Investigation of GTP Hydrolysis Catalyzed by Interferon-g-induced hGBP1 (Human Guanylate Binding Protein 1). *J. Biol. Chem.* 281, 28627–28635.

(23) Ghosh, A., Praefcke, G. J. K., Renault, L., Wittinghofer, A., and Herrmann, C. (2006) How guanylate-binding proteins achieve assembly-stimulated processive cleavage of GTP to GMP. *Nature* 440, 101–104.

(24) Prakash, B., Renault, L., Praefcke, G. J. K., Herrmann, C., and Wittinghofer, A. (2000) Triphosphate structure of guanylate-binding protein 1 and implications for nucleotide binding and GTPase mechanism. *EMBO J.* 19, 4555–4564.

(25) Kunzelmann, S., Praefcke, G. J. K., and Herrmann, C. (2005) Nucleotide Binding and Self-Stimulated GTPase Activity of Human Guanylate-Binding Protein 1 (hGBP1). *Methods Enzymol.* 404, 512–527.

(26) Gasper, R., Meyer, S., Gotthardt, K., Sirajuddin, M., and Wittinghofer, A. (2009) It takes two to tango: regulation of G proteins by dimerization. *Nat. Rev. Mol. Cell. Biol.* 10, 423–429.

(27) Vöpel, T., Syguda, A., Britzen-Laurent, N., Kunzelmann, S., Lüdemann, M. B., Dovengerds, C., Stürzl, M., and Herrmann, C. (2010) Mechanism of GTPase activity induced self-assembly of human guanylate binding protein 1. *J. Mol. Biol.* 400, 63–70.

(28) Vöpel, T., Kunzelmann, S., and Herrmann, C. (2009) Nucleotide dependent cysteine reactivity of hGBP1 uncovers a

domain movement during GTP hydrolysis. *FEBS Lett.* 583, 1923–1927.

(29) Syguda, A., Bauer, M., Benschaid, U., Ostler, N., Naschberger, E., Ince, S., Stürzl, M., and Herrmann, C. (2012) Tetramerisation of human guanylate binding protein 1 is mediated by coiled coil formation of the C-terminal  $\alpha$ -helices. *FEBS J.* 279, 2544–2554.

(30) Martin, R. E., Pannier, M., Diederich, F., Gramlich, V., Hubrich, M., and Spiess, H. W. (1998) Determination of End-to-End Distances in a Series of TEMPO Diradicals of up to 2.8 nm Length with a New Four-Pulse Double Electron Electron Resonance Experiment. *Angew. Chem. Int. Ed.* 37, 2833–2837.

(31) Pannier, M., Veit, S., Godt, A., Jeschke, G., and Spiess, H. W. (2000) Dead-Time Free Measurement of Dipole-Dipole Interactions between Electron Spins. *J. Magn. Reson.* 142, 331–340.

(32) Klare, J. P., and Steinhoff, H.-J. (2009) Spin Labeling EPR. *Photosynth. Res.* 102, 377–390.

(33) Meyer, S., Böhme, S., Krüger, A., Steinhoff, H.-J., Klare, J. P., and Wittinghofer, A. (2009) Kissing G domains of MnmE monitored by X-ray crystallography and pulse EPR Spectroscopy. *PLoS Biol.* 7, e1000212.

(34) Böhme, S., Meyer, S., Krüger, A., Steinhoff, H.-J., Wittinghofer, A., and Klare, J. P. (2010) Stabilization of G domain conformations in the tRNA modifying MnmE/GidA complex observed with DEER spectroscopy. *J. Biol. Chem.* 285, 16991–17000.

(35) Kravets, E., Degrandi, D., Weidtkamp-Peters, S., Konermann, C., Felekyan, S., Darkazanli, J. M., Praefcke, G. J., Seidel, C. A., Schmitt, L., Smits, S. H., and Pfeffer, K. J. (2012) The GTPase activity of murine guanylate-binding protein 2 (mGBP2) controls the intracellular localization and recruitment to the parasitophorous vacuole of *Toxoplasma gondii*. *J. Biol. Chem.* 287, 27452–27466.

(36) Rothwell, P. J., Berger, S., Kensh, O., Felekyan, S., Antonik, M., Wöhr, B. M., Restle, T., Goody, R. S., and Seidel, C. A. M. (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.

(37) Sindbert, S., Kalinin, S., Nguyen, H., Kienzler, A., Bannwarth, W., Appel, B., Müller, S., and Seidel, C. A. M. (2011) Accurate Distance Determination of Nucleic Acids via Förster Resonance Energy Transfer: Implications of Dye Linker Length and Rigidity. *J. Am. Chem. Soc.* 133, 2463–2480.

(38) Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182, 319–326.

(39) Altenbach, C., <https://sites.google.com/site/altenbach/labview-programs/epr-programs/short-distances>.

(40) Jeschke, G., Chechik, V., Ionita, P., Godt, A., Zimmermann, H., Banham, J. E., Timmel, C. R., Hilger, D., and Jung, H. (2006) DeerAnalysis2006 - a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.* 30, 473–498.

(41) Tikhonov, A. N., and Arsenin, V. Y. (1977) *Solutions of Ill-Posed Problems*, Wiley, New York.

(42) Polyhach, Y., Bordignon, E., and Jeschke, G. (2011) Rotamer libraries of spin labelled cysteines for protein studies. *Phys. Chem. Chem. Phys.* 13, 2356–2366.

(43) O'Connor, D. V., and Phillips, D. (1984) *Time-Related Single Photon Counting*, Academic Press, New York.

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

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

(46) Kalinin, S., and Johansson, L. B. (2004) Energy Migration and Transfer Rates Are Invariant to Modeling the Fluorescence Relaxation by Discrete and Continuous Distributions of Lifetimes. *J. Phys. Chem. B* 108, 3092–3097.

(47) Livesey, A. K., and Skilling, J. (1985) Maximum Entropy Theory. *Acta Crystallogr., Sect. A* 41, 113–122.

- (48) Brochon, J. C. (1994) Maximum entropy method of data analysis in time-resolved spectroscopy. *Methods Enzymol.* 240, 262–311.
- (49) Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953) Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* 21, 1087–1092 S.
- (50) Hastings, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109 S.
- (51) Lakowicz, J. R. (2006) *Principles of Fluorescence Spectroscopy*, 3rd ed., Springer, New York.
- (52) Gregory, P. (2005) *Bayesian Logical Data Analysis for the Physical Sciences*, Cambridge University Press, New York.
- (53) Krieger, E., Koraimann, G., and Vriend, G. (2002) Increasing the precision of comparative models with YASARA NOVA - a self-parameterizing force field. *Proteins* 47, 393–402.
- (54) Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J., and Lesk, A. M. (2006) MUSTANG: A multiple structural alignment algorithm. *Proteins* 64, 559–574.
- (55) Wehner, M., Kunzelmann, S., and Herrmann, C. (2012) The guanine cap of hGBP1 is responsible for dimerization and self-activation of GTP hydrolysis. *FEBS J.* 279, 203–210.
- (56) Dale, R. E., Eisinger, J., and Blumberg, W. E. (1979) The orientational freedom of molecular probes orientation factor in intramolecular energy transfer. *Biophys. J.* 26, 161–193.
- (57) Abdullah, N., Balakumari, M., and Sau, A. K. (2010) Dimerization and Its Role in GMP Formation by Human Guanylate Binding Proteins. *Biophys. J.* 99, 2235–2244.
- (58) Nantais, D. E., Schwemmler, M., Stickney, J. T., Vestal, D. J., and Buss, J. E. (1996) Prenylation of an interferon-gamma-induced GTP-binding protein: the human guanylate binding protein, huGBP1. *J. Leukoc. Biol.* 60, 423–431.
- (59) Britzen-Laurent, N., Bauer, M., Berton, V., Fischer, N., Syguda, A., Reipschläger, S., Naschberger, E., Herrmann, C., and Stürzl, M. (2010) Intracellular Trafficking of Guanylate-Binding Proteins Is Regulated by Heterodimerization in a Hierarchical Manner. *PLoS One* 5, e14246.
- (60) Hancock, J. F., Paterson, H., and Marshall, C. J. (1990) A polybasic domain or palmitoylation is required in addition to the CAAX motif to localize p21ras to the plasma membrane. *Cell* 63, 133–139.
- (61) Wright, L. P., and Philips, M. R. (2006) CAAX modification and membrane targeting of Ras. *J. Lipid Res.* 47, 883–891.
- (62) Low, H. H., and Löwe, J. (2006) A bacterial dynamin-like protein. *Nature* 444, 766–769.
- (63) Ford, M. G. J., Jenni, S., and Nunnari, J. (2011) The crystal structure of dynamin. *Nature* 477, 561–566.
- (64) Chappie, J. S., Mears, J. A., Fang, S., Leonard, M., Schmid, S. L., Milligan, R. A., Hinshaw, J. E., and Dyda, F. (2011) A pseudoatomic model of the dynamin polymer identifies a hydrolysis-dependent powerstroke. *Cell* 147, 209–222.
- (65) Gao, S., von der Malsburg, A., Dick, A., Faelber, K., Schröder, G. F., Haller, O., Kochs, G., and Daumke, O. (2011) Structure of myxovirus resistance protein A reveals intra- and intermolecular domain interactions required for the antiviral function. *Immunity* 35, 514–525.

## Supplement B

Supporting Information for:

Triphosphate induced dimerization of human guanylate binding protein 1 involves association of the C-terminal helices - a joint

DEER and FRET study

*Tobias Vöpel<sup>†,§,#</sup>, Carola S. Hengstenberg<sup>†,§</sup>, Thomas-Otavio Peulen<sup>§,§</sup>, Yathrib Ajaï<sup>§</sup>, Claus A.M. Seidel<sup>\*§</sup>, Christian Herrmann<sup>\*†</sup> and Johann P. Klare<sup>\*‡</sup>*

*§TV, CSH and TOP contributed equally.*

<sup>†</sup>Physical Chemistry I, Faculty of Chemistry and Biochemistry, Ruhr-University Bochum, Universitätsstr. 150, 44780 Bochum, Germany

<sup>§</sup>Chair for Molecular Physical Chemistry, Heinrich-Heine University, Universitätsstr. 1, 40225 Düsseldorf, Germany

<sup>‡</sup>Macromolecular Structure Group, Department of Physics, University of Osnabrück, Barbarastr. 7, 49076 Osnabrück, Germany

<sup>#</sup>Present address: Physical Chemistry II, Faculty of Chemistry and Biochemistry, Ruhr-University Bochum, Universitätsstr. 150, 44780 Bochum, Germany

*\*Corresponding authors: [cseidel@hhu.de](mailto:cseidel@hhu.de), Tel: +49 211 81 15881, Fax: +49 211 81 12803; [chr.herrmann@rub.de](mailto:chr.herrmann@rub.de), Tel: +49 234 32 24173, Fax: +49 234 32 14785; [jklare@uos.de](mailto:jklare@uos.de), Tel: +49 541 969 2664, Fax: +49 541 969 2656*

## Supplement B

### Materials and Methods

#### DEER data analysis

Calculation of a distance distribution  $P(r)$  from experimental DEER data is carried out as follows: The experimental raw data (dipolar evolution function)  $S(t)$  is a product of the form factor  $F(t)$  and a background  $B(t)$  from the homogeneous distribution of surrounding molecules carrying dipolarly coupled spins. Consequently, the form factor is obtained by division with the background (assuming a 3D homogeneous distribution) fitted to the raw data after the decay of initial dipolar oscillations.

From any given distance distribution the calculation of the according form factor  $F(t)$  is straight forward<sup>S1,S2</sup>. Contrarily, the inverse problem of calculating a distance distribution  $P(r)$  from  $F(t)$  is ill-posed, i.e. small variations in  $F(t)$  can cause large variations in the resulting distance distribution  $P(r)$ .

To solve such an ill-posed problem the solution is stabilized by adding a second measure (here the curvature/smoothness of  $P(r)$ ) in addition to the quality of the fit to  $F(t)$ . Furthermore the problem is solved under the side constraint  $P(r) > 0$  for all distances  $r$ .

To find a proper weighting factor between these two measures, namely the fit quality and the smoothness of  $P(r)$ , for a successful solution the so-called regularization parameter  $\alpha$  is used. Briefly, for one regularization parameter  $\alpha$ ,  $P(r)$  is calculated by minimizing the objective function:

$$G_{\alpha}(P) = \|S(t) - D(t)\|^2 + \alpha \cdot \left\| \frac{d^2}{dr^2} P(r) \right\|_{\alpha}^2,$$

with the mean square deviation  $\rho(\alpha) = \|S(t) - D(t)\|^2$ , where  $D(t)$  is the simulated dipolar evolution

function, and the “smoothness” of the distance distribution  $\eta(\alpha) = \left\| \frac{d^2}{dr^2} P(r) \right\|_{\alpha}^2$ .

To determine a proper  $\alpha$  as part of finding a solution for the calculation of  $P(r)$ , the mathematical possibility which is least susceptible to experimental noise is the so called L-curve criterion. For a set of solutions calculated with different  $\alpha$ , the plot of the log(“smoothness”,  $\eta$ ) over the log(“deviation”,  $\rho$ ) of the fit from  $F(t)$  typically has an L-shaped form with the ideal  $\alpha$  lying at the kink.<sup>S3</sup> Graphs according to this criterion are used here to determine adequate  $\alpha$  values as implemented in DeerAnalysis2011.<sup>S3</sup>

Calculation of the number of interacting spins has been carried out as described in Jeschke et al.<sup>S3</sup> As all DEER experiments have been performed at X-band frequencies, possible effects due to orientation selection can be largely neglected. Just in extreme cases where the spin label side chain takes a single defined orientation with respect to the protein (i.e. strong immobilization of the spin label side chain due to tertiary contacts) orientation selection has to be taken into account. We are confident (e.g. based on the intermediate mobilities observed in the cw EPR spectra taken at room temperature, data not shown) that this is not the case for the labeling positions chosen in this study. Thus, a number of interacting spins of  $\sim 2$  indeed corresponds to two interacting spins. Our “calculation” (> 90% dimerization) accounts for uncertainties of this measure, e.g. that the modulation depth can be determined unequivocally only if the DEER trace does not show modulations at its end.<sup>S3</sup>

## Supplement B

### ***Rotamer Library Analysis***

In the rotamer library approach (RLA) the canonical ensemble of possible spin label side chain conformations is modeled by a discrete set of 210 precalculated rotamers.<sup>S4</sup> From the RLA a conformational distribution of the spin label side chain at any chosen position in the otherwise fixed protein structure can be determined as described in detail in Ref. S10. In brief, the superposition of the spin label side chain's backbone atoms onto the protein backbone at the respective position provides the orientation of the label side chain with respect to the protein structure and allows to calculate a resulting energy for the spin label-protein interaction from the Lennard Jones potential using the MD force field CHARMM27.<sup>S5</sup> Subsequent Boltzmann weighting and normalization by the partition function yields a probability for each rotamer which is then multiplied by the probability of the spin label side chain to exhibit each conformation. This results in the final rotamer probability distribution at the site of interest. Between two such probability distributions at two positions in the protein, a distance distribution is calculated as the histogram of all pairwise inter-spin distances weighted by the product of their respective probabilities.

### ***FRET measurements***

Ensemble time-correlated single-photon-counting (eTCSPC) measurements were performed using an IBH-5000U (IBH, Scotland) system. The excitation source was a 470 nm diode laser (LDH-P-C 470, Picoquant, Berlin, Germany) operating at 8 MHz for donor excitation. The emission wavelength was set to 520 nm for donor emission. The corresponding monochromator slits were set to 2 nm (excitation path) and 16 nm (emission path) resolution. An additional 500nm cut-off filter was used to reduce the contribution of the scattered light. All measurements were performed at room temperature at magic-angle. The concentrations of the donor-labelled protein were kept below 1  $\mu$ M (for details see legend of Figure S4). The fluorescence measurements were performed at a total protein concentration in a range of 8.5-10  $\mu$ M with a GppNHp-concentration of 500  $\mu$ M. Under these experimental conditions the fraction of proteins not forming dimers is in the range of 50-60%. Fluorescence intensity decay curves were fitted using the iterative re-convolution approach.<sup>S6</sup> To correct instrumental non-linearities the response to uncorrelated light was recorded and considered in the fitting procedure by multiplying the model-function with the normalized/smoothed uncorrelated instrumental response.

### ***FRET-accessible volume calculations***

To describe the behaviour of the fluorophore labels on the macromolecules accessible volume (AV) simulations were performed similar to as described previously.<sup>S7</sup> The AV-algorithm finds fluorophore positions in space for which (1) the fluorophore does not clash with macromolecules and (2) it is possible to route the linker (approximated as a flexible pipe of length  $L_{link}$ ) from the linker attachment point to the fluorophore position. All allowed positions are considered as equally probable which allows one to define an accessible volume for the dye (AV). We used typical parameters for the linkage width ( $w_{link}=4.5$  Å) from Muschielok *et al.*<sup>S7</sup> The linkage lengths ( $L_{link}$ ) were estimated from the fully extended conformations of each linker using the Hyperchemsoftware<sup>S8</sup> and are listed in the table shown below. To take the three quite different spatial dimensions of Alexa 647 into account, we used its real physical dimensions for each calculation of a position distribution and performed three

## Supplement B

independent AV simulations with three different radii  $R_{\text{dye}(i)}$  and superimposed them. Thus, the obtained position distribution represents an average weighted by the number of allowed positions. The following table shows the values used for  $R_{\text{dye}(i)}$ ,  $L_{\text{link}}$  and  $w_{\text{link}}$ :

**Table S1: Parameters used in the AV-calculations**

<b>Donor properties in AV-calculations</b>	
$L_{\text{link}}$	20 Å
$w_{\text{link}}$	4.5 Å
$R_{\text{dye}}$	3.5 Å
<b>Acceptor properties in AV-calculations</b>	
$L_{\text{link}}$	23 Å
$w_{\text{link}}$	4.5 Å
$R_{\text{dye}(1)}$	11 Å
$R_{\text{dye}(2)}$	3 Å
$R_{\text{dye}(1)}$	1.5 Å

The dye-linker diffusion ( $\sim 100$  ns) is slow compared to the fluorescence lifetime of the dyes ( $\sim 4$  ns).<sup>S9</sup> Thus, the distance distributions measured by time-correlated single photon counting, either obtained by fitting Gaussians distance distributions or by deconvolution, can be directly compared to the distributions as expected by the AV-simulations. Whereas the measured mean donor-acceptor distance  $\langle R_{DA} \rangle_{\text{exp}}$  is compared to expected mean donor-acceptor distance based on the accessible volume calculations -  $\langle R_{DA} \rangle_{\text{AV}}$

$$\langle R_{DA} \rangle_{\text{AV}} = \left\langle \left| \mathbf{R}_{D(i)} - \mathbf{R}_{A(j)} \right| \right\rangle_{i,j} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \left| \mathbf{R}_{A(j)} - \mathbf{R}_{D(i)} \right| \quad (\text{S1})$$

## Supplement B

### *Ensemble TCSPC-fitting*

The fits approximately range 250 channels before the maximum of the instrument response functions (IRF) to the first time channel with less than 150 detected photons. The maximum number of counts in the peak was typically 50,000. Due to local quenching the fluorescence decay of the donor in the absence of FRET is already often multi-exponential:

$$F_{D(0)}(t) = \sum_i x_D^{(i)} \exp(-t / \tau_{D(0)}^{(i)}) \quad (\text{S2})$$

Thus, the time resolved fluorescence intensity decays of donor-/acceptor-labeled protein-complex (FRET-sample) were fitted globally with the decays of the donor-/unlabeled protein-complexes (donor only sample, DOnly).

Generally it is reasonable to assume that quenching the donor radiative lifetime is not changed by quenching. Hence, the FRET-rate ( $k_{FRET}$ ) is only determined by the donor-acceptor distance and their relative orientation.<sup>S10</sup> In the presence of FRET, the donor fluorescence decay can be expressed using the donor-acceptor distance distribution  $p(R_{DA})$ :

$$F_D(t) = \sum_i x_D^{(i)} \int_{R_{DA}} p(R_{DA}) \exp\left(-\frac{t}{\tau_{D(0)}^{(i)}} \left[1 + (R_0 / R_{DA})^6\right]\right) dR_{DA} \quad (\text{S3a})$$

Here we either assumed Gaussian distribution of donor-acceptor distances ( $p(R_{DA})$ ) with a mean of  $\langle R_{DA} \rangle$  and a half-width of  $\sigma_{DA}$  or determine  $p(R_{DA})$  by deconvolution of the fluorescence decays by using the maximum-entropy method.<sup>S11,S12</sup> As is illustrated in figure S5 the assumption of Gaussian distributed distances is an good approximation within the AV-framework. For Gaussian distributed donor-acceptor distances the lifetime-decay is expressed as follows:

$$F_D(t) = \sum_i x_D^{(i)} \int_{R_{DA}} \frac{1}{\sqrt{2\pi}\sigma_{DA}} \exp\left(-\frac{(R_{DA} - \langle R_{DA} \rangle_{\text{exp}})^2}{2\sigma_{DA,\text{exp}}^2}\right) \exp\left(-\frac{t}{\tau_{D(0)}^{(i)}} \left[1 + (R_0 / R_{DA})^6\right]\right) dR_{DA} \quad (\text{S3b})$$

In addition, a fraction of Donor-only molecules ( $x_{D\text{Only}}$ ) and a constant offset  $c$  was considered:

$$F(t) = (1 - x_{D\text{Only}})F_D(t) + x_{D\text{Only}}F_{D(0)}(t) + c \quad (\text{S4})$$

## Supplement B

**Table S2: Specific GTPase Activities**

Mutant	specific activity (min <sup>-1</sup> )
wt	22.8 <sup>a</sup>
Cys-9	56.7 ± 1.2
Cys-9/N18C	64.1 ± 0.2
Cys-9/N18C <sup>SL</sup>	39.9 ± 0.7
Cys-9/C225	31.1 ± 0.2
Cys-9/C225 <sup>SL</sup>	65.4 ± 0.2
Cys-9/K567C	57.4 ± 2.9
Cys-9/K567C <sup>SL</sup>	48.5 ± 1.3
Cys-9/Q577C	43.4 ± 0.8
Cys-9/Q577C <sup>SL</sup>	71.5 ± 1.8

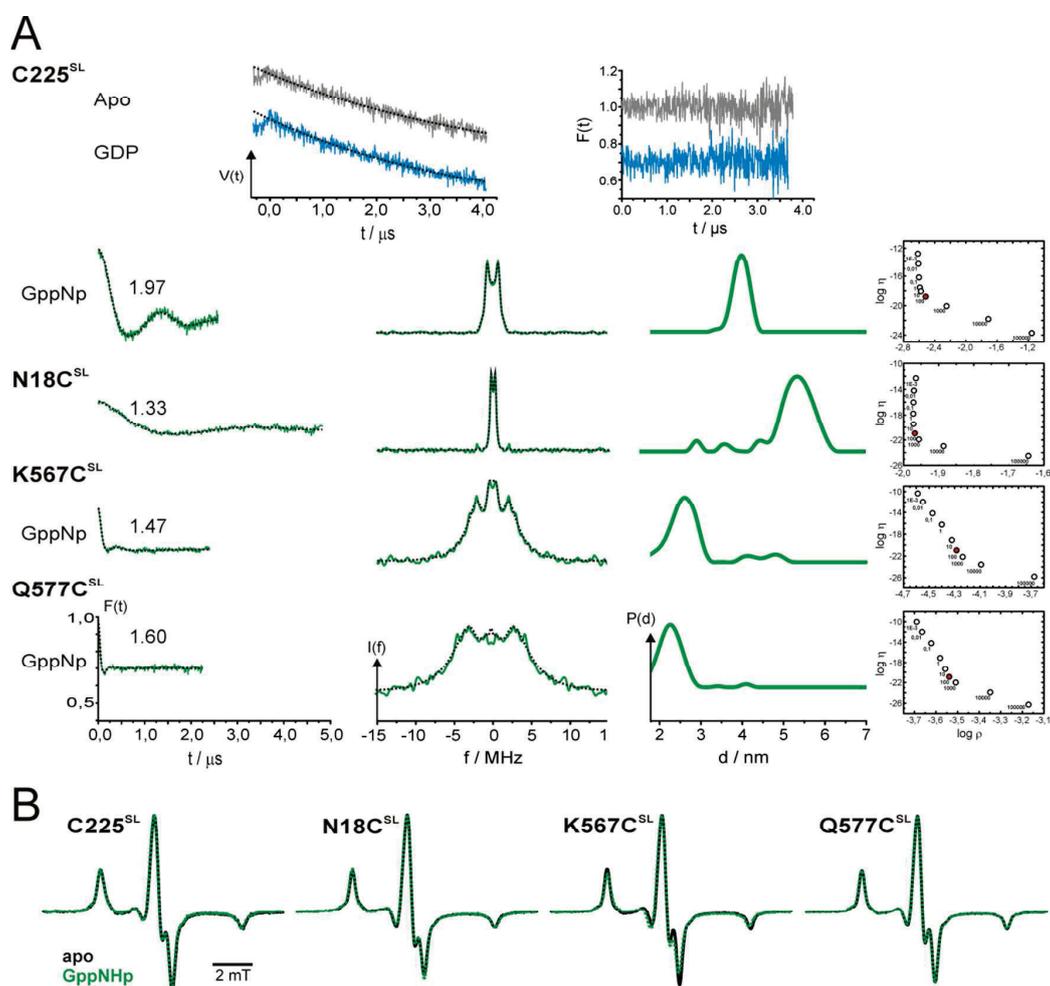
<sup>a</sup>Taken from [S13]

**Table S3: FRET-measurements - obtained fitting parameters**

		Q18C <sup>D</sup> -Q18C <sup>A</sup>	Q577C <sup>D</sup> -Q577C <sup>A</sup>	
		1-state	2-State	1-state
Donor-sample	$\chi_r^2$	1.092	1.038	1.100
	$x_{D1}$	0.11	0.09	0.09
	$\tau_{D1}$ [ns]	0.33	2.54	2.54
	$x_{D2}$	0.75	0.91	0.91
	$\tau_{D2}$ [ns]	4.04	4.26	4.26
	$x_{D3}$	0.14		
	$\tau_{D3}$ [ns]	1.74		
FRET-sample	$x_{D\text{Only}}$	0.73	0.69	0.4
	$w_{DA}$ [Å]	14	15.4	36.2
	$\langle R_{DA,1} \rangle$ [Å]	61	36	30
	$\langle R_{DA,2} \rangle$ [Å]		67	

## Supplement B

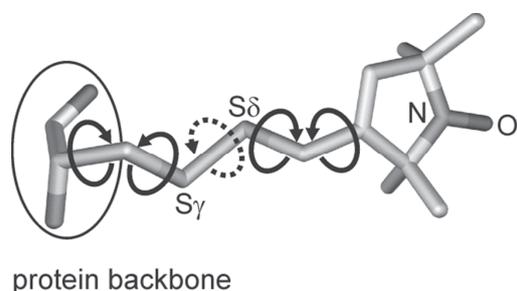
**Figure S1: A. DEER data, B. cw EPR spectra recorded at 25°C**



**Figure S1: A.** DEER characterization of nucleotide-dependent inter spin distances of singly labeled hGBP1 (C225<sup>SL</sup>, N18C<sup>SL</sup>, K567C<sup>SL</sup> and Q577C<sup>SL</sup>) in the apo state (grey, black), with GDP (blue) and GppNHp (green). Data for the apo and GDP state is only shown for C225<sup>SL</sup>, where in the left panel the raw dipolar evolution data and the respective background fits (broken line) are given. In both cases, the dipolar evolution trace obtained from the experiment can be entirely fitted with a background function, as can also be seen from the residuals shown in the right hand panel. For GppNHp: Left panel, background corrected dipolar evolution data and fits to the experimental data obtained by Tikhonov regularization (broken lines). Numbers given are the number of interacting spins calculated based on the modulation depths of the dipolar evolution traces. Second-left panel, dipolar spectra (Fourier transform of the dipolar evolution data); Second-right panel, distance distributions obtained by Tikhonov regularization; Right panel, L curves. The regularization parameter corresponding to the data shown in the other panels is indicated by the red filled circle. **B.** Low temperature (160 K) cw EPR spectra for C225<sup>SL</sup>, N18C<sup>SL</sup>, K567C<sup>SL</sup> and Q577C<sup>SL</sup> in the apo- (black) and GppNHp-bound state (green, dotted). The absence of line broadening in the spectra recorded in the presence of GppNHp indicates that no distances  $\lesssim 17$  Å are present.

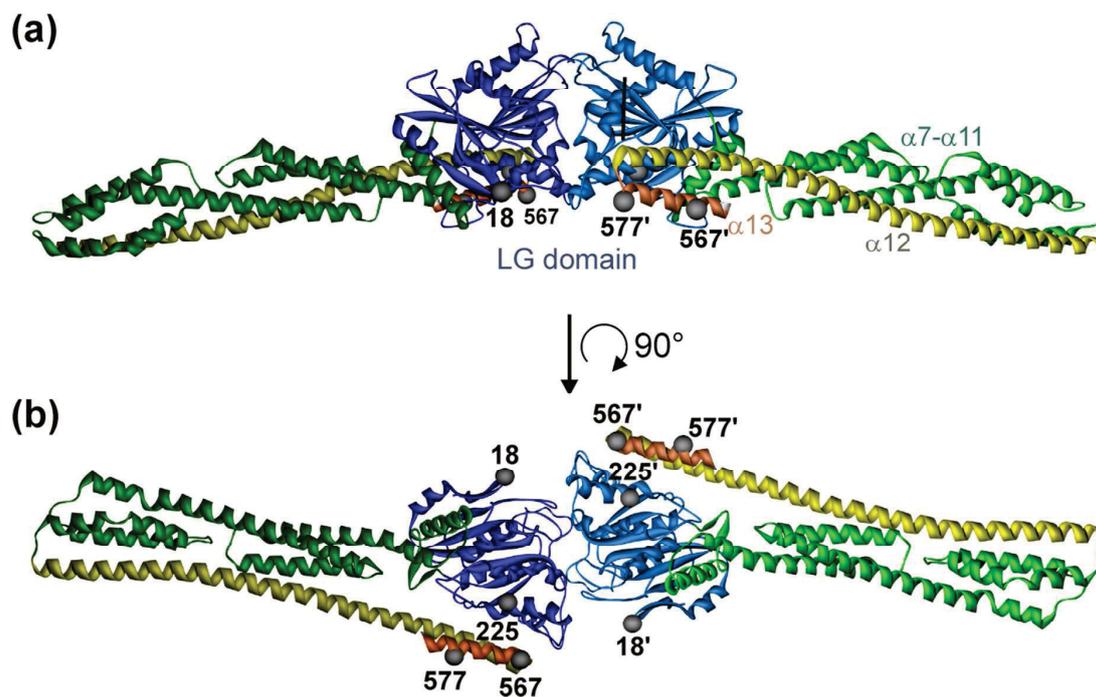
## Supplement B

**Figure S2: Structure of the spin label side chain ( $C^{SL}$ )**



**Figure S2:** Structure of the spin label (R1) side chain. Bend arrows indicate rotatable bonds. The  $S_{\gamma}$ - $S_{\delta}$  bond has a higher torsion barrier, thus rotation is slower compared to the other bonds, as indicated by the dotted arrow.

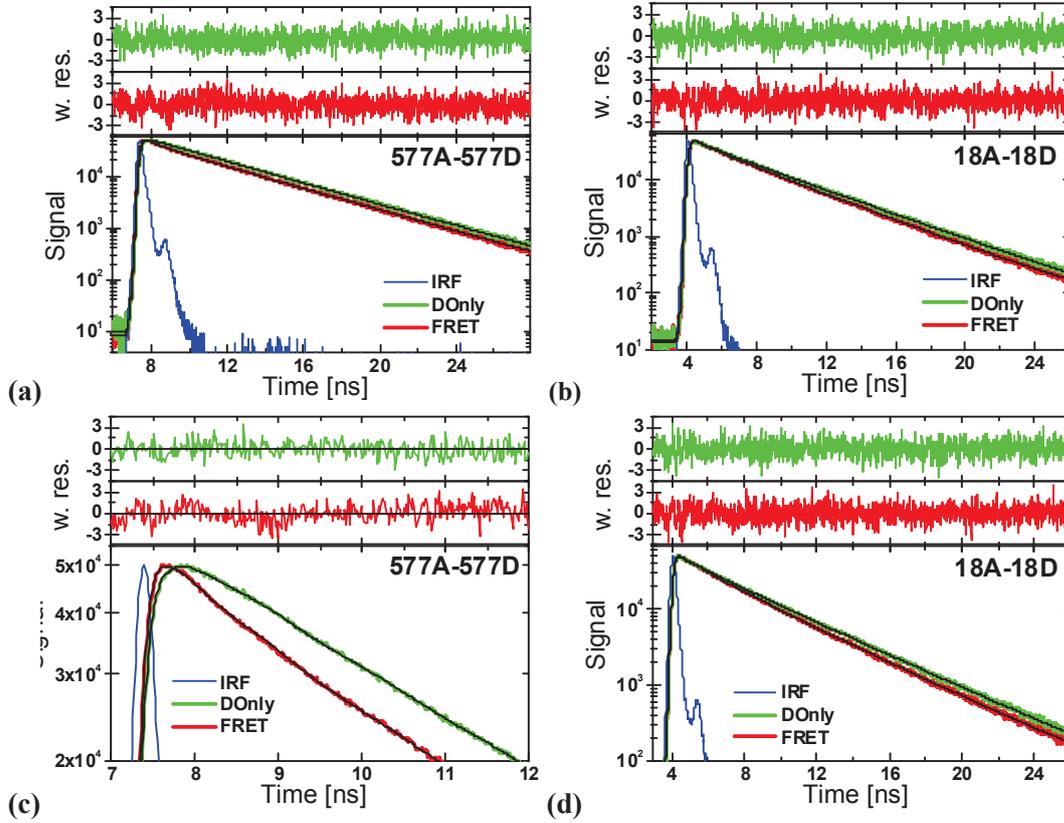
**Figure S3: Initial Dimer Model**



**Figure S3:** Model of the full-length dimer obtained by superimposing the LG domains of the full-length crystal structure in the GppNHp bound state (pdb: 1F5N)<sup>S14</sup> onto the LG domains in the dimer structure obtained in the presence of GppNHp (PDB ID 2BC9).<sup>S15</sup> (a) side view, (b) bottom view. Spin labeled residues are marked by grey spheres at the positions of their  $C_{\alpha}$  atoms.

## Supplement B

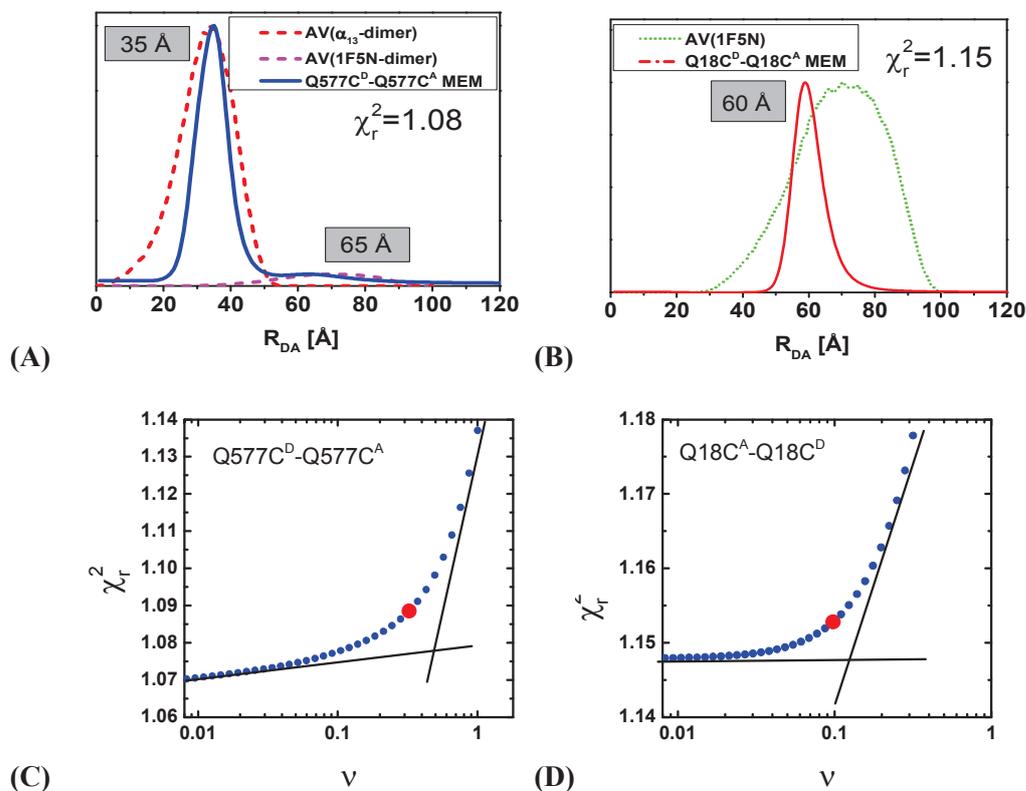
**Figure S4: FRET measurements**



**Figure S4: Gaussian-fits** Experimental donor (D) lifetime decay (Alexa488) in presence of GppNHp at magic angle emission for mixtures with acceptor (A) labeled samples (Alexa647 - FRET-decay, red) or unlabeled proteins (U) (donor-only decay /DOnly-decay, green). The DOnly-decays were fitted by a formal lifetime-decay with either two or three lifetimes, the FRET-decays are fitted by Gaussian distance-distributions using the donor-lifetimes determined by the DOnly-decay. The weighted residuals of the donor-decay and the FRET-decay are displayed in the upper panel in green and red, respectively. The concentrations of the donor, acceptor and unlabeled proteins are denoted as  $c_D$ ,  $c_A$  and  $c_U$  respectively, the GppNHp-concentration is given by  $c_{GppNHp}$ . The fraction of donor molecules (either donor-molecules not forming dimer at the given conditions or donor-molecules forming dimers with unlabeled proteins) is given by  $x_{DOnly}$  and a free fitting parameter. The fractions and lifetimes of the donor-decay are  $x_i$  and  $\tau_i$ , respectively. The reduced  $\chi_r^2$  of the donor-decay and the FRET-decay are  $\chi_{rD}^2$  and  $\chi_{rF}^2$  respectively (a/c) Q577C/Q577C-dimer – DOnly-decay ( $c_D=0.53 \mu\text{M}$ ,  $c_U=8.8\mu\text{M}$ ,  $c_{GppNHp}=500\mu\text{M}$  - green)  $x_1=0.92$   $\tau_1=4.2$   $x_2=0.08$   $\tau_2=2.6\text{ns}$ , FRET-decay ( $c_A=7.1 \mu\text{M}$ ,  $c_D=0.53$ ,  $c_U=0.7\mu\text{M}$ ,  $c_{GppNHp}=500\mu\text{M}$ ) – mean-distance  $\langle R_{DA,1} \rangle = 36 \text{ \AA}$  ( $x_1=92\%$ ),  $\langle R_{DA,1} \rangle = 67 \text{ \AA}$  ( $x_2=8\%$ ),  $w_{DA}=15.4 \text{ \AA}$ ,  $x_{DOnly}=0.69$   $\chi_{rD}^2=1.04$ . (b/d) N18C/N18C-dimer - DOnly-decay ( $c_D=0.5 \mu\text{M}$ ,  $c_U=8.5\mu\text{M}$ ,  $c_{GppNHp}=600\mu\text{M}$ )  $x_1=0.75$   $\tau_1=4.04$   $x_2=0.14$   $\tau_2=1.76$   $x_3=0.11$   $\tau_3=0.33$ , FRET-decay ( $c_A=5.1 \mu\text{M}$ ,  $c_D=0.5$ ,  $c_U=2.6\mu\text{M}$ ,  $c_{GppNHp}=600\mu\text{M}$ ) mean-distance  $\langle R_{DA} \rangle = 61 \text{ \AA}$ ,  $w_{DA}=14 \text{ \AA}$ ,  $x_{DOnly}=0.73$   $\chi_{rF}^2=1.10$ .

## Supplement B

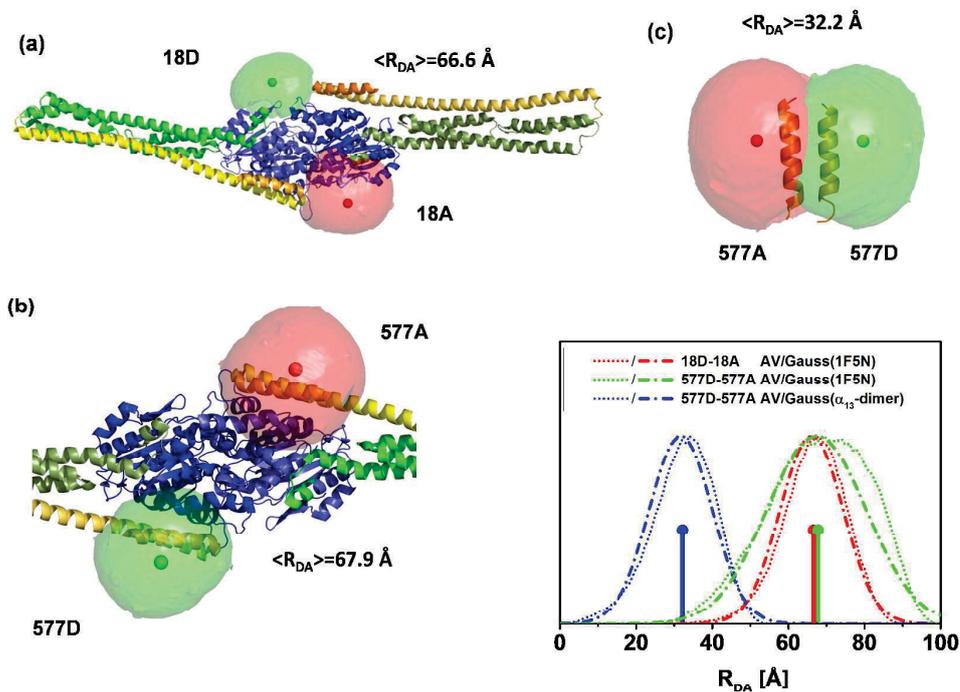
**Figure S5: Maximum entropy distance distributions**



**Figure S5: Maximum-entropy distance distributions of the fluorescence decays** – (A/B) the solid lines represent the deconvoluted donor-acceptor distance distributions, the dotted lines correspond to the expected distributions determined by the accessible-volume algorithm for the  $\alpha 13$ -dimer ( $Q577C^D$ - $Q577C^A$ ) and the full-length dimer ( $N18C^D$ - $N18C^A$ ). The mean  $\langle R_{DA} \rangle$  of the distance distributions are 35/65 Å and 63 Å for the  $Q577C^D$ - $Q577C^A$  dimer and the  $N18C^D$ - $N18C^A$  dimer, respectively, being in line with the distances determined by fitting of Gaussian distance distributions. (C/D) the regularization parameter/weight of the entropy was chosen according to the L-curve criterion<sup>S16</sup>.

## Supplement B

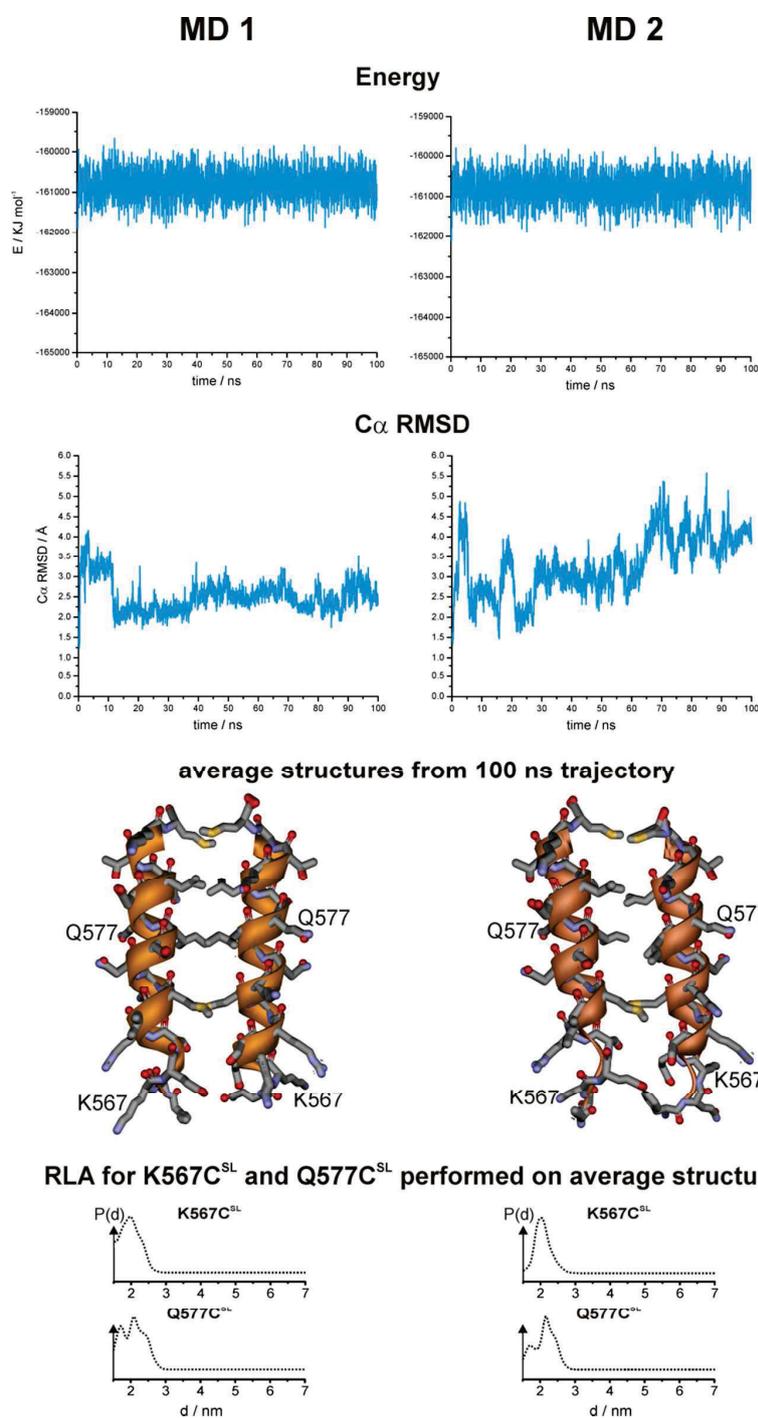
**Figure S6: Expected dye distributions and mean donor-acceptor distances**



**Figure S6:** Expected dye distributions and mean donor-acceptor distances  $\langle R_{DA} \rangle = \langle \|r_D - r_A\|_2 \rangle$  for different structural models: (a,b) full-length dimer model obtained by superimposing the LG domains the crystal structure in the GppNHp bound state (pdb: 1F5N)<sup>S14</sup> onto the LG domains in the dimer structure obtained in the presence of GppNHp (PDB ID 2BC9)<sup>S15</sup>. (c)  $\alpha$ 13-helix dimer obtained by MD-simulations. The dye distributions for the donor-fluorophore Alexa488 and the acceptor fluorophore Alexa647 are depicted as green and red surfaces respectively. (d) Comparison of distance distributions obtained by accessible volume calculations and Gaussian distributions with identical  $\langle R_{DA} \rangle$  and  $hw_{DA}$ . The deviations are generally smaller than the experimental error justifying the use of simple Gaussian-distributions as a model-function of the fits.

## Supplement B

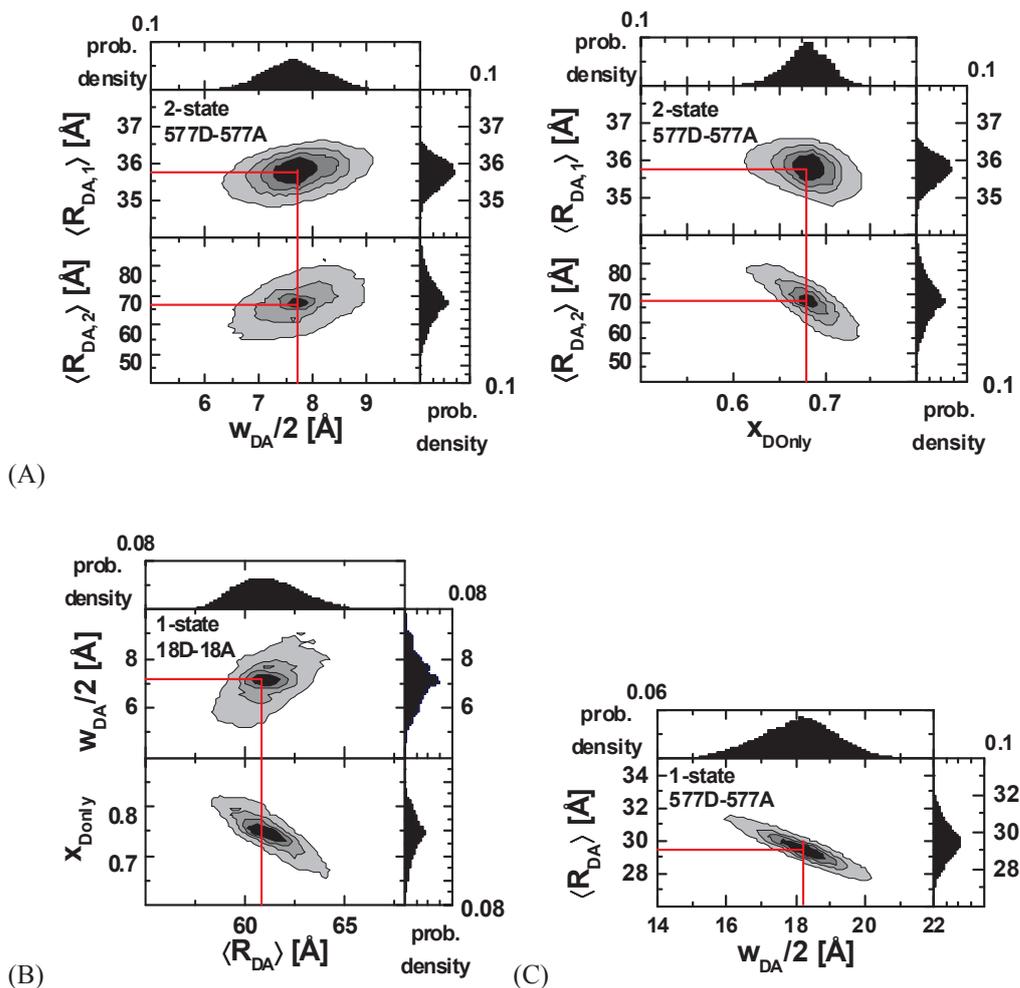
Figure S7: Results of the  $\alpha 13$  dimer MD simulations



**Figure S7:** Results of the MD simulations. From top to bottom: Energies ( $\text{kJ mol}^{-1}$ ) vs. time, C $\alpha$  root means square deviation (RMSD) vs. time, average structure calculated from 100 ns trajectories, RLA for K567C<sup>SL</sup> and Q577C<sup>SL</sup> performed on the average structure shown above.

## Supplement B

Figure S8: Marginal distributions



**Figure S8:** Marginal distributions of the experimental data for different models (1-Gaussian, 2-Gaussian) for the samples 577D-577A and 18D-18A, (A) no significant correlations between the distance  $\langle R_{DA,1} \rangle$  and the donor-fraction  $x_{Donly}$  is visible, the larger distance  $\langle R_{DA,2} \rangle$  is correlated to the donor-only fraction, both distances show only weak correlations with the global donor-acceptor width  $w_{DA}$ , (B) fitting the same data with only one distance a strong correlation between the fitted distance and  $w_{DA}$  is visible. (C) Due to the large distance a weak correlation between the donor-fraction  $x_{Donly}$  and the distance  $\langle R_{DA} \rangle$  is visible.

## Supplement B

### Supplementary References

- [S1] Martin, R. E., Pannier, M., Diederich, F., Gramlich, V., Hubrich, M. & Spiess, H. W. (1998). Determination of End-to-End Distances in a Series of TEMPO Diradicals of up to 2.8 nm Length with a New Four-Pulse Double Electron Resonance Experiment. *Angew. Chem. Int. Ed.* 37, 2833-2837.
- [S2] Pannier, M., Veit, S., Godt, A., Jeschke, G. & Spiess, H. W. (2000). Dead-Time Free Measurement of Dipole-Dipole Interactions between Electron Spins. *J. Magn. Res.* 142, 331-340.
- [S3] Jeschke, G., Chechik, V., Ionita, P., Godt, A., Zimmermann, H., Banham, J. E., Timmel, C. R., Hilger, D. & Jung, H. (2006). DeerAnalysis2006 - a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.* 30, 473-498.
- [S4] Polyhach, Y., Bordignon, E. & Jeschke, G. (2011). Rotamer libraries of spin labelled cysteines for protein studies. *Phys. Chem. Chem. Phys.* 13, 2356-2366.
- [S5] Mackerell, A. D., Jr., Feig, M. & Brooks, C. L., III (2004). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25, 1400-1415.
- [S6] O'Connor, D. V. & Phillips, D. (1984). *Time-correlated Single Photon Counting*. Academic Press, New York.
- [S7] Muschielok, A., Andrecka, J., Jawhari, A., Bruckner, F., Cramer, P. & Michaelis, J. (2008) A nano-positioning system for macromolecular structural analysis. *Nature Meth.* 5, 965-971.
- [S8] HyperChem (TM) Professional v. 7.51 (1115 NW 4th Street, Gainesville, Florida 32601, USA).
- [S9] Sindbert, S., Kalinin, S., Nguyen, H., Kienzler, A., Clima, L., Bannwarth, W., Appel, B., Müller, S. & Seidel, C.A.M. (2011) Accurate Distance Determination of Nucleic Acids via Förster Resonance Energy Transfer: Implications of Dye Linker Length and Rigidity. *J. Am. Chem. Soc.* 133, 2463-2480.
- [S10] Kalinin, S. & Johansson, L.B.-Å. Energy Migration and Transfer Rates Are Invariant to Modeling the Fluorescence Relaxation by Discrete and Continuous Distributions of Lifetimes. (2004) *J. Phys. Chem. B* 108, 3092-3097.
- [S11] Livesey, A. K.; Skilling J. Maximum Entropy Theory. *Acta Crystallogr. Sect. A* 1985, 41, 113-122.
- [S12] Brochon, J. C. Maximum entropy method of data analysis in time-resolved spectroscopy. *Methods Enzymol.* 1994, 240, 262-311.
- [S13] Vöpel, T., Kunzelmann, S. & Herrmann, C. (2009). Nucleotide dependent cysteine reactivity of hGBP1 uncovers a domain movement during GTP hydrolysis. *FEBS Lett.* 583, 1923-1927.
- [S14] Prakash, B., Praefcke, G. J. K., Renault, L., Wittinghofer, A. & Herrmann, C. (2000). Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins. *Nature* 403, 567-571.
- [S15] Ghosh, A., Praefcke, G. J. K., Renault, L., Wittinghofer, A. & Herrmann, C. (2006). How guanylate-binding proteins achieve assembly-stimulated processive cleavage of GTP to GMP. *Nature* 440, 101-104.
- [S16] Hansen, P.C. & Oleary, D.P. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.*, 14 (1993), pp. 1487-1503

## **Supplement C**

**Guanylate binding 1 proteins (GBPs) directly attack  
*T. gondii* via supramolecular complexes**

1                                   **Guanylate binding proteins (GBPs)**

2                                   **directly attack *T. gondii* via supramolecular complexes**

3  
4 Elisabeth Kravets<sup>1,4</sup>, Daniel Degrandi<sup>1,4</sup>, Qijun Ma<sup>2,4</sup>, Thomas Peulen<sup>2</sup>, Suren Felekyan<sup>2</sup>, Ralf  
5 Kühnemuth<sup>2</sup>, Stefanie Weidtkamp-Peters<sup>3</sup>, Claus A.M. Seidel<sup>2,5\*</sup>, Klaus Pfeffer<sup>1,5\*</sup>

6  
7 <sup>1</sup> Institute of Medical Microbiology and Hospital Hygiene

8 <sup>2</sup> Institute for Molecular Physical Chemistry

9 <sup>3</sup> Center of Advanced Imaging

10 Heinrich-Heine-University Düsseldorf, Universitätsstr. 1, D-40225 Düsseldorf, Germany

11 <sup>4</sup> Co-first authors

12 <sup>5</sup> Co-senior authors

13  
14 \*Corresponding authors:

15 Prof. Dr. Klaus Pfeffer  
16 Institute of Medical Microbiology and Hospital Hygiene  
17 Heinrich-Heine-University Düsseldorf  
18 Tel.: +49 211 8112459; Fax: +49 211 8115906  
19 E-mail: klaus.pfeffer@hhu.de.  
20

21 Prof. Dr. Claus A. M. Seidel  
22 Institute for Molecular Physical Chemistry  
23 Heinrich-Heine-University Düsseldorf  
24 Tel.: +49 211 8115881; Fax: +49 211 81-12803;  
25 E-mail: cseidel@hhu.de  
26

## Supplement C

### 27 **Abstract**

28 GBPs are essential for immunity against intracellular pathogens, especially for *T. gondii* control. Here,  
29 the molecular interactions of murine GBPs (mGBP1/2/3/5/6), homo- and hetero-multimerization  
30 properties of mGBP2 and its function in parasite killing were investigated by mutational,  
31 Multiparameter Fluorescence Image Spectroscopy, and live cell microscopy methodologies. Control of  
32 *T. gondii* replication by mGBP2 requires GTP hydrolysis and isoprenylation thus, enabling reversible  
33 oligomerization in vesicle-like structures. mGBP2 undergoes structural transitions between monomeric,  
34 dimeric and oligomeric states visualized by quantitative FRET analysis. mGBPs reside in at least two  
35 discrete subcellular reservoirs and attack the parasitophorous vacuole membrane (PVM) as orchestrated,  
36 supramolecular complexes forming large, densely packed multimers comprising up to several thousand  
37 monomers. This dramatic mGBP enrichment results in the loss of PVM integrity, followed by a direct  
38 assault of mGBP2 upon the plasma membrane of the parasite. These discoveries provide vital dynamic  
39 and molecular perceptions into cell-autonomous immunity.

40

## Supplement C

### 41 **Introduction**

42 IFN $\gamma$  is an immunomodulatory cytokine that rapidly activates potent host cell effector mechanisms to  
43 confront a variety of intracellular pathogens (Decker et al., 2002). Some of the most abundantly IFN $\gamma$   
44 induced proteins are the 65-kDa guanylate-binding proteins (GBPs), which mediate cell-autonomous  
45 immunity (MacMicking, 2012, Degrandi et al., 2013, Pilla et al., 2014, Meunier et al., 2015). GBPs are  
46 related to the dynamin super family of GTPases (Praefcke and McMahon, 2004) and are highly  
47 conserved throughout the vertebrate lineage (Vestal and Jeyaratnam, 2011). The human genome harbors  
48 seven GBPs and at least one pseudogene, whereas the mouse genome contains 11 GBPs and two  
49 pseudogenes (Kresse et al., 2008, Olszewski et al., 2006). The gene loci of murine GBPs (mGBPs) are  
50 tandemly organized in clusters on chromosomes 3 and 5 (Degrandi et al., 2007, Kresse et al., 2008).

51 GBPs contain a conserved GTPase-domain which binds guanine nucleotides with low affinities. This  
52 induces nucleotide dependent GBP multimerization and cooperative hydrolysis of GTP via GDP to  
53 GMP (Praefcke et al., 2004, Ghosh et al., 2006, Kravets et al., 2012, Prakash et al., 2000b). Some GBPs  
54 are isoprenylated, endowing them with the ability to associate with intracellular membranous  
55 compartments (Vestal et al., 2000, Degrandi et al., 2013).

56 Murine GBPs (mGBPs) exert a major impact on cell-autonomous restriction of *Toxoplasma gondii*  
57 (Yamamoto et al., 2012, Degrandi et al., 2007, Selleck et al., 2013, Degrandi et al., 2013). *T. gondii* is  
58 an apicomplexan protozoan parasite with a broad host range, is distributed worldwide and causes  
59 serious and often fatal infections in immunocompromised hosts (Gazzinelli et al., 2014). *T. gondii*  
60 infection experiments in mice deficient for a cluster of mGBPs on chromosome 3 (Yamamoto et al.,  
61 2012) or solely for mGBP1 or mGBP2 (Degrandi et al., 2013, Selleck et al., 2013) prove that mGBPs  
62 are essential immune effector molecules mediating antiparasitic resistance. In several cell types distinct  
63 mGBPs accumulate at the parasitophorous vacuole membrane (PVM) of *T. gondii* (Degrandi et al.,  
64 2007, Kravets et al., 2012, Degrandi et al., 2013).

65 In previous studies, introduction of point mutations into the key positions of the conserved motifs of the  
66 GTPase-domain (R48A, K51, E99A, D182N) and the isoprenylation site of mGBP2 (C586S), clearly  
67 showed that nucleotide binding, multimerization, GTP-hydrolysis and membrane anchoring, are

## Supplement C

68 essential for localization in vesicle-like structures (VLS) and for the recruitment of mGBP2 to the PVM  
69 of *T. gondii* (Kravets et al., 2012, Degrandi et al., 2013). However, the assembly of homo- and hetero-  
70 mGBP multimers, their composition in distinct subcellular compartments, localization-dependent  
71 multimerization as well as their capacity to control replication of *T. gondii* in living cells remained  
72 enigmatic.

73 Therefore quantitative live-cell-imaging technologies were employed revealing seminal information on  
74 localization, interaction, concentration, structure and dynamics of biomolecules. To investigate the  
75 structure, composition and interaction of proteins, Förster resonance energy transfer (FRET) (Giepmans  
76 et al., 2006) is combined with Multiparameter fluorescence image spectroscopy (MFIS) (Kudryavtsev  
77 et al., 2007, Weidtkamp-Peters et al., 2009), which enables unique advances in FRET imaging. In  
78 MFIS, a variety of fluorescence parameters is monitored simultaneously with picosecond accuracy,  
79 allowing the determination of many fluorescence parameters in a *pixel-wise* analysis such as number of  
80 photons, anisotropies, fluorescence lifetimes, and signal ratios by statistically most efficient estimators  
81 (Sisamakris et al., 2010) and to plot distinct parameters in MFIS pixel frequency histograms. The  
82 combination of MFIS and FRET experiments (MFIS-FRET) enables a quantitative analysis of the  
83 biophysical properties of homomeric and heteromeric molecular complexes in living cells (Stahl et al.,  
84 2013). This allows the identification and selection of pixel populations with unique properties for a  
85 detailed *pixel-integrated* analysis. Importantly, live cell measurements with MFIS can achieve the  
86 resolution and precision of traditional *in vitro* measurements of molecule ensembles with respect to the  
87 number of resolved species and rate constants.

88 Here, by advanced biophysical MFIS-FRET technology, it is demonstrated that the GTPase activity and  
89 isoprenylation of mGBP2 are prerequisites for its multimerization. The multimerization is essential for  
90 control of *T. gondii* replication. Colocalization and MFIS analysis of mGBPs showed intermolecular  
91 interaction of mGBP2 with itself, with mGBP1 and mGBP3, but not with mGBP6 in VLS in living  
92 cells. Interestingly, the interaction partnerships were recapitulated at the PVM of *T. gondii*. Moreover,  
93 characteristic interaction affinities of mGBP complexes were individually quantified. For the first time,  
94 we show that in the process of attacking *T. gondii*, mGBP2 directly targets the plasma membrane of the

## Supplement C

95 parasite after disruption and permeabilization of the PVM. These investigations enable a discrete  
96 understanding of the dynamics and intracellular interactions of mGBP effector molecules in *T. gondii*  
97 host defense.

98

## Supplement C

### 99 **Results**

100 *Multimerization of mGBP2 WT and mutants, determined by intracellular homo-FRET MFIS analysis*

101 Site-directed mutagenesis of mGBP2 revealed that GTP-binding and hydrolysis as well as C-terminal  
102 isoprenylation affect the localization of mGBP2 in the cell (Degrandi et al., 2013, Kravets et al., 2012).  
103 However, the role of the GTPase activity and isoprenylation on the multimerization ability of mGBP2 in  
104 living cells is unknown.

105 Therefore, MFIS-FRET measurements and fluorescence-anisotropy-based homo-FRET analysis were  
106 employed in living IFN- $\gamma$  stimulated mGBP2<sup>-/-</sup> MEFs reconstituted either with GFP-fused mGBP2 WT  
107 protein (hereafter referred to as G-mGBP2 MEFs) or with one of the GTPase-domain mutants (R48A,  
108 K51A, E99A, D182N) or with the isoprenylation mutant (C586S) (Fig. 1A).

109 The mean steady-state anisotropy of GFP in the cytosol was experimentally determined as  $\langle r_D \rangle_{\text{cytosol}} =$   
110 0.328, which is in agreement with the value predicted by the Perrin equation (Lakowicz, 2006) using the  
111 known mean global rotational diffusion time  $\rho_{\text{global}} \approx 15$  ns for freely diffusing GFP. When GFP is fused  
112 to mGBP2, two opposing effects need to be considered (Fig.1B). First, its rotational freedom is  
113 restricted and therefore  $r_D$  increases; second, homo-FRET between G-mGBP2 complexes reduces  $r_D$  by  
114 depolarization of the total GFP signal. Consequently, the average steady-state anisotropy of WT G-  
115 mGBP2 in the cytosol  $\langle r_D \rangle_{\text{cytosol}}$  remained comparable to the value for free GFP (Fig. 1C). In contrast,  
116 the GFP signal intensity ( $S_{G,G}$ ) in VLS increased significantly, indicating an enrichment of mGBP2  
117 molecules in these structures (Fig. 1C) accompanied by a significant reduction of the average anisotropy  
118  $\langle r_D \rangle_{\text{VLS}}$ , suggesting an increased mGBP2 homo-multimerization (Fig. 1A, C).

119 The nucleotide binding and hydrolysis impaired K51A mutant does not localize in VLS (Kravets et al.,  
120 2012). This mutant showed a higher average anisotropy ( $\langle r_D \rangle_{\text{cytosol}} = 0.336$ ) as compared to the cytosolic  
121 WT mGBP2 (Fig. 1A, C) due to the absence of homo-FRET, proving its incapability to form multimers.  
122 Next, the mean anisotropies of averages over whole MEFs  $\langle r_D \rangle_{\text{cell}}$  were determined (Fig. 1D). The  
123 hydrolytically impaired mGBP2 mutants R48A and E99A (Kravets et al., 2012) showed significantly  
124 increased  $\langle r_D \rangle_{\text{cell}}$  values (Fig. 1A, D), further proving that the GTPase activity is essential for

## Supplement C

125 multimerization in living cells. The nucleotide binding deficient mGBP2 mutant D182N showed  
126 significantly increased  $\langle r_D \rangle_{\text{cell}}$  value (Fig. 1A, D) as compared to WT mGBP2 and mutants R48A and  
127 E99A reflects the low multimerization capability of this mutant. The recombinant isoprenylation mutant  
128 (C586S) did not show altered nucleotide binding, hydrolysis activity or multimerization of mGBP2 in  
129 cell-free analyses (Fig. S2). Nevertheless, this mutant did not localize in VLS (Degrandi et al., 2013)  
130 and showed anisotropy values comparable to the dysfunctional K51A mutant (Fig. 1D).

131 Altogether, these data provide compelling evidence that nucleotide binding and membrane  
132 anchoring are prerequisites for multimerization of mGBP2 in living cells. The degree of multimerization  
133 of mGBP2 increases from cytosol to VLS.

### 134 *Multimerization of mGBP2 WT and mutants at the PVM of T. gondii*

135 mGBPs were reported to be involved in rupture of *T. gondii* PVMs few hours after infection and are  
136 important for *T. gondii* control *in vivo* (Degrandi et al., 2013, Selleck et al., 2013, Yamamoto et al.,  
137 2012). Previously, it could be determined that the GTPase activity as well as isoprenylation regulate the  
138 recruitment of mGBP2 to the PVM of *T. gondii* (Degrandi et al., 2013, Kravets et al., 2012). The next  
139 step therefore was to elucidate the impact of the GTPase activity and the isoprenylation of mGBP2 on  
140 the ability to multimerize at the PVM and to control intracellular *T. gondii* replication. Hence, G-  
141 mGBP2 MEFs as well as MEFs expressing GTPase and isoprenylation mutants were infected with *T.*  
142 *gondii* and analyzed by MFIS homo-FRET assays. Also, the ratio of replicative units, so called rosettes,  
143 versus single parasites was determined 32 hours after infection (Fig. 2).

144 A marked decrease of fluorescence intensities of WT mGBP2 in the cytosol of infected cells (Fig. 2A,  
145 B) compared to uninfected cells (Fig. 1C) concurrent with a strong increase of the mGBP2  
146 concentration at the PVM of *T. gondii* was observed along with a further decrease in anisotropy (Fig.  
147 2A, B; Fig. S1). This raises the question on a distinct composition of the mGBP2 complexes at the  
148 PVM, which will be addressed below by pixel-integrated MFIS analysis.

149 As shown previously, the enzymatically dysfunctional K51A and the isoprenylation C586S mutants  
150 showed nearly no recruitment to the PVM (Kravets et al., 2012, Degrandi et al., 2013). Interestingly, as

## Supplement C

151 shown here, the corresponding anisotropies (Fig. 2A, B, C) did not significantly change in comparison  
152 to the uninfected situation (Fig. 1). These mutants were incapacitated in controlling *T. gondii* replication  
153 (Fig. 2D). The R48A and E99A mutants, which have reduced capacity to recruit to the PVM (Kravets et  
154 al., 2012), showed slightly increased anisotropy at the PVM as compared to WT mGBP2 (Fig. 2C) and  
155 a reduced capability to restrict *T. gondii* growth (Fig. 2D). For the D182N mutant a higher anisotropy at  
156 the PVM in comparison to WT mGBP2 could be determined, suggesting a lower degree of  
157 multimerization. This correlated with insufficient control of *T. gondii* growth, comparable to the K51A  
158 and C586S mutants (Fig. 2D).

159 In summary, it can be concluded that at the PVM the enrichment of mGBP2 is increased compared to  
160 VLS. Nucleotide binding, GTPase activity as well as membrane anchoring regulate the multimerization  
161 capability of mGBP2 at the PVM and are prerequisites for the control of *T. gondii* replication.

### 162 *Colocalization and hetero-FRET studies of mGBPs.*

163 Several members of the mGBP family localize in VLS in IFN $\gamma$  stimulated cells (Degrandi et al., 2007).  
164 However, it is unclear whether co-compartmentalization of mGBPs and molecular interactions between  
165 them in VLS occur. For this purpose, G-mGBP2 MEFs were cotransduced with mCherry fusion  
166 proteins of mGBP1, mGBP2, mGBP3, mGBP5, and mGBP6 (hereafter referred to as G-mGBP2/mCh-  
167 mGBPx) and confocal imaging studies were performed. (Fig. 3, Fig. S3). All of the analyzed mGBPs  
168 showed a vesicular distribution except for mGBP5 (Fig. 3). A correlation of localization could be  
169 computed employing the Pearson's coefficient,  $P$ . G-mGBP2/mCh-mGBP2 MEFs showed the most  
170 pronounced colocalization indicating that the fluorescence tags do not affect protein localization ( $P =$   
171  $0.758 \pm 0.093$ ). Confocal images revealed a high correlation of G-mGBP2 positive VLS with mCh-  
172 mGBP1 ( $P = 0.516 \pm 0.132$ ) and mCh-mGBP3 VLS ( $P = 0.65 \pm 0.121$ ). mCh-mGBP5 ( $P = 0.108 \pm$   
173  $0.104$ ) and mCh-mGBP6 ( $P = 0.338 \pm 0.126$ ) scarcely overlapped with G-mGBP2. Thus, the subcellular  
174 reservoir of mGBP1, mGBP2 and mGBP3 differed from mGBP6, whereas mGBP5 showed no  
175 compartmentalization.

## Supplement C

176 To elucidate whether the colocalization of mGBPs is due to specific protein interactions, MFIS-hetero-  
177 FRET measurements were performed using G-mGBP2 as donor and mCh-mGBPx as acceptors (Fig. 4).  
178 In the FRET analysis GFP and mCherry fluorescence intensities ( $F_G$  and  $F_R$ ) and the mean  
179 fluorescence-weighted donor lifetime  $\langle\tau_D\rangle_f$  were determined for each pixel (Fig. 4A). By displaying the  
180 frequency of pixels in color scales for the two localizations (red: cytosol, green: VLS), the VLS-  
181 population exhibits a correlated shift in the MFIS 2D-histogram of the FRET indicators  $F_G/F_R$  and  $\langle\tau_D\rangle_f$   
182 towards smaller values with respect to the population in the cytosol. This is a clear indicator for the  
183 presence of hetero-FRET, which proves the interaction between molecules. Furthermore, GFP  $r_D$  was  
184 plotted versus  $\langle\tau_D\rangle_f$  as well as the G-mGBP2 concentration ( $C_{G-mGBP2}$ ) derived from  $F_G$  (see Methods,  
185 section 12) (Fig. 4B, C, S4A). A  $\langle\tau_D\rangle_f - r_D$  diagram is essential to determine homo- and hetero-  
186 oligomerization between mGBPs sensed by hetero- and homo-FRET. Fig. 4B illustrates the  
187 interpretation of a  $\langle\tau_D\rangle_f - r_D$  diagram based on the Perrin equation to visualize the effects on a donor-  
188 reference data set (green circle) by selective hetero- (red sphere) or homo-FRET (yellow sphere) or  
189 simultaneous homo- and hetero-FRET (orange sphere). Comparing G-mGBP2 MEFs (Fig. 4C) with G-  
190 mGBP2/mCh-mGBP2 MEFs, both homo- and hetero-FRET were visible for the latter cells indicated by  
191 a simultaneous reduction of  $\langle\tau_D\rangle_f$  and an increase of  $r_D$ . Moreover, analyzing the cells individually, the  
192 anisotropy dropped with increasing G-mGBP2 concentrations. The variation of mGBP2 concentrations  
193 between individual cells allowed the estimation of the spatially resolved apparent dissociation constant  
194 ( $K_{D,app}$ ) of the mGBP2 homomultimer of approx. 9  $\mu\text{M}$  in the VLS (Fig. 4C, upper right panel, black  
195 curve). Note that any interactions interfering with G-mGBP2 homomerization will result in a  $K_{D,app}$ -  
196 curve shifted upwards (purple curve).

197 To attain an overview of all experimental data, we computed the averaged values of  $\langle\tau_D\rangle_f$  and  
198 fluorescence intensity weighted anisotropy  $\langle r_D \rangle_{loc}$  for all cells of the specified FRET pair (Fig. 4C,  
199 lower panel). Both in cytosol and in VLS, the strongest fluorescence lifetime reduction compared to the  
200 donor-only sample could be measured for combinations of G-mGBP2 with mCh-mGBP2 and to a lesser  
201 extent for mCh-mGBP1 and mCh-mGBP3 (Fig. 4A, C), proving that mGBP1, 2, and 3 do not only  
202 colocalize but also directly interact. Although no detectable lifetime reduction could be observed

## Supplement C

203 between G-mGBP2 and mCh-mGBP5, data showed a higher anisotropy compared to the donor  
204 reference, indicating interference of mGBP5 with mGBP2 homomerization (Fig. 4C). No fluorescence  
205 lifetime reduction (Fig. 4C, left panel) or interaction-induced anisotropy increase (Fig. 4C, right panel)  
206 could be observed for mGBP2 and mGBP6 coexpressing cells.

207 To elucidate the reason for the donor lifetime reduction in VLS by determining the fraction of FRET-  
208 active complexes ( $x_{FRET}$ ) together with their FRET properties given by the rate constants of FRET  
209 ( $k_{FRET}$ ), pixel-integrated MFIS-FRET analysis was applied by computing the FRET-induced donor-  
210 quenching decay  $\varepsilon_{mix}(t)$  (eqs. 1-5) to graphically display the FRET effect (Fig. 4D). The larger drop of  
211  $\varepsilon_{mix}(t)$  (Fig. 4D, upper panel) directly shows the difference in  $x_{FRET}$  which proves that more interacting  
212 mGBP2 complexes reside in the VLS than in the cytosol. Next the characteristic  $k_{FRET}$  of both  
213 populations in the cytoplasm and the VLS was determined. The formally fitted decay curves (equations  
214 1-5) of FRET-active complexes plotted in an  $\varepsilon_{(D,A)}(t)$  diagram (Fig. 4D, lower panel) clearly differ for  
215 cytosol and VLS suggesting a higher degree of multimerization of mGBP2 in VLS. The  $\varepsilon_{mix}(t)$ -curve of  
216 a representative cell expressing G-mGBP2/mCh-mGBP6 (Fig. 4E) had random fluctuations around 1,  
217 which is consistent with the data in Figure 4A and C showing no FRET events and confirms the absence  
218 of heteromeric complexes.

219 In summary, in the cytosol and VLS mGBP2 forms homo-multimers and hetero-multimers with mGBP1  
220 and mGBP3, but not with mGBP6.

221 *Colocalization and hetero-FRET studies of mGBPs at the PVM of T. gondii.*

222 Individual members of the mGBP family are able to recruit to the PVM (Degrandi et al., 2007). To  
223 investigate the colocalization of several mGBPs at the PVM, G-mGBP2/mCh-mGBPx MEFs were  
224 infected with *T. gondii*. (Fig. 5). A colocalization of all investigated mGBPs with mGBP2 could be  
225 detected at distinct PVMs for each pairwise combination of proteins.

226 To investigate whether the colocalized mGBPs interact at the PVM, MFIS-FRET measurements were  
227 applied in G-mGBP2/mCh-mGBPx MEFs (Fig. 6). A strong decrease of both FRET indicators, GFP  
228 fluorescence lifetimes  $\langle \tau_D \rangle_f$  and intensity ratio  $F_G/F_R$ , could be detected in the cytosol and at the PVM

## Supplement C

229 of G-mGBP2/mCh-mGBP1 and G-mGBP2/mCh-mGBP2 MEFs and, to a lesser extent, in G-  
230 mGBP2/mCh-mGBP3 MEFs (Fig. 6A, B).

231 For individual cells, MFIS diagrams plotting the  $r_D$  values against donor lifetimes  $\langle \tau_D \rangle_f$  and G-mGBP2  
232 concentrations were generated (Fig. 6B upper panels, Fig. S4B). The  $K_{D,app}$ -curves describing the  
233 relationship between  $r_D$  and  $C_{G-mGBP2}$  in uninfected cells (Fig. 4C) fitted also very well to the infected  
234 situation (Fig. 6B). The averaged values of  $\langle \tau_D \rangle_f$ ,  $\langle r_D \rangle_{loc}$  and  $C_{G-mGBP2}$  over individual cells are depicted  
235 in Figure 6B (lower panels). An even stronger reduction in  $\langle \tau_D \rangle_f$  was observed at the PVM for  
236 combinations of G-mGBP2 with mCh-mGBP2 and to a lesser extent with mCh-mGBP1 and mCh-  
237 mGBP3 as compared to the VLS in uninfected cells (Fig. 4C), proving that the observed colocalization  
238 at the PVM (Fig. 5) enables direct protein interactions. For G-mGBP2/mCh-mGBP5 MEFs the situation  
239 is more complex: in the cytosol the anisotropy was slightly increased but the donor lifetime was  
240 unchanged, whereas at the PVM an increase in anisotropy was absent (Fig. 6B, lower right panel). In G-  
241 mGBP2/mCh-mGBP6 MEFs no interactions were detected, neither in the cytosol nor at the PVM.

242 The FRET-related donor quenching  $\varepsilon_{mix}(t)$  of one representative G-mGBP2/mCh-mGBP2 cell (Fig. 6C)  
243 exhibited a larger drop, which indicates a higher  $x_{FRET}$ , i.e. more interacting protein complexes were  
244 located at the PVM compared to VLS in uninfected cells (Fig. 4D). Nevertheless, their slopes ( $k_{FRET}$ ) of  
245  $\varepsilon_{(D,A)}(t)$  are comparable within the precision of the analysis (Fig. 6C, green dashed line), suggesting an  
246 unchanged local environment in the oligomer. Furthermore, the  $\varepsilon_{mix}(t)$  diagram for one representative  
247 G-mGBP2/mCh-mGBP6 cell revealed no interaction between these mGBPs.

248 In conclusion, mGBP2, besides its homo-interaction, directly interacts with mGBP1, and, to a lesser  
249 extent, with mGBP3 at the PVM. Although other mGBPs, such as mGBP5 and mGBP6 were recruited  
250 to the same PVMs, no direct interaction could be detected suggesting the formation of specific mGBP  
251 supramolecular complexes.

252 *Quantitative species-resolved pixel-integrated MFIS-FRET analysis of mGBPs multimers.*

## Supplement C

253 In addition to the formal analysis by eqs. 1-5 (Fig. S5) of the hetero-FRET data, an additional inspection  
254 of the time-resolved donor anisotropy ( $r_D(t)$ ) (Fig. 7A) revealed that cells with a higher mGBP2  
255 concentration ( $C_{\text{mGBP2}}$ ) exhibited a larger drop in initial anisotropy, which is evidence for ultrafast  
256 depolarization processes due to the formation of densely packed mGBP2 homo-oligomers with multiple  
257 GFPs. These processes were too fast to be resolved by hetero-FRET analysis (Fig. 6C), but combining  
258 both homo- and hetero-FRET, global pattern based, pixel-integrated MFIS-FRET analysis could be  
259 performed to resolve the individual mGBP species and to characterize the composition of FRET-active  
260 homo- and hetero-complexes of mGBP2 (eqs 6-7) for the distinct localizations. The obtained species  
261 fraction of mGBP2 monomers, homo- or hetero-dimers and oligomers are displayed in Figure 7B. The  
262 homo- and hetero-dimer formation is very similar in G-mGBP2 MEFs and G-mGBP2/mCh-mGBP1, 2  
263 or 3 MEFs as expected for the highly conserved GTPase-domains of mGBPs. Dimeric complexes are  
264 primarily formed with a small fraction of monomers in the cytosol (Fig. 7B, middle panel, see methods,  
265 eq. 11). The obtained  $K_{D,dim}$  of  $\sim 24$  nM is close to previous biochemical studies (Kravets et al., 2012). In  
266 the VLS an equilibrium of mGBP dimers and oligomers existed which was shifted towards oligomers  
267 with increasing protein concentration so that, the fraction of oligomers at the PVM is even higher than  
268 in the VLS. However, the dissociation constants for oligomerization  $K_{D,oligo}$  differ significantly between  
269 the mGBPs:  $70 \mu\text{M}$  for G-mGBP2/mCh-mGBP1,  $8 \mu\text{M}$  for G-mGBP2/mCh-mGBP2 and  $208 \mu\text{M}$  G-  
270 mGBP2/mCh-mGBP3 (Fig. 7B, lower panel).

271 Global analysis of G-mGBP2 MEFs and G-mGBP2/mCh-mGBP2 MEFs, revealed the heterogeneity in  
272 size of the mGBP2 oligomers via the broad distribution of FRET rate constants for small and large  
273 oligomers,  $k_{Olig,s}$  and  $k_{Olig,l}$ , respectively (Fig. 7C). While  $k_{Olig,s}$  did not change with increasing protein  
274 concentration,  $k_{Olig,l}$  increased and reached a saturation level of  $\sim 15 \text{ ns}^{-1}$  at  $\sim 50 \mu\text{M}$  (Fig. 7C, red line),  
275 which is expected for a maximal local packing of FRET acceptors around the donor (see methods  
276 section 14) and proved the growth of oligomers. Notably FRET senses only the local environment in a  
277 distance range limited to  $\sim 10$  nm, however the continuous increase in brightness suggests also the  
278 formation of larger oligomers. Therefore we introduced scanning fluorescence intensity distribution  
279 analysis (FIDA) ((Kask et al., 2000), methods section 15) to determine the mean number and brightness

## Supplement C

280 of the large oligomers for all pixels of the PVM in one infected MEF. The obtained oligomer brightness  
281 allowed us to derive the mean number of mGBP2 units in an oligomer using the specific brightness of  
282 one GFP under these measurement conditions. With increasing local mGBP2 concentration, scanning  
283 FIDA suggests also an increasing oligomer size (Fig. 7D). The mean number of mGBP2 monomer units  
284 in the oligomer ranges between 1000 and 6000 at the PVM. Remarkably the FRET rate constants in  
285 large oligomers  $k_{Olig,s}$  saturated at approximated 2000 monomer units, which corresponds to a total local  
286 concentration of mGBP2 monomer units of  $\sim 30 \mu\text{M}$  (Fig. 7D).

287 In summary, with increasing protein concentration the fraction of mGBP2 dimers decreases due to the  
288 formation of large oligomers of heterogeneous size. The formation of mGBP2 homo-oligomers is  
289 preferred over heteromers with mGBP1 and mGBP3 as  $K_{D,oligo}$  dropped by a factor of 9 and 25,  
290 respectively. The mean size of large mGBP2 oligomers can reach up to several thousand monomer  
291 units.

### 292 *mGBP2 directly targets the parasite membrane*

293 mGBP2 was shown to rapidly accumulate at the PVM after active invasion of the parasite in IFN $\gamma$   
294 activated cells (Degrandi et al., 2013). To further investigate the spatio-temporal behavior of mGBP2,  
295 3D live cell imaging was performed in mGBP2<sup>-/-</sup> MEFs stably expressing G-mGBP2 or mCh-mGBP2  
296 (Fig. 8 and Video S1-3). mGBP2 localized in VLS of heterogeneous size, morphology, and velocity  
297 within the cytosol. In IFN $\gamma$  stimulated uninfected cells the diameter of VLS reaches up to several  
298 microns. No obvious directional movement could be observed (Video S1). After *T. gondii* infection of  
299 IFN $\gamma$  stimulated MEFs, mGBP2 accumulated rapidly at the PVM (Fig. 8A, B and Video S1). Image  
300 analysis revealed that accumulation initiated simultaneously at different sites around the PVM (Fig.  
301 8B). Quantification of the overall G-mGBP2 fluorescence in regions containing the PVM and the  
302 remaining cell revealed a constant reduction of the cytosolic and VLS G-mGBP2 concentrations after  
303 infection, paired with a reciprocal increase at the PVM (Fig. 8C). Thus, accumulation of mGBP2 at the  
304 PVM occurs by redistribution of the protein, leading to a depletion of mGBP2 reservoirs and a  
305 reduction of the number of VLS (Fig. 8D) within the cytosol. However, no directional movement of  
306 VLS towards the parasite could be observed (Video S1).

## Supplement C

307 After accumulation of mGBP2 at the PVM of *T. gondii*, different fates of the parasite could be observed  
308 within the recording period by live cell imaging. mGBP2 remained at the PVM for more than 16 hours  
309 without any noticeable change in PVM or parasite morphology (not shown), mGBP2 penetrated through  
310 the PVM into the vacuolar space and accumulated at the parasite membrane (Fig. 8E and Video S2), or  
311 the mGBP2-associated PVM acquired a rounded shape immediately followed by disruption of the PVM  
312 and subsequent accumulation of mGBP2 at the parasite membrane (Fig. 8F and Video S3). Importantly,  
313 the behavior of mGBP2 was independent of the mCherry or GFP fusion.

314 These observations show direct evidence that mGBP2 promotes PVM permeabilization and disruption  
315 and provide novel evidence that mGBP2 translocates into the PV space targeting the plasma membrane  
316 of the parasite, presumably delivering a direct attack on the parasite.

317

318 **Discussion**

319 The localization, molecular dynamics, interactions, and the formation of mGBP supramolecular  
320 complexes in the context of defense against *T. gondii* could be directly visualized in living cells using  
321 MFIS and live cell imaging within this study. Our data demonstrate that GTP binding and hydrolysis as  
322 well as membrane anchoring enable the pre-assembly of multimeric complexes containing mGBP2 in  
323 VLS. mGBP2/mGBP2, mGBP2/ mGBP1 and mGBP2/mGBP3 complexes in the form of dimers and  
324 multimers with distinct composition are recruited at considerably high concentrations (10 - 200  $\mu$ M) to  
325 the PVM of *T. gondii*. Moreover, the GTPase activity and isoprenylation of mGBP2 are crucial for the  
326 control of *T. gondii* proliferation within the PV. Eventually, mGBP2 multimers target the plasma  
327 membrane of *T. gondii*, thus establishing the immune function of GBPs to directly attacking  
328 intracellular pathogens.

329 To extract structural information from the MFIS-FRET data (Kalinin et al., 2012), we performed Monte  
330 Carlo sampling of the donor-acceptor conformational space of the mGBP2 dimer to compute the  
331 expected FRET parameters (methods section 11, Fig. S7B). The sterically accessible volume of flexibly  
332 attached fluorescent proteins (green (GFP) and red (mCherry)) are depicted as fuzzy clouds. The  
333 prediction that more than 60% of all D-A configurations are FRET-inactive due to their large distances  
334 between the fluorophores is confirmed by the formal MFIS-FRET analysis (Fig. S5D). Our data argue  
335 that GTP binding is a prerequisite to induce dimer-and multimerization of mGBP2 in living cells.  
336 Indeed, the simulated FRET parameters of the mGBP2 homodimer (Fig. S7B-D) interacting via the  
337 GTPase domains are in good agreement with MFIS pixel integrated analysis (Fig. 4D, 6C, Fig. S5).  
338 Moreover, the K51A mutant, which is predicted to be predominantly nucleotide-free (Kravets et al.,  
339 2012), shows higher anisotropy values compared to WT, is entirely delocalized in the cytosol, and is  
340 monomeric in living cells (this study). However, GTPase-domain dimerization is not sufficient to  
341 determine the targeting of mGBP2 to the PVM.

342 Interestingly, individual murine and human GBPs (hGBPs) harbor C-terminal CaaX motifs (GBP1,  
343 GBP2, GBP5), targeting them for isoprenylation, which provides anchorage to different membranous

## Supplement C

344 compartments distributed within the host cell (Degrandi et al., 2007, Britzen-Laurent et al., 2010, Vestal  
345 et al., 2000). As described for hGBP1, the dimerization of the GTPase-domains enables contact  
346 formation between the two C-terminal  $\alpha$ 13 helices resulting in a juxtaposition which is crucial for their  
347 membrane localization through the attached farnesyl groups (Vöpel et al., 2014). The purified CaaX  
348 mutant of mGBP2 (C586S) shows GTP binding and hydrolysis properties as well as nucleotide  
349 dependent dimerization like the WT protein (Fig. S2). However, the C586S mGBP2 mutant renders the  
350 protein non-functional and it is found ubiquitously within the cytosol. Noteworthy, the isoprenylation  
351 mutant C586S shows similar localization and anisotropy values as the K51A mutant in living cells, also  
352 indicating a monomeric species. Altogether these studies suggest an assembly mechanism for mGBP2  
353 complexes in living cells that connects the GTPase activity of mGBP2 with membrane association  
354 leading to the stabilization of mGBP2 multimers, which is essential for its biological function.  
355 Moreover, MFIS measurements with high-precision FRET and brightness analysis allowed us to  
356 characterize the dynamic equilibrium between mGBP2 multimers. Their size distribution is  
357 heterogeneous ranging from dimers to large multimers (Fig. 7B and C). The dependence of FRET rate  
358 constants on the mGBP2 concentration and their saturation level proves dense packing of the mGBP2  
359 protomers in multimers (methods section 14) as suggested for the related mechanochemical GTPase  
360 dynamin forming large helical oligomers (Faelber et al., 2011). While FRET characterizes the molecular  
361 environment of GFPs, scanning FIDA shows that the average number of mGBP2 units in the oligomers  
362 can reach several thousands. Considering the predicted size of the mGBP2 monomer ( $\sim 4 \times 6 \times 12$  nm,  
363 according to PDB-ID 1F5N of hGBP1), it is expected that the oligomers should reach a size of several  
364 hundred nanometers. Remarkably, confocal live cell imaging (Fig. 8E and Video S2) resolves the  
365 enrichment of mGBP2 at the PV membrane resulting in a rough surface with elongated very bright  
366 features, that are sufficiently large to be resolved by far field confocal microscopy.

367 Fig. 9 provides a scheme derived from the observed mGBP interactions in living cells with molecular  
368 resolution at various stages after *T. gondii* infection. Our hetero-FRET data of MFIS measurements  
369 clearly reveal interactions of mGBP2 in multimers with itself, mGBP1, and, to a lesser extent, with  
370 mGBP3, but not with mGBP6. However, the interplay between mGBP2 and mGBP5 is different. The  
371 two proteins can be coprecipitated (Fig. S6), but the complex shows no FRET (Fig. 4 and 6). In the

## Supplement C

372 cytosol and VLS, the observation that fluorescence anisotropy of G-mGBP2 increased while its donor  
373 lifetime remained unchanged suggests an interaction of mGBP2 and mGBP5 via adaptor molecules so  
374 that they are not in close proximity and hence FRET inactive. It is noteworthy that, upon infection,  
375 oligomerization and accumulation of the mGBPs in VLS is reversible, so that the VLS serve as protein  
376 reservoir to accomplish a fast attack of the parasite after infection.

377 Both mGBP1 and mGBP2 have been implicated in *T. gondii* defense in single gene deficient mice  
378 (Degrandi et al., 2013, Selleck et al., 2013). Since mGBP1 still recruits to *T. gondii* in mGBP2<sup>-/-</sup> cells  
379 (Degrandi et al., 2013), the high level of colocalization and interaction between mGBP1 and mGBP2  
380 and their important roles in *T. gondii* control strongly argue for a cooperative effect at the PVM of *T.*  
381 *gondii*. Interestingly, reconstitution of mGBP2 in mGBP<sup>chr3</sup>-deleted MEFs did not allow a sufficient  
382 control of *T. gondii* replication, while reconstitution of mGBP1 partially restored the WT phenotype  
383 (Yamamoto et al., 2012). Although more studies on the hierarchy of mGBPs are needed to fully  
384 understand the individual roles of each GTPase, this might hint that mGBP2 acts in concert with  
385 mGBP1 and possibly other mGBPs to exert its molecular antiparasitic activity.

386 The dissociation constant  $K_{D,oligo}$  of mGBP2/mGBP3 heteromers is 25 times larger than that of  
387 mGBP2/mGBP2 homomers. Thus, it is not surprising that mGBP3 colocalized only partially in the  
388 same VLS (Fig. 3, Fig. 9). Strikingly, mGBP6, which also localizes in VLS and recruits to the PVM of  
389 *T. gondii*, is predominantly found in different VLS and shows no interaction with mGBP2 by FRET and  
390 co-IP. The different localizations of mGBP multimers argue for distinct individual functional roles in *T.*  
391 *gondii* immunity to be elucidated in the future.

392 Recently, an essential function for the cassette of autophagy proteins, including Atg7, Atg3, and the  
393 Atg12-Atg5-Atg16L1 complex was demonstrated in cellular anti-*T. gondii* immunity by facilitating IRG  
394 and GBP recruitment to the PVM (Ohshima et al., 2014, Choi et al., 2014, Haldar et al., 2014). This  
395 function appears to be independent of the classical autophagy degradation pathway (Zhao et al., 2008),  
396 but rather to play a role in the delivery of effectors to pathogen containing vacuoles (Selleck et al.,  
397 2013). Performing live cell imaging and MFIS analysis it could be shown that mGBP2 is loaded on the  
398 PVM of *T. gondii* within a few minutes post-infection, assembling a machinery of supramolecular

## Supplement C

399 complexes with mGBP1 and mGBP3. Additionally, we unambiguously show that mGBP2 directly  
400 targets the membrane of the parasite after penetration or disruption of the PVM. Interestingly, GBP  
401 proteins, especially mGBP2, were shown to stimulate caspase-11-dependent pyroptosis in macrophages  
402 infected with Gram-negative bacteria which reside in vacuoles. There, GBP dependent induction of  
403 lysis of the pathogen-containing vacuoles and release of cytoplasmic LPS leads to the activation of the  
404 noncanonical inflammasome (Pilla et al., 2014, Meunier et al., 2014). Strikingly, a novel study suggests  
405 a direct bacteriolytic function of mGBPs, releasing pathogen-associated molecular patterns into the  
406 cytosol, resulting in activation of the AIM2 inflammasome (Man et al., 2015, Meunier et al., 2015).  
407 Thus, based on our observations, it is likely that mGBPs promote not only disruption of the PVM, but  
408 also directly induce lysis of the plasma membrane of *T. gondii*. The hierarchy of events which might be  
409 involved in *T. gondii* targeting and elimination, such as autophagic degradation (Choi et al., 2014)  
410 and/or inflammasome activation (Ewald et al., 2014, Gorfou et al., 2014, Meunier et al., 2014, Meunier et  
411 al., 2015), have yet to be determined.

412 These studies define mGBP2 as an important effector molecule of innate immunity in the host parasite  
413 interaction with apicomplexan parasites such as *T. gondii*, by providing seminal insight into its  
414 supramolecular assembly and cellular function. Further studies will be performed to address the  
415 question how this information can be exploited for anti-parasitic therapy.

## Supplement C

### 416 **Material and Methods**

#### 417 **1. Expression Constructs**

418 The WT ORF of mGBP2 (NCBI accession numbers: mGBP-2, NM\_010260.1) was subjected to site  
419 directed mutagenesis (QuikChange II Mutagenesis kit, Stratagene) for generation of GTPase mutants  
420 R48A, K51A, E99A and D182N (Kravets et al., 2012) and isoprenylation mutant C586S (Degrandi et  
421 al., 2013) in the pEGFP-C2 plasmid (Clontech). The respective genes were then cloned into the pWPXL  
422 plasmid (Trono lab) as N-terminal GFP-fusion constructs. For the cloning of mCherry constructs, the  
423 pWPXL plasmid was modified by exchanging of the gene for GFP by the gene for mCherry. The ORFs  
424 of mGBP1 (NM\_010259.2), mGBP2, mGBP3 (NM\_001289492.1), mGBP5 (NM\_153564.2), mGBP6  
425 (NM\_194336.2) were then cloned into the modified pWPXL plasmid as N-terminal mCherry-fusion  
426 constructs. The lentiviral envelope vector pLP/VSVG (Invitrogen) and the packaging vector psPAX2  
427 (Trono lab) were used for the lentiviral genetic transfer.

#### 428 **2. Cell culture and transduction**

429 MEFs were cultured in Dulbecco's modified Eagle's medium (DMEM, Invitrogen/Gibco) supplemented  
430 with 10% (v/v) heat-inactivated low endotoxin fetal bovine serum (FBS, Cambrex), 100 U/ml penicillin,  
431 100 µg/ml streptomycin, 2 mM L-glutamine (Biochrom) and 0.05 mM β-mercaptoethanol  
432 (Invitrogen/Gibco). Human foreskin fibroblasts (HS27, ATCC CRL-1634) were hold in culture in  
433 Iscove's Modified Dulbecco's Medium (IMDM, Invitrogen/Gibco) with the same supplementations.  
434 293FT cells were cultivated in DMEM supplemented with 10% FBS, 100 U/ml penicillin, and 100  
435 µg/ml streptomycin. All recombinant lentiviruses were produced by transient transfection of 293FT  
436 cells according to standard protocols. Briefly, subconfluent 293FT cells were cotransfected with 20 µg  
437 of a plasmid vector, 10 µg of psPAX2, and 5 µg of pLP/VSVG by calcium chloride precipitation in FBS  
438 free medium. After 6 h medium was changed (10% FBS), and supernatants with recombinant lentivirus  
439 vectors were harvested 48 h later. Alternatively, the trasfection was performed utilizing the jetPRIME®  
440 trasfection reagent (Polyplus) in medium supplemented with 10% FBS. MEFs were seeded in 24 well  
441 plates (Corning Incorporated) and transduced with 600 µl of lentivirus with 25 µg Polybrene

## Supplement C

442 (Millipore). After 4 h of incubation medium was changed. The transduction efficacy was analyzed by  
443 flowcytometry. Subsequently, GFP or GFP/mCherry positive cells were sorted and cultivated.

444 Tachyzoites from *T. gondii* strain ME49 were maintained by serial passage in confluent monolayers of  
445 HS27 cells. After infection of fibroblasts, parasites were harvested and passaged as described previously  
446 (Degrandi et al., 2007).

### 447 **3. Infection of murine MEFs with *T. gondii***

448 Cells were stimulated with 200 U/mL IFN $\gamma$  (R&D Systems) 16 h before infection. For  
449 immunofluorescence, MEFs were cultured in 24-well plates (Falcon, BD Biosciences) on cover slips ( $\emptyset$   
450 13 mm, VWR International) and inoculated with freshly harvested *T. gondii* at a ratio of 50:1. To  
451 remove extracellular parasites, cells were washed with PBS.

### 452 **4. Immunofluorescence analysis**

453 Cells were fixed in 4% paraformaldehyde (PFA, Sigma-Aldrich) permeabilized with 0.02% saponin  
454 (Calbiochem-Merck) and blocked in 0.002% saponin with 2% goat serum (DaKoCytomation). The outer  
455 membrane of *T. gondii* was visualized by anti-SAG1 (Abcam) at a concentration of 1/700. As secondary  
456 reagents, 1/200 concentrated Cy2-conjugated goat anti-rabbit IgG and Cy3-conjugated goat anti-mouse  
457 IgG plus IgM (Jackson ImmunoResearch Laboratories) were used. Nuclei were counterstained with  
458 1/2500 4',6-diamidino-2-phenylindole (DAPI, Invitrogen). The cover slips were fixed in fluorescence  
459 mounting medium (Fluoromount-G, Southern Biotechnology Associates). Fluorescence was visualized  
460 using a LSM780 confocal microscope (Zeiss). Image analysis and processing was performed with the  
461 ZEN (Zeiss) and Imaris (Bitplane) softwares.

### 462 **5. Confocal live cell imaging**

463 Live cell imaging was performed using an LSM780 confocal microscope (Zeiss, Germany) at 37°C with  
464 8% CO $_2$  and humidity saturated air. Cells were cultured and imaged on imaging dishes CG (MoBiTec,  
465 Germany) with Phenol-free cell culture media. Image analysis was performed with the software ZEN  
466 (Zeiss), Imaris (Bitplane) and AutoquantX3 (MediaCy/Bitplane).

## Supplement C

### 467 **6. MFIS setup**

468 MFIS experiments (Kudryavtsev et al., 2007, Weidtkamp-Peters et al., 2009) were performed with a  
469 confocal laser scanning microscope (Olympus FV1000, IX81 inverted microscope) additionally  
470 equipped with a single photon counting device with picosecond time-resolution (PicoQuant Hydra Harp  
471 400, PicoQuant, Berlin, Germany). GFP was excited at 485 nm with a linearly polarized, pulsed (32  
472 MHz) diode laser (LDH-D-C-485) at 0.4  $\mu$ W at the objective (60x water immersion, Olympus  
473 UPlanSApo NA 1.2, diffraction limited focus). mCherry was excited at 559 nm with a continuous-wave  
474 laser (FV1000) at 0.54  $\mu$ W at the objective. The emitted light was collected in the same objective and  
475 separated into its perpendicular and parallel polarization (PBS 101, Thorlabs). GFP fluorescence was  
476 then detected by SPADs (PD5CTC, Micro Photon Devices, Bolzano, Italy) in a narrow range of its  
477 emission spectrum (bandpass filter: HC520/35 (AHF, Tübingen, Germany)). mCherry fluorescence was  
478 detected by HPDs (HPMC-100-40, Becker&Hickl, Berlin, Germany), of which the detection  
479 wavelength range was set by the bandpass filters (HC 607/70, AHF). Images were taken with 20  $\mu$ s  
480 pixel time and a resolution of 276 nm/pixel. With 485nm excitation, series of 40-100 frames were  
481 merged to one image and further analyzed with custom-designed software (Widengren et al., 2006).

### 482 **7. Pixel-wise MFIS analysis of fluorescence parameters**

483 From the recorded GFP ( $S_G$ ) and mCherry ( $S_R$ ) signal intensities, background intensities of the regions  
484 where no cells localize were subtracted to determine fluorescence intensities of GFP ( $F_G$ ) and mCherry  
485 ( $F_R$ ) respectively. To determine fluorescence anisotropy ( $r_D$ ) and fluorescence-weighted donor lifetimes  
486 ( $\langle \tau_D \rangle_f$ ) in each pixel, the histograms presenting the decay of fluorescence intensity after the excitation  
487 pulse were built with 256 bins and 128 ps per bin. The fitting procedures were described previously  
488 (Stahl et al., 2013, Kravets et al., 2012).

### 489 **8. Formal pixel-integrated MFIS-FRET analysis**

490 In each obtained MFIS image, pixels in the VLS and in the cytosol in uninfected cells, and pixels at the  
491 PVM and in the cytosol in infected cells were separately selected according to fluorescence photon  
492 number (Fig. 1a, 2a, 4a and 6a). Photons from each pixel selection were integrated to an intensity decay

## Supplement C

493 histogram with 1024 bins and 32 ps per bin. The pixel-integrated histograms were formally fitted to  
494 quantitatively determine FRET parameters. In the model, fluorescence decay of FRET sample ( $f_{mix}(t)$ ) is  
495 the sum of FRET-quenched donor decay ( $f_{(D,A)}(t)$ ) weighted by its species fraction  $x_{FRET}$  and  
496 unquenched donor decay ( $f_{(D,0)}(t)$ ) weighted by  $(1 - x_{FRET})$ :

$$f(t) = (1 - x_{FRET}) \cdot f_{(D,0)}(t) + x_{FRET} \cdot f_{(D,A)}(t) \quad (1)$$

497 Here,  $f_{(D,0)}(t)$  could be pre-determined from donor-only measurements using a bi-exponential fit model:

$$f_{(D,0)}(t) = \sum_m x_{D0}^{(m)} \cdot \exp(-t \cdot k_{D0}^{(m)}) \quad (2)$$

498 in which  $m=2$  because fluorescent proteins in living cells usually show a bi-exponential decay (Suhling  
499 et al., 2002). Fit parameters in  $f_{(D,0)}(t)$  include two normalized pre-exponential factors  $x_{D0}^{(m)}$  ( $\sum x_{D0}^{(m)} = 1$ )  
500 and two decay rate constants,  $k_{D0}^{(m)}$ . These pre-determined parameters from donor-only measurements  
501 were then set as global restraints. The quenched donor decay  $f_{(D,A)}(t)$  in equation (1) is given by:

$$f_{(D,A)}(t) = f_{(D,0)}(t) \cdot \exp(-t \cdot k_{FRET}) \quad (3)$$

502 where  $k_{FRET}$  is the FRET rate constant. The fitted parameters in the  $1 - k_{FRET}$  model (eq. 1-3) are  $x_{FRET}$  and  
503  $k_{FRET}$ . This formal analysis revealed that mGBPs exhibit distinct FRET features in different cellular  
504 compartments (supplementary Fig. 5).

### 505 9. $\varepsilon_{mix}(t)$ and $\varepsilon_{(D,A)}(t)$ diagrams

506 FRET-related donor quenching histogram ( $\varepsilon_{mix}(t)$ ) was plotted to directly separate different molecular  
507 species and visualize FRET efficiency in the pixel-integrated data.  $\varepsilon_{mix}(t)$  is calculated as the ratio  
508 between normalized fluorescence decay of FRET sample,  $f_{mix}(t)$ , and of donor-only sample,  $f_{(D,0)}(t)$ :

$$\varepsilon_{mix}(t) = \frac{f_{mix}(t)}{f_{(D,0)}(t)} = x_{FRET} \varepsilon_{(D,A)}(t) + (1 - x_{FRET}) \quad (4)$$

509 The drop on a  $\varepsilon_{mix}(t)$  diagram represents the species fraction of FRET-active complex,  $x_{FRET}$ .

## Supplement C

510 In equation (4),  $\varepsilon_{(D,A)}(t)$  is the ratio between  $f_{(D,A)}(t)$  (eq. 3) and  $f_{(D,0)}(t)$  (eq. 2) and describes the time-  
511 dependent occurrence of the FRET process.

$$\varepsilon_{(D,A)}(t) = \frac{f_{(D,A)}(t)}{f_{(D,0)}(t)} = \exp(-t \cdot k_{FRET}) \quad (5)$$

512 To directly compare different experiments,  $\varepsilon_{(D,A)}(t)$  diagrams were plotted in Figure 4d. A steeper slope  
513 in  $\varepsilon_{(D,A)}(t)$  diagram indicates that the experiment showed higher  $k_{FRET}$ .

### 514 10. Pattern based pixel-integrated MFIS-FRET analysis

515 To resolve three characteristic protein species, namely mGBP monomer (with specie fraction  $x_{mono}$ ),  
516 dimer ( $x_{di}$ ) and oligomers ( $x_{oligo}$ ) by analyzing time-resolved anisotropy  $r_{mix}(t)$  (eq. 6) and time-resolved  
517 FRET-induced donor decay  $\varepsilon_{mix}(t)$  (eq. 7) for homo- and hetero-FRET, respectively, both decays were  
518 fitted with a linear combination of three species-specific patterns.

519 *Homo-FRET.* The  $r_{mix}(t)$  of homo-FRET data was fitted with:

$$r_{mix}(t) = r_0 \cdot \left( x_{mono} + x_{di} \left( \int p(k_{di}) e^{-2 \cdot k_{di} \cdot t} dk_{di} \right) + x_{oligo} \left( x_s e^{-2 \cdot k_{olig,s} \cdot t} + x_l e^{-2 \cdot k_{olig,l} \cdot t} \right) \right) e^{-t/\rho_{global}} \quad (6)$$

520 Here  $p(k_{di})$  is the FRET-rate distribution of mGBP2 dimer complex as determined by the  
521 conformational sampling of the sterically allowed space (see Monte Carlo sampling of the donor-  
522 acceptor conformational space of mGBP2 dimer and supplementary Fig. 7d, e).  $k_{olig,s}$  and  $k_{olig,l}$  are  
523 formally assigned as the FRET rate constants of mGBPs oligomers of small and large sizes respectively,  
524 and  $x_s$  and  $x_l$  are their normalized fractions. It has to be considered that energy can be transferred in  
525 forward and backward direction which doubles the rate constants. The monomer is described by a  
526 constant offset, because there is no FRET. The fundamental anisotropy  $r_0$  for GFP molecules is known  
527 as 0.38. The global rotational correlation time  $\rho_{global}$  was estimated as 15 ns given the molecular weight  
528 of G-mGBP2 fusion protein. Oligomer species which produced ultrafast decay components in  $r_{mix}(t)$   
529 resulted in a drop in the initial anisotropy (Fig. 6d). With the knowledge of  $r_0$  they can be determined  
530 together with other species in homo-FRET data.

## Supplement C

531 *Hetero-FRET*. An analogous analysis was applied to the hetero-FRET data. The  $\varepsilon_{mix}(t)$  (equation 4) was  
532 fitted with:

$$\varepsilon_{mix}(t) = x_{mono} + x_{di} \left( \int p(k_{di}) e^{-k_{di}t} dk_{di} \right) + x_{oligo,s} e^{-t k_{oligo,s}} \quad (7)$$

533 in which  $x_{oligo,s}$  denotes the species fraction of small oligomers. In the case of hetero-FRET, donor  
534 molecules in large oligomers (with species fraction  $x_{oligo,l}$ ) could be strongly quenched by nearby  
535 acceptors up to nearly 100% and thus became irresolvable owing to the finite width of the instrument  
536 response function. Therefore the information of large oligomers in hetero-FRET data needed to be  
537 recovered according to the homo-FRET data. In the latter, the species fractions of small and large  
538 oligomers were found equal in various compartments. Based on the relation  $x_{oligo,s} = x_{oligo,l}$  the large  
539 oligomer fractions in hetero-FRET data were extrapolated. Moreover, such a coherent behavior between  
540 small and large oligomers indicated that they have a common origin; and the broad distribution of their  
541 rate constants showed that oligomers may consist of a variety number of units. Hence, it is more  
542 meaningful to combine both oligomer species and generally sort protein species as monomer, dimer and  
543 oligomer as displayed in Figure 6c. The fits were performed by custom software programmed in  
544 MATLAB.

### 545 **11. Monte Carlo sampling of the donor-acceptor conformational space of mGBP2 dimer**

546 Based on the hGBP1 crystal structure (Prakash et al., 2000a) homology models of the G-mGBP2 (PDB-  
547 ID: 1F5N, 4EUL) and mCherry-mGBP2 fusion protein (PDB-ID: 1F5N, 2H5Q) (supplementary Table  
548 2) were constructed using MODELLER (Fiser and Sali, 2003). The homology models were protonated  
549 using PDB-ID 2PQR (Dolinsky et al., 2007). Then the protonated full-length protein models were  
550 mapped to a reduced representation solely consisting of the C-, C $_{\alpha}$ -, N-, O- and the hydrogen atoms  
551 forming the NH-O bonds. The repulsion between the atom pairs (O, N), (C, O) and (C, N) were  
552 modeled as repulsive quadratic potential (Kalinin et al., 2012) and the existing hydrogen bonds as  
553 simple scaled attractive potential (1/r) preserving secondary structural elements. The sampling was  
554 performed on the  $\varphi$  and  $\psi$  torsion angles. In each iteration step the torsion angle of one amino acid was  
555 changed by random value taken from a Gaussian-distribution with a width of 0.025 rad. The sampling

## Supplement C

556 of the conformational space was restricted to the linkage region. Thus, only the internal coordinates of  
557 the connecting linker were altered while the internal coordinates of the beta-barrels as well as the  
558 internal coordinates of the mGBP2 model were kept constant. Given the sampled conformation of the  
559 mCh-mGBP2 and the G-mGBP2 constructs a putative head-to-head dimer structure was created by  
560 superimposing the LG-domains onto the LG-domains in the dimer structure of hGBP1 in presence of  
561 GppNHp (PDB-ID: 2BC9) and discarding conformations with clashes (Vöpel et al., 2014). To calculate  
562 the donor-acceptor distance,  $R_{sim}$ , in every simulated structure, on each fluorophore, two C $\alpha$ -atoms on  
563 the beta-barrel (Asn122 and Asn147 on GFP, Tyr125 and Glu149 on mCherry) were chosen  
564 (supplementary Fig. 7a, supplementary Table 2), so that the connecting vector of the two atoms is a  
565 good approximation of the transition dipole. The distance between the middle points of the connecting  
566 vectors of the donor and acceptor is taken as the distance between the chromophores ( $R_{DA, sim}$ ).  
567 Supplementary Table 3 lists out the detailed calculation steps to determine the ( $R_{DA, sim}$ ) and orientation  
568 factor ( $\kappa^2$ ). For each simulated mGBP2 dimer conformation, its  $k_{di}$  value was calculated according to:

$$k_{di} = (3/2) \cdot \kappa^2 \cdot (1/\tau_{D(0)}) \cdot (R_0/R_{DA, sim})^6 \quad (8)$$

569 in which  $\tau_{D(0)} = 1/k_0$  is 2.6 ns and the Förster radius ( $R_0$ ) of GFP and mCherry is 52Å including  $\kappa^2 =$   
570 2/3.

### 571 12. Determination of mGBP protein concentrations and binding curves

572 *mGBP protein concentrations.* The protein concentration is monitored via the fluorescence intensity of  
573 the fused fluorescent proteins. The detection volume of the MFIS microscope was calibrated by  
574 Fluorescence Correlation Microscopy (FCS) measurements of Rhodamine 110 (Rh110) to determine its  
575 shape and size. The fitting model applied to the obtained FCS curve assumes a 3-dimensional Gaussian-  
576 shaped volume, and a single diffusing species including transitions to a triplet state as shown in  
577 (Weidtkamp-Peters et al., 2009). From the Rh110 diffusion time of 32  $\mu$ s and aspect ratio of 7, the  
578 detection volume  $V_{det-GFP}$  of GFP was determined as 0.5 fl. The detection volume of mCherry  $V_{det-mCherry}$   
579 is larger (0.8 fl) because of the longer wavelength. The brightness of GFP or mCherry in living cells

## Supplement C

580 was individually characterized from FCS measurements of freely diffusing fluorescent proteins in the  
581 cytosol. By fitting the same model function as in Rh110 experiment, it was found that with 0.54  $\mu\text{W}$  of  
582 559 nm laser excitation at the objective, mCherry brightness is  $Q_{mCherry} = 0.1$  kcpm in the cytosol and  
583 that with 0.4  $\mu\text{W}$  of 485 nm laser excitation, GFP brightness is  $Q_{GFP} = 0.56$  kcpm in the cytosol.

584 The average mCherry fluorescence intensity of an image with mCherry excitation was first corrected for  
585 detector dead time (Becker, 2005), and then used to calculate the mCherry concentration with the  
586 determined detection volume and the mCherry brightness.

587 The average GFP fluorescence intensity of an image with GFP excitation was first corrected for detector  
588 dead time, and then the obtained intensity ( $S_{G,G}^m$ ) was further corrected for quenching effect due to  
589 FRET:

$$S_{G,G}^u = \frac{S_{G,G}^m}{(1 - x_{FRET}) + x_{FRET} \cdot (1 - E)} \quad (9)$$

590  $S_{G,G}^u$  denotes unquenched GFP fluorescence intensity in the absence of hetero-FRET and was used to  
591 calculate the GFP concentration. In equation 10, the FRET-active species fraction ( $x_{FRET}$ ) is obtained  
592 from fitting of each measurement in pixel-integrated MFIS-FRET analysis using the  $1 - k_{FRET}$  model (eq.  
593 1-3). FRET efficiency,  $E$ , was calculated as:

$$E = 1 - \frac{\sum_m x_{D0}^{(m)} \cdot (k_{D0}^{(m)} + k_{FRET})^{-1}}{\sum_m x_{D0}^{(m)} \cdot (k_{D0}^{(m)})^{-1}} \quad (10)$$

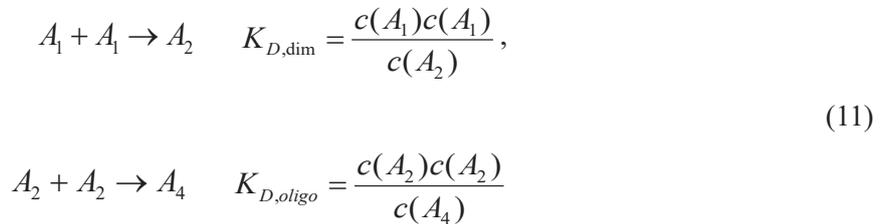
594 Please refer to Formal pixel-integrated MFIS-FRET analysis for explanations on the symbols in  
595 equations 9 and 10.

### 596 13. Determination of dissociation constants.

597 To quantify the dependence of the dimeric species fraction on the total protein concentration (initial  
598 increase, stationary phase followed by a decrease) the simplest possible model was used to approximate

## Supplement C

599 such a behavior. In this model the formation of a dimer and a subsequent formation of a tetramer  
600 formed by two dimers was assumed. The formation of a dimer and a tetramer can be described by two  
601 reactions with corresponding dissociation equilibrium constants:



602 For given of equilibrium constants and a total protein concentration  $c_T = c(A_1) + 2 \cdot c(A_2) + 4 \cdot c(A_4)$   
603 the species concentrations  $c(A_1), c(A_2), c(A_4)$  were determined numerically by solving the fourth  
604 polynomial equation  $c_T(A_1)$  by Ridders method. Given the observed species fractions the equilibrium  
605 constants were optimized by the limited memory Broyden–Fletcher–Goldfarb–Shanno (BFGS) method.  
606 This model of stepwise oligomer formation was extended by the stepwise binding of dimer in a non-  
607 cooperative fashion (i.e. all equilibrium constants are equal to  $K_{D,\text{oligo}}$ ) up to a dodecamer. If the total  
608 concentration of all oligomers (4-12) is used to display the binding isotherm, one obtains an only  
609 slightly broadened binding isotherm compared to the tetramer system. If this binding isotherm is fitted  
610 with the simpler tetramer model, a binding constant for dimer binding  $K_{D,\text{app-oligo}}$  is obtained, which is  
611 slightly (factor 1.6) larger than the simulated value.

612 As no information on the cooperativity of binding and the spatially resolved GTP concentration was  
613 available, the formation of higher order oligomers was approximated by the minimal tetramer model for  
614 the following reasons: (1) FRET only senses its local environment (i.e. a limited oligomer size) thus the  
615 contribution of each monomer unit to the measured signal decreases with increasing oligomer size. (2)  
616 This simple model reduces the number of fitting parameters to an adequate level given the spread of the  
617 data-points. To conclude, a simple model with a Langmuir binding isotherm (i.e. non-cooperative  
618 binding) describes all experiments very well. The simulation showed that the obtained apparent  
619 dissociation constant  $K_{D,\text{app-oligo}}$  is a good approximation for the true  $K_{D,\text{oligo}}$ .

## Supplement C

620 Note that the observed reduction in steady-state anisotropy ( $r_D$ ) for cells of high mGBP2 concentration  
621 as displayed in Figure 4c, was mainly due to the large drop in the initial anisotropy of their time-  
622 resolved anisotropy ( $r_D(t)$ ) as plotted in Figure 6d. Therefore the  $K_{D,app}$  value (9  $\mu\text{M}$ ) derived from  $r_D$  in  
623 fact reports the mGBP2 oligomerization processes that could produce such ultrafast depolarizing effect,  
624 and is very close to the 8  $\mu\text{M}$  obtained by fitting  $r_D(t)$  with the species-resolved model. Hence, the two  
625 independent approaches interrogating the same oligomerization process delivered very consistent  
626 results, verifying the reliability of the analyses.

627

### 628 14. Maximum FRET rate constants

629 Due to its inverse sixth-power distance dependence (eq. 12), FRET depends on molecular proximity and  
630 cannot occur between remotely located fluorescent proteins. Consequently, in large mGBP oligomers,  
631 the FRET-induced donor quenching will eventually saturate regardless of the presence of more  
632 acceptors simply because they are too distant. If assuming that the mGBP proteins are arranged  
633 homogeneously in mGBP oligomers, the maximum  $k_{FRET}$  can be estimated following the ideas of T.  
634 Förster (Förster, 1949).

635 Here, the case of a single donor is considered, the FRET rate constant  $k_{FRET}$  from the donor to  $N$   
636 surrounding acceptors is given by eq. 12 using the parameter in eq. 8.

$$k_{FRET,max} = \frac{1}{\tau_{D(0)}} \sum_{k=1}^N \left( \frac{R_0}{R_{DA,k}} \right)^6 \quad (13)$$

637 with  $R_{DA,k}$  being the distance between the donor and the  $k$ -th acceptor  $k_{FRET}$ . Assuming that the  
638 acceptors that attached on mGBPs are homogeneously distributed around the donor and closed packed  
639 with a minimum inter-fluorophore distance  $R_{min}$ , which is  $\sim 26$  Å given by the molecular dimensions of  
640 fluorescent proteins, a similar estimation of the maximum  $k_{FRET}$  as in (Förster, 1949) can be performed.

641 To determine the maximum FRET-rate at which a donor molecule is quenched by multiple acceptors  
642 some spatial properties of the donor and the acceptor have to be considered. Firstly it was assumed that

## Supplement C

643 at saturation protein concentrations the space around the donor is fully occupied by acceptors. Secondly,  
644 it was assumed that the space that is occupied by the donor cannot be occupied by the acceptor. Thirdly,  
645 it was assumed that the acceptors are located at defined positions in space. This assumption differs from  
646 Förster's analysis for a random distribution of point like acceptors around a donor. If a donor is  
647 homogenously surrounded by acceptors which are separated at least by a distance of  $R_{min}$  from the  
648 donor the FRET-rate constant is given by convenient analytical approximation:

$$\begin{aligned} k_{FRET,max} &= \frac{1}{\tau_{D(0)}} \int_{R_{min}}^{\infty} p(R_{DA}) \left( \frac{R_0}{R_{DA}} \right)^6 dR_{DA} \\ &= \frac{1}{\tau_{D(0)}} \frac{(R_0)^6}{R_{mol}^3 R_{min}^3} \end{aligned} \quad (14)$$

649  $R_{mol}$  is the mean radius of a mCherry in mCh-mGBPs fusion protein, approximated as 31 Å given the  
650 molecular structure. Therefore according to equation 13, the maximum  $k_{FRET}$  was estimated as  $\sim 15 \text{ ns}^{-1}$ .

651  
652 **15. Scanning fluorescence intensity distribution analysis (FIDA) for determination of oligomer**  
653 **size**

654 To investigate the size (composition) of mGBP2 oligomer locating at the PVM which can exceed the  
655 detectable range of FRET technique ( $> 10 \text{ nm}$ ), FIDA from (Kask et al., 2000) was adapted for imaging  
656 measurements and employed in infected G-mGBP2 expressing cells. Given the recorded photon trace in  
657 the image line of selected PVM area, 20  $\mu\text{s}$  binned new sliding with 2.5  $\mu\text{s}$  ( $1/8 \times \text{pixel time}$ ) steps  
658 intensity traces were computed. Then a corresponding 2D matrix of green versus red photon counts  
659 from all the time windows is generated and analysed by 2D FIDA. The average brightness,  $\langle Q_{oligo} \rangle$ , and  
660 average number,  $\langle N_{oligo} \rangle$ , of the mGBP2 oligomers could be determined. The average number of  
661 mGBP2 units (Fig. 7D) per oligomer  $\langle N_{mGBP2} \rangle$  is calculated as the ratio of obtained  $\langle Q_{oligo} \rangle$  to single  
662 GFP brightness  $Q_{GFP}$ :

## Supplement C

$$\langle N_{mGBP2} \rangle = \frac{\langle Q_{oligo} \rangle}{Q_{GFP}} \quad (15)$$

663 Based on this two average numbers of oligomers and mGBP2 units per pixel and knowing the excitation  
664 volume of the setup, the average mGBP2 concentration  $\langle c_{mGBP2} \rangle$  is calculated as

$$\langle c_{mGBP2} \rangle_{FIDA} = \frac{\langle N_{oligo} \rangle \langle N_{mGBP2} \rangle}{N_A \cdot V_{det}} \quad (16)$$

665 where  $N_A = 6.022 \times 10^{23} \text{ mol}^{-1}$  is the Avogadro's number and  $V_{det} = 0.5 \text{ fl}$  – excitation volume of the used  
666 laser. The mGBP2 concentration calculated from scanning FIDA was compared with that directly  
667 derived from the GFP intensity as a control. Figure S7F shows the good agreement between both  
668 methods.

669

## Supplement C

### 670 **Acknowledgements**

671 We thank Julia Hartmann and Karin Buchholz for excellent experimental assistance. Q.M and T.P.  
672 thank the International Helmholtz Research School of Biophysics and Soft Matter (IHRS BioSoft) for  
673 funding. The work was supported by grants of the Deutsche Forschungsgemeinschaft to C.A.M.S. and  
674 K.P. and the Jürgen Manchot Foundation to K.P.. All authors declare no conflict of interests.

675

676 **References**

- 677 Becker, W 2005. *Advanced time-correlated single-photon counting techniques*, Springer Science and  
678 Business Media.
- 679 Britzen-Laurent, N, Bauer, M, Berton, V, Fischer, N, Syguda, A, Reipschlager, S, Naschberger, E,  
680 Herrmann, C & Sturzl, M. 2010. Intracellular Trafficking of Guanylate-Binding Proteins Is  
681 Regulated by Heterodimerization in a Hierarchical Manner. *Plos One*  
682 **5**:doi:10.1371/journal.pone.0014246.
- 683 Choi, J, Park, S, Biering, SB, Selleck, E, Liu, CY, Zhang, X, Fujita, N, Saitoh, T, Akira, S, Yoshimori,  
684 T, Sibley, LD, Hwang, S & Virgin, HW. 2014. The Parasitophorous Vacuole Membrane of  
685 *Toxoplasma gondii* Is Targeted for Disruption by Ubiquitin-like Conjugation Systems of  
686 Autophagy. *Immunity* **40**:924-935. doi:10.1016/j.immuni.2014.05.006.
- 687 Decker, T, Stockinger, S, Karaghiosoff, M, Muller, M & Kovarik, P. 2002. IFNs and STATs in innate  
688 immunity to microorganisms. *J Clin Invest* **109**:1271-7. doi:10.1172/JCI15770.
- 689 Degrandi, D, Konermann, C, Beuter-Gunia, C, Kresse, A, Wurthner, J, Kurig, S, Beer, S & Pfeffer, K.  
690 2007. Extensive characterization of IFN-induced GTPases mGBP1 to mGBP10 involved in host  
691 defense. *J Immunol* **179**:7729-40.
- 692 Degrandi, D, Kravets, E, Konermann, C, Beuter-Gunia, C, Klumpers, V, Lahme, S, Wischmann, E,  
693 Mausberg, AK, Beer-Hammer, S & Pfeffer, K. 2013. Murine guanylate binding protein 2  
694 (mGBP2) controls *Toxoplasma gondii* replication. *Proc Natl Acad Sci U S A* **110**:294-9.  
695 doi:10.1073/pnas.1205635110.
- 696 Dolinsky, TJ, Czodrowski, P, Li, H, Nielsen, JE, Jensen, JH, Klebe, G & Baker, NA. 2007. PDB2PQR:  
697 expanding and upgrading automated preparation of biomolecular structures for molecular  
698 simulations. *Nucleic Acids Res* **35**:W522-5. doi:10.1093/nar/gkm276.
- 699 Ewald, SE, Chavarria-Smith, J & Boothroyd, JC. 2014. NLRP1 Is an Inflammasome Sensor for  
700 *Toxoplasma gondii*. *Infection and Immunity* **82**:460-468. doi:10.1128/iai.01170-13.
- 701 Faelber, K, Posor, Y, Gao, S, Held, M, Roske, Y, Schulze, D, Haucke, V, Noe, F & Daumke, O. 2011.  
702 Crystal structure of nucleotide-free dynamin. *Nature* **477**:556-60. doi:10.1038/nature10369.
- 703 Fiser, A & Sali, A. 2003. Modeller: generation and refinement of homology-based protein structure  
704 models. *Methods Enzymol* **374**:461-91. doi:10.1016/S0076-6879(03)74020-8.
- 705 Förster. 1949. Experimentelle und theoretische Untersuchung des zwischenmolekularen Übergangs von  
706 Elektronenanregungsenergie. *Z. Naturforschung* **4a**:321-327.
- 707 Gazzinelli, RT, Mendonca-Neto, R, Lilue, J, Howard, J & Sher, A. 2014. Innate resistance against  
708 *Toxoplasma gondii*: an evolutionary tale of mice, cats, and men. *Cell Host Microbe* **15**:132-8.  
709 doi:10.1016/j.chom.2014.01.004.
- 710 Ghosh, A, Praefcke, GJK, Renault, L, Wittinghofer, A & Herrmann, C. 2006. How guanylate-binding  
711 proteins achieve assembly-stimulated processive cleavage of GTP to GMP. *Nature* **440**:101-  
712 104. doi:10.1038/nature04510.
- 713 Giepmans, BNG, Adams, SR, Ellisman, MH & Tsien, RY. 2006. Review - The fluorescent toolbox for  
714 assessing protein location and function. *Science* **312**:217-224. doi:10.1126/science.1124618.
- 715 Gorfou, G, Cirelli, KM, Melo, MB, Mayer-Barber, K, Crown, D, Koller, BH, Masters, S, Sher, A,  
716 Leppla, SH, Moayeri, M, Saeij, JP & Grigg, ME. 2014. Dual role for inflammasome sensors  
717 NLRP1 and NLRP3 in murine resistance to *Toxoplasma gondii*. *MBio*  
718 **5**:doi:10.1128/mBio.01117-13.
- 719 Haldar, AK, Piro, AS, Pilla, DM, Yamamoto, M & Coers, J. 2014. The E2-like conjugation enzyme  
720 Atg3 promotes binding of IRG and Gbp proteins to Chlamydia- and *Toxoplasma*-containing  
721 vacuoles and host resistance. *PLoS One* **9**:e86684. doi:10.1371/journal.pone.0086684.
- 722 Kalinin, S, Peulen, T, Sindbert, S, Rothwell, PJ, Berger, S, Restle, T, Goody, RS, Gohlke, H & Seidel,  
723 CA. 2012. A toolkit and benchmark study for FRET-restrained high-precision structural  
724 modeling. *Nat Methods* **9**:1218-25. doi:10.1038/nmeth.2222.
- 725 Kask, P, Palo, K, Fay, N, Brand, L, Mets, U, Ullmann, D, Jungmann, J, Pschorr, J & Gall, K. 2000.  
726 Two-dimensional fluorescence intensity distribution analysis: theory and applications. *Biophys*  
727 *J* **78**:1703-13. doi:10.1016/S0006-3495(00)76722-1.
- 728 Kravets, E, Degrandi, D, Weidtkamp-Peters, S, Ries, B, Konermann, C, Felekyan, S, Dargazanli, JM,  
729 Praefcke, GJ, Seidel, CA, Schmitt, L, Smits, SH & Pfeffer, K. 2012. The GTPase activity of

## Supplement C

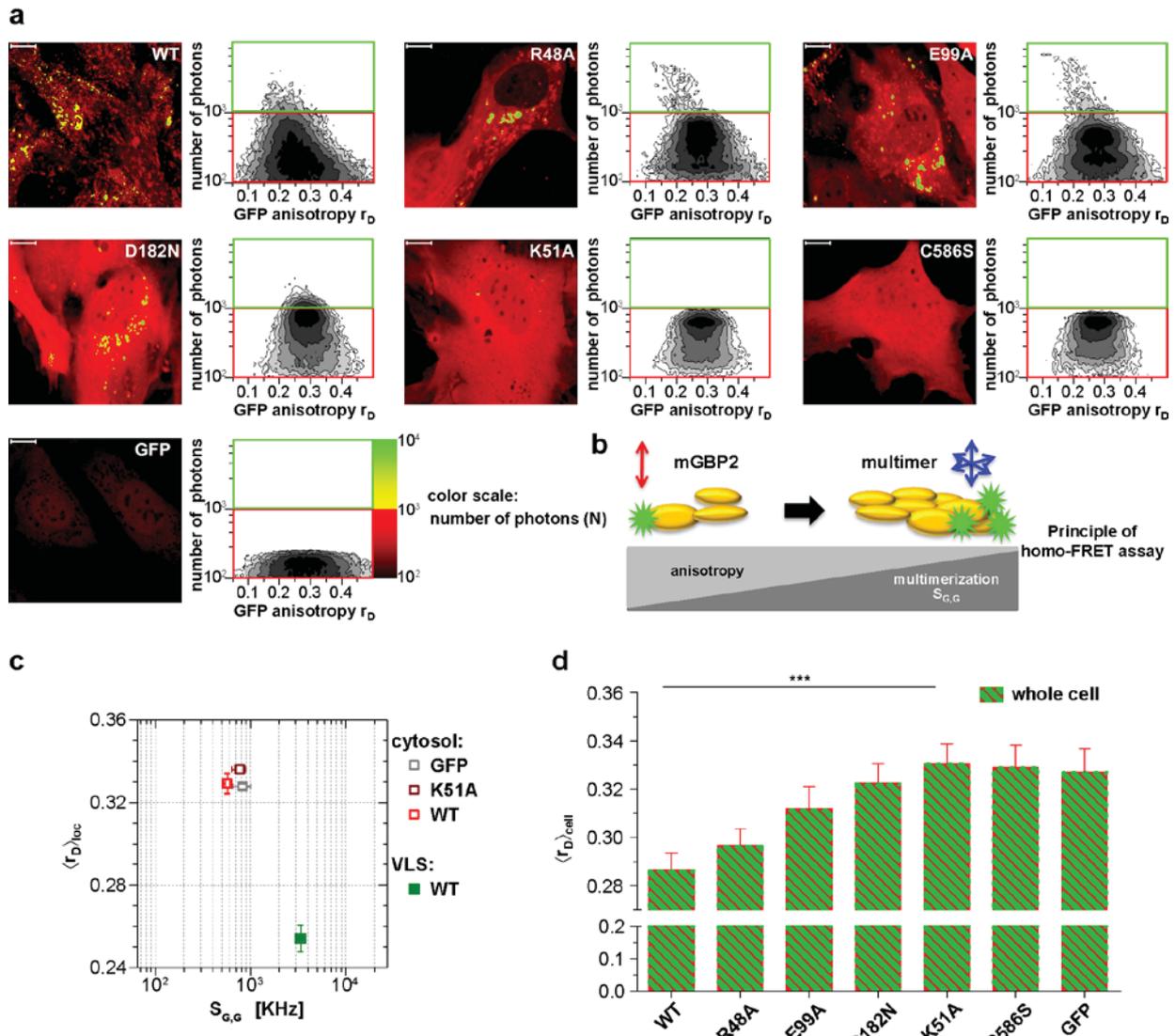
- 730 murine guanylate-binding protein 2 (mGBP2) controls the intracellular localization and  
731 recruitment to the parasitophorous vacuole of *Toxoplasma gondii*. *J Biol Chem* **287**:27452-66.  
732 doi:10.1074/jbc.M112.379636.
- 733 Kresse, A, Konermann, C, Degrandi, D, Beuter-Gunia, C, Wuerthner, J, Pfeffer, K & Beer, S. 2008.  
734 Analyses of murine GBP homology clusters based on in silico, in vitro and in vivo studies.  
735 *BMC Genomics* **9**:158. doi:10.1186/1471-2164-9-158.
- 736 Kudryavtsev, V, Felekyan, S, Wozniak, AK, Konig, M, Sandhagen, C, Kuhnemuth, R, Seidel, CA &  
737 Oesterhelt, F. 2007. Monitoring dynamic systems with multiparameter fluorescence imaging.  
738 *Anal Bioanal Chem* **387**:71-82. doi:10.1007/s00216-006-0917-0.
- 739 Lakowicz, JR 2006. *Principles of Fluorescence Spectroscopy*, New York, Springer.
- 740 MacMicking, JD. 2012. Interferon-inducible effector mechanisms in cell-autonomous immunity. *Nat*  
741 *Rev Immunol* **12**:367-82. doi:10.1038/nri3210.
- 742 Man, SM, Karki, R, Malireddi, RK, Neale, G, Vogel, P, Yamamoto, M, Lamkanfi, M & Kanneganti,  
743 TD. 2015. The transcription factor IRF1 and guanylate-binding proteins target activation of the  
744 AIM2 inflammasome by *Francisella* infection. *Nat Immunol* **16**:467-75. doi:10.1038/ni.3118.
- 745 Meunier, E, Dick, MS, Dreier, RF, Schurmann, N, Kenzelmann Broz, D, Warming, S, Roose-Girma, M,  
746 Bumann, D, Kayagaki, N, Takeda, K, Yamamoto, M & Broz, P. 2014. Caspase-11 activation  
747 requires lysis of pathogen-containing vacuoles by IFN-induced GTPases. *Nature* **509**:366-70.  
748 doi:10.1038/nature13157.
- 749 Meunier, E, Wallet, P, Dreier, RF, Costanzo, S, Anton, L, Ruhl, S, Dussurgey, S, Dick, MS, Kistner, A,  
750 Rigard, M, Degrandi, D, Pfeffer, K, Yamamoto, M, Henry, T & Broz, P. 2015. Guanylate-  
751 binding proteins promote activation of the AIM2 inflammasome during infection with  
752 *Francisella novicida*. *Nat Immunol* **16**:476-84. doi:10.1038/ni.3119.
- 753 Ohshima, J, Lee, Y, Sasai, M, Saitoh, T, Su Ma, J, Kamiyama, N, Matsuura, Y, Pann-Ghill, S, Hayashi,  
754 M, Ebisu, S, Takeda, K, Akira, S & Yamamoto, M. 2014. Role of mouse and human autophagy  
755 proteins in IFN-gamma-induced cell-autonomous responses against *Toxoplasma gondii*. *J*  
756 *Immunol* **192**:3328-35. doi:10.4049/jimmunol.1302822.
- 757 Olszewski, MA, Gray, J & Vestal, DJ. 2006. In silico genomic analysis of the human and murine  
758 guanylate-binding protein (GBP) gene clusters. *J Interferon Cytokine Res* **26**:328-52.  
759 doi:10.1089/jir.2006.26.328.
- 760 Pilla, DM, Hagar, JA, Haldar, AK, Mason, AK, Degrandi, D, Pfeffer, K, Ernst, RK, Yamamoto, M,  
761 Miao, EA & Coers, J. 2014. Guanylate binding proteins promote caspase-11-dependent  
762 pyroptosis in response to cytoplasmic LPS. *Proc Natl Acad Sci U S A* **111**:6046-51.  
763 doi:10.1073/pnas.1321700111.
- 764 Praefcke, GJ, Kloep, S, Benscheid, U, Lilie, H, Prakash, B & Herrmann, C. 2004. Identification of  
765 residues in the human guanylate-binding protein 1 critical for nucleotide binding and  
766 cooperative GTP hydrolysis. *J Mol Biol* **344**:257-69. doi:10.1016/j.jmb.2004.09.026.
- 767 Praefcke, GJ & McMahon, HT. 2004. The dynamin superfamily: universal membrane tubulation and  
768 fission molecules? *Nat Rev Mol Cell Biol* **5**:133-47. doi:10.1038/nrm1313.
- 769 Prakash, B, Praefcke, GJ, Renault, L, Wittinghofer, A & Herrmann, C. 2000a. Structure of human  
770 guanylate-binding protein 1 representing a unique class of GTP-binding proteins. *Nature*  
771 **403**:567-71. doi:10.1038/35000617.
- 772 Prakash, B, Renault, L, Praefcke, GJ, Herrmann, C & Wittinghofer, A. 2000b. Triphosphate structure of  
773 guanylate-binding protein 1 and implications for nucleotide binding and GTPase mechanism.  
774 *EMBO J* **19**:4555-64. doi:10.1093/emboj/19.17.4555.
- 775 Selleck, EM, Fentress, SJ, Beatty, WL, Degrandi, D, Pfeffer, K, Virgin, HWt, Macmicking, JD &  
776 Sibley, LD. 2013. Guanylate-binding protein 1 (Gbp1) contributes to cell-autonomous  
777 immunity against *Toxoplasma gondii*. *PLoS Pathog* **9**:e1003320.  
778 doi:10.1371/journal.ppat.1003320.
- 779 Sisamakias, E, Valeri, A, Kalinin, S, Rothwell, PJ & Seidel, CAM. 2010. Accurate Single-Molecule Fret  
780 Studies Using Multiparameter Fluorescence Detection. *Methods in Enzymology, Vol 475: Single*  
781 *Molecule Tools, Pt B* **475**:455-514. doi:10.1016/S0076-6879(10)75018-7.
- 782 Stahl, Y, Grabowski, S, Bleckmann, A, Kuhnemuth, R, Weidtkamp-Peters, S, Pinto, KG, Kirschner,  
783 GK, Schmid, JB, Wink, RH, Hulsewede, A, Felekyan, S, Seidel, CAM & Simon, R. 2013.  
784 Moderation of Arabidopsis Root Stertness by CLAVATA1 and ARABIDOPSIS CRINKLY4  
785 Receptor Kinase Complexes. *Current Biology* **23**:362-371. doi:10.1016/j.cub.2013.01.045.

## Supplement C

- 786 Suhling, K, Siegel, J, Phillips, D, French, PM, Leveque-Fort, S, Webb, SE & Davis, DM. 2002. Imaging  
787 the environment of green fluorescent protein. *Biophys J* **83**:3589-95. doi:10.1016/S0006-  
788 3495(02)75359-9.
- 789 Vestal, DJ, Gorbacheva, VY & Sen, GC. 2000. Different subcellular localizations for the related  
790 interferon-induced GTPases, MuGBP-1 and MuGBP-2: implications for different functions? *J*  
791 *Interferon Cytokine Res* **20**:991-1000. doi:10.1089/10799900050198435.
- 792 Vestal, DJ & Jeyaratnam, JA. 2011. The Guanylate-Binding Proteins: Emerging Insights into the  
793 Biochemical Properties and Functions of This Family of Large Interferon-Induced Guanosine  
794 Triphosphatase. *Journal of Interferon and Cytokine Research* **31**:89-97.  
795 doi:10.1089/jir.2010.0102.
- 796 Vöpel, T, Hengstenberg, CS, Peulen, TO, Ajaj, Y, Seidel, CA, Herrmann, C & Klare, JP. 2014.  
797 Triphosphate induced dimerization of human guanylate binding protein 1 involves association  
798 of the C-terminal helices: a joint double electron-electron resonance and FRET study.  
799 *Biochemistry* **53**:4590-600. doi:10.1021/bi500524u.
- 800 Weidtkamp-Peters, S, Felekyan, S, Bleckmann, A, Simon, R, Becker, W, Kuhnemuth, R & Seidel, CA.  
801 2009. Multiparameter fluorescence image spectroscopy to study molecular interactions.  
802 *Photochem Photobiol Sci* **8**:470-80. doi:10.1039/b903245m.
- 803 Widengren, J, Kudryavtsev, V, Antonik, M, Berger, S, Gerken, M & Seidel, CA. 2006. Single-molecule  
804 detection and identification of multiple species by multiparameter fluorescence detection. *Anal*  
805 *Chem* **78**:2039-50. doi:10.1021/ac0522759.
- 806 Yamamoto, M, Okuyama, M, Ma, JS, Kimura, T, Kamiyama, N, Saiga, H, Ohshima, J, Sasai, M,  
807 Kayama, H, Okamoto, T, Huang, DC, Soldati-Favre, D, Horie, K, Takeda, J & Takeda, K.  
808 2012. A cluster of interferon-gamma-inducible p65 GTPases plays a critical role in host defense  
809 against *Toxoplasma gondii*. *Immunity* **37**:302-13. doi:10.1016/j.immuni.2012.06.009.
- 810 Zhao, Z, Fux, B, Goodwin, M, Dunay, IR, Strong, D, Miller, BC, Cadwell, K, Delgado, MA, Ponpuak,  
811 M, Green, KG, Schmidt, RE, Mizushima, N, Deretic, V, Sibley, LD & Virgin, HW. 2008.  
812 Autophagosome-independent essential function for the autophagy protein Atg5 in cellular  
813 immunity to intracellular pathogens. *Cell Host Microbe* **4**:458-69.  
814 doi:10.1016/j.chom.2008.10.003.

815

816

817 **Figures**

818

819 **FIGURE 1. Intracellular homo-multimerization of WT and mutant mGBP2.**

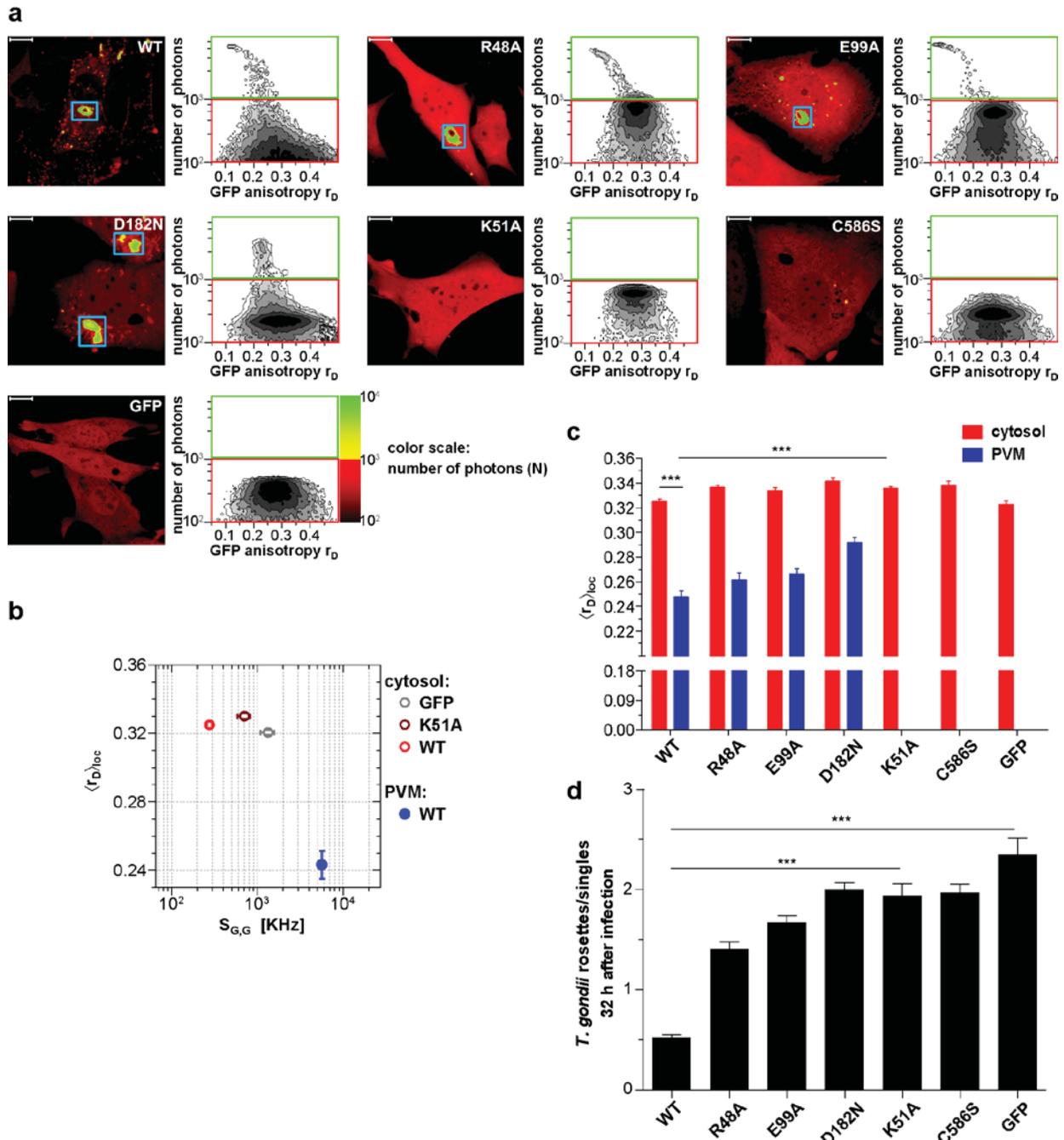
820 All cells were pre-treated with IFN $\gamma$  for 16h prior investigation (A) Left panel. GFP fluorescence  
 821 intensity ( $S_{G,G}$ ) images of GBP2 $^{-/-}$  MEFs expressing G-mGBP2-WT (G-mGBP2 MEFs) or mutants  
 822 (R48A, K51, E99A, D182N, C586S) highlighted with selections of pixels within different cellular  
 823 compartments. Right panel. MFIS 2D-histograms of GFP anisotropy ( $r_D$ ) on x axis vs. photon number  
 824 per pixel on y axis, the frequency of pixels color coded from white (lowest) to black (highest). This  
 825 allows the identification and selection of pixel populations with unique fluorescence properties for a  
 826 detailed subsequent pixel integrated analysis. The pixels with low photon numbers (below 1000) are  
 827 selected in red boxes (defined as cytosol) and those with more than 1000 photons in green boxes

## Supplement C

828 (defined as VLS). Bars, 10  $\mu\text{m}$ . (B) Scheme of the principle of homo-FRET assays. Compared to G-  
829 mGBP2 monomers,  $r_D$  in G-mGBP2 multimers decreases due to depolarization of GFP fluorescence  
830 while GFP  $S_{G,G}$  increases. (C) Mean values of  $\langle r_D \rangle_{\text{loc}}$  and  $S_{G,G}$  in cytosol and VLS were plotted for G-  
831 mGBP2-WT, and the K51A mutant and GFP in the cytosol. (D) Mean anisotropy of averages over  
832 whole cells  $\langle r_D \rangle_{\text{cell}}$  for G-mGBP2 WT and mutant proteins. GFP expressing cells served as controls  
833 (\*\*\*)  $P < 0.0001$ .

834

## Supplement C

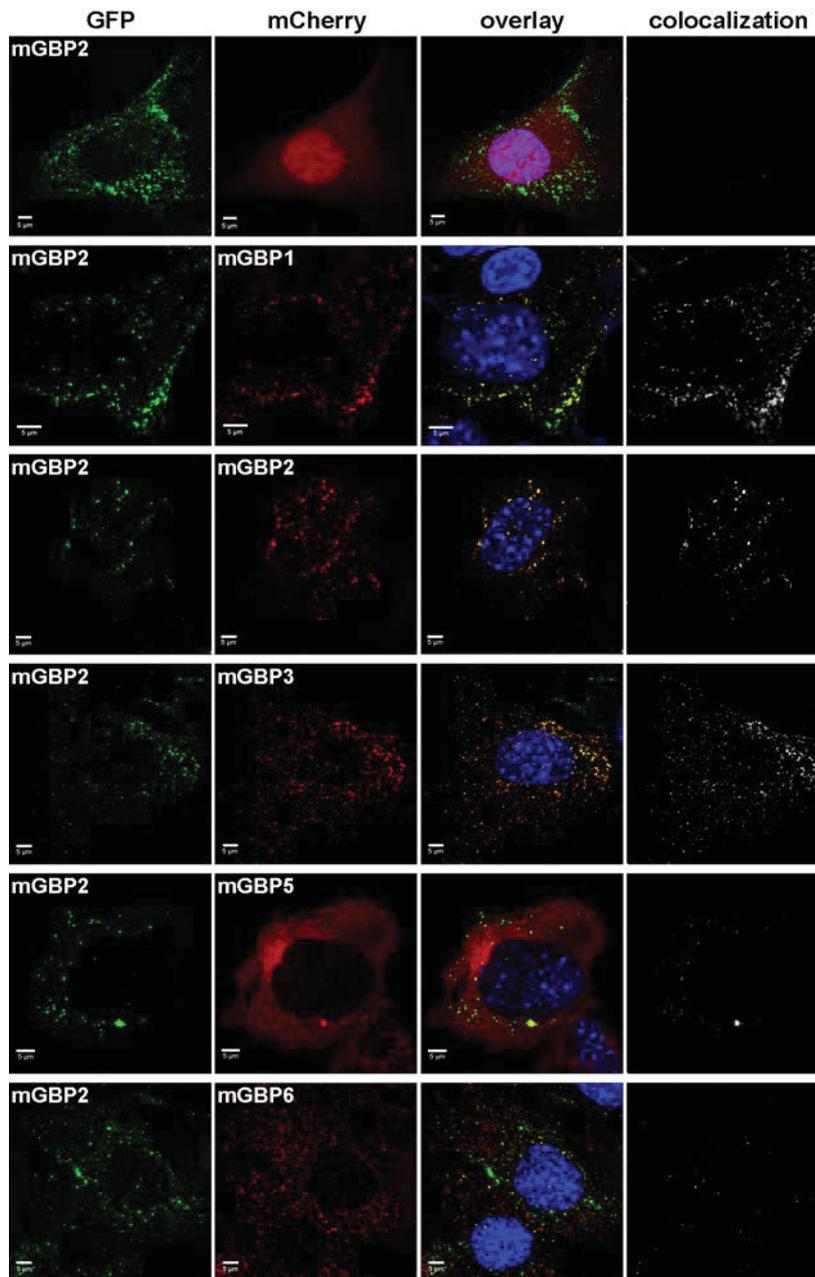


835  
 836 **FIGURE 2. Intracellular homo-multimerization of WT and mutant mGBP2 at the PVM of *T.***  
 837 ***gondii* and parasite inhibition.**  
 838 Cells were pre-treated with IFN $\gamma$  for 16h prior infection with *T. gondii* ME49 (A) Left panel. GFP  
 839 fluorescence intensity images of G-mGBP2-WT or mutants MEFs highlighted with selections of pixels  
 840 with low and high numbers of photons. Blue boxes mark the PVM area. Bars, 10  $\mu$ m. Right panel.  
 841 MFIS 2D-histograms of GFP  $r_D$  on x axis vs. photon number per pixel on y axis. The pixels with low  
 842 photon numbers (below 1000) are selected in red boxes and the pixels containing more than 1000  
 843 photons in green boxes. (B) Mean values of  $\langle r_D \rangle_{loc}$  and mean GFP  $S_{G,G}$  were plotted for G-mGBP2-WT

## Supplement C

844 in the cytosol and at the PVM of *T. gondii* and for the K51A mutant and GFP in the cytosol. (C) Mean  
845 anisotropy  $\langle r_D \rangle_{loc}$  of WT and mutants in the cytosol and at the PVM (blue boxes in A). GFP expressing  
846 cells served as controls (\*\* $P < 0.0001$ ). (D) Replication inhibitory capacity of G-mGBP2-WT and  
847 mutants. After fixation *T. gondii* were stained with the  $\alpha$ -SAG1 antibody and the cell nuclei with DAPI.  
848 Slides were analyzed by confocal microscopy. Replication inhibition was calculated by the ratio of *T.*  
849 *gondii* single parasites versus replicative units (rosettes) in at least 100 infected cells (\*\* $P < 0.0001$ ).  
850

## Supplement C



851

852 **FIGURE 3. Intracellular colocalization of mGBP proteins.**

853 Subcellular localization of mGBPs was analyzed in G-mGBP2 coexpressing one of the mCh-mGBPs (1,

854 2, 3, 5 or 6). mCherry expressing cells served as controls. Cells were pre-treated with IFN $\gamma$  for 16 h.

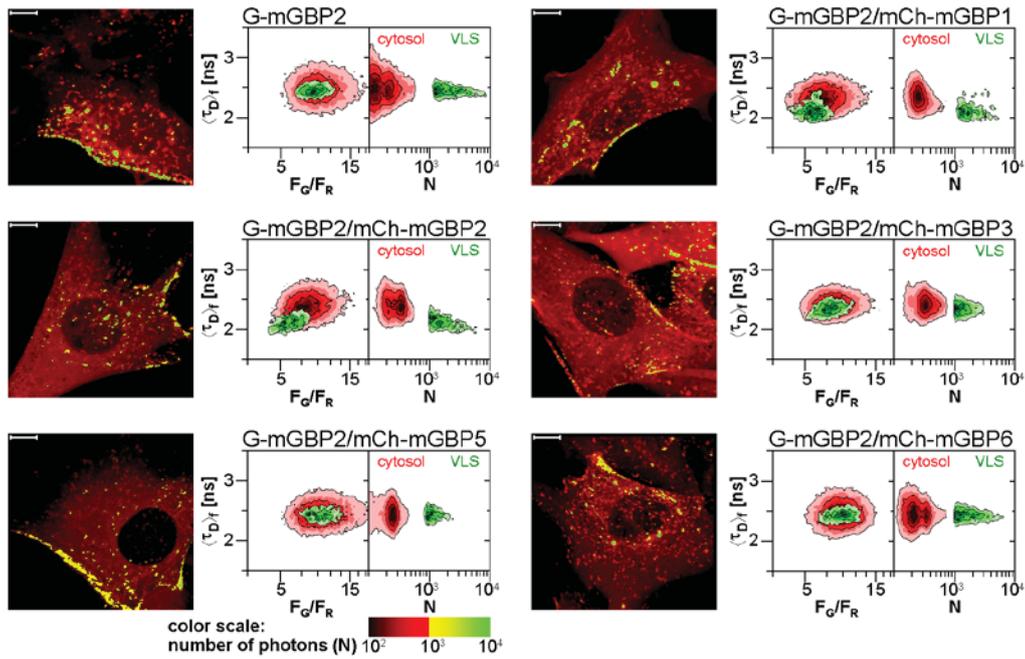
855 After fixation, nuclei were stained with DAPI. Glass slides were analyzed by confocal microscopy.

856 Bars, 5  $\mu$ m. Colocalization analysis was performed with Imaris (Bitplane).

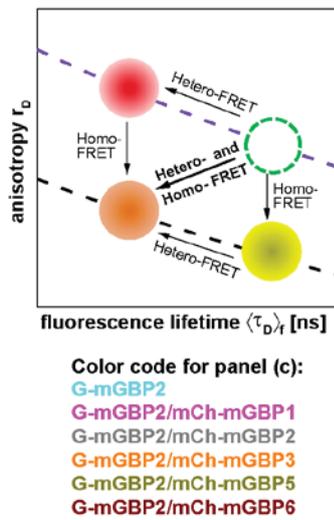
857

## Supplement C

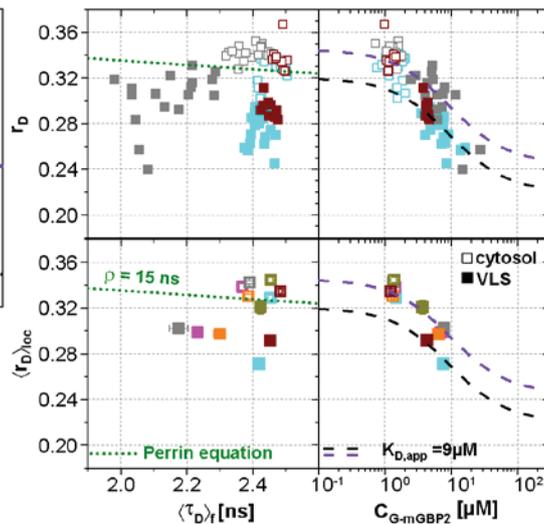
a



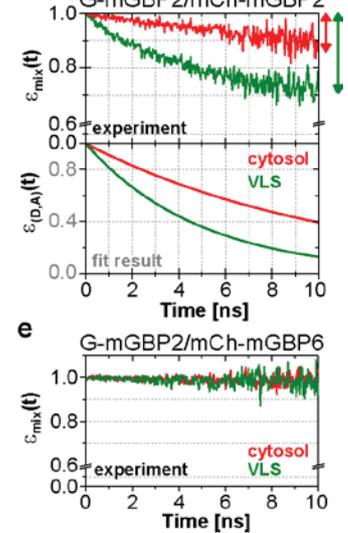
b



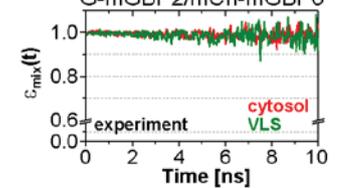
c



d



e



858  
859

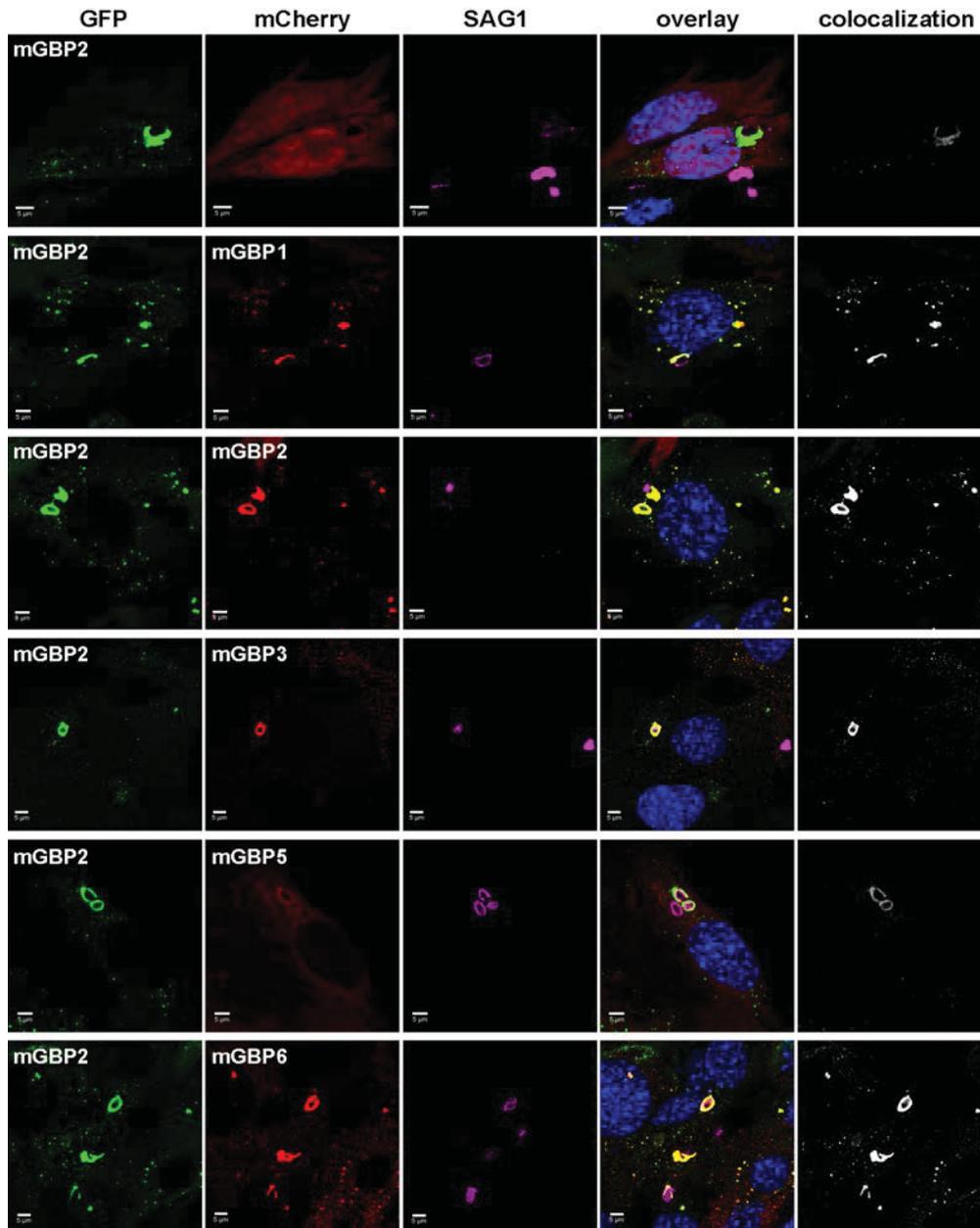
**FIGURE 4. Intracellular homo- and hetero-multimerization of mGBPs.**

860 All cells were pre-treated with IFN $\gamma$  for 16h prior investigation (A) Left panels. GFP fluorescence  
 861 intensity images of G-mGBP2 or G-mGBP2/mCh-mGBP(1,2,3,5,6) MEFs highlighted with selections  
 862 of pixels with different intensities. Bars, 10  $\mu$ m. Right panels. Two MFIS 2D-histograms of GFP  
 863 fluorescence lifetimes ( $\langle \tau_{D,f} \rangle$ ) on y axes, GFP/mCherry fluorescence intensity ratios ( $F_G/F_R$ ) or photon  
 864 number per pixel ( $N$ ) on x axes. The pixel populations locating in cytosol ( $N < 1000$ : red island) and  
 865 VLS ( $N > 1000$ : green island) were separated according to photon numbers. (B) Schematic 2D MFIS  
 866 plot detailing the effects of hetero- and/or homo-FRET on a reference data set (green circle). The

## Supplement C

867 average GFP  $\langle \tau_D \rangle_f$  is plotted on the x axis from short to long, while the average steady-state  $r_D$  is plotted  
868 on the y axis. For detailed explanation refer to results section. (C) Upper panel. For individual G-  
869 mGBP2, G-mGBP2/mCh-mGBP2 or G-mGBP2/mCh-mGBP6 MEFs, mean values of  $r_D$  in the cytosol  
870 (empty squares) and in the VLS (solid squares) were plotted against  $\langle \tau_D \rangle_f$  and G-mGBP2 concentrations  
871 ( $C_{G\text{-mGBP2}}$ ). Lower panel. Mean anisotropy  $\langle r_D \rangle_{loc}$  values (average over all cells weighted by  $C_{G\text{-mGBP2}}$ )  
872 were plotted against  $\langle \tau_D \rangle_f$  or  $C_{G\text{-mGBP2}}$ . The two left panels contain an overlay calculated according to  
873 the Perrin equation:  $r_D = r_0 / (1 + \langle \tau_D \rangle_f / \rho_{global})$  with GFP fundamental anisotropy  $r_0 = 0.38$  and  
874 rotational correlation time  $\rho_{global} = 15$  ns. The two right panels are overlaid with function curves plotting  
875  $r_D = r_{max} - (r_{max} - r_{min}) \cdot C_{G\text{-mGBP2}} / (C_{G\text{-mGBP2}} + K_{D,app})$  which assumes a mGBP2 Langmuir binding  
876 model with an apparent dissociation constant  $K_{D,app}$ . In all the donor-only experiments the formation of  
877 mGBP2 homo-multimers could be described by  $K_{D,app} = 9$   $\mu\text{M}$ ,  $r_{max} = 0.32$  and  $r_{min} = 0.22$  (black  
878 curve). If other interaction processes interfere with homo-FRET between G-mGBP2 proteins, this curve  
879 is shifted upwards (violet curve) while keeping  $K_{D,app}$  invariant ( $r_{max} = 0.345$  and  $r_{min} = 0.245$ ). (D, E)  
880  $\varepsilon_{mix}(t)$  and  $\varepsilon_{(D,A)}(t)$  diagrams of a representative G-mGBP2/mCh-mGBP2 MEF (D) and G-  
881 mGBP2/mCh-mGBP6 MEF (E). The drop in  $\varepsilon_{mix}(t)$  curves, as marked by the arrows, represents the  
882 species fractions of FRET-active complexes ( $x_{FRET}$ ) in the VLS (green) and in the cytosol (red). In (D),  
883 the FRET rate constant ( $k_{FRET}$ ) in the cytosol is  $0.09$  ns<sup>-1</sup> and in the VLS  $0.20$  ns<sup>-1</sup>.  
884

## Supplement C

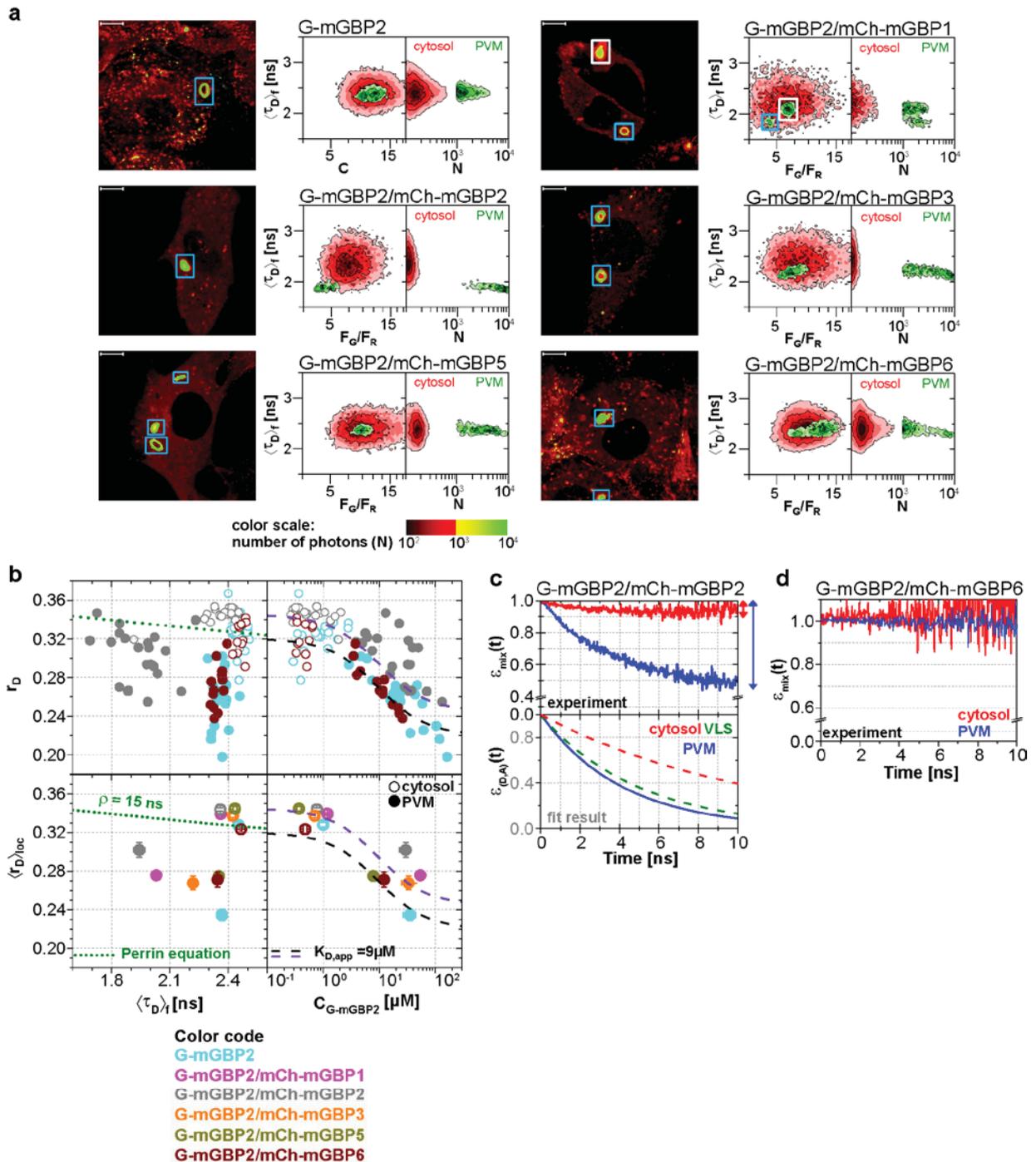


885

886 **FIGURE 5. Intracellular colocalization at the PVM of *T. gondii* and enrichment of mGBP**  
887 **proteins.** Recruitment and colocalization of mGBPs was analyzed in G-mGBP2/mCh-mGBP(1,2,3,5,6)  
888 MEFs. mCherry expressing cells served as controls. Cells were stimulated with IFN $\gamma$  for 16 h and  
889 subsequently infected with *T. gondii* for 2 h. After fixation, *T. gondii* were stained with an  $\alpha$ -SAG1  
890 antibody and cell and *T. gondii* nuclei with DAPI. Glass slides were analyzed by confocal microscopy.  
891 Bars, 5  $\mu$ m. Colocalization analysis was performed with Imaris (Bitplane).

892

## Supplement C

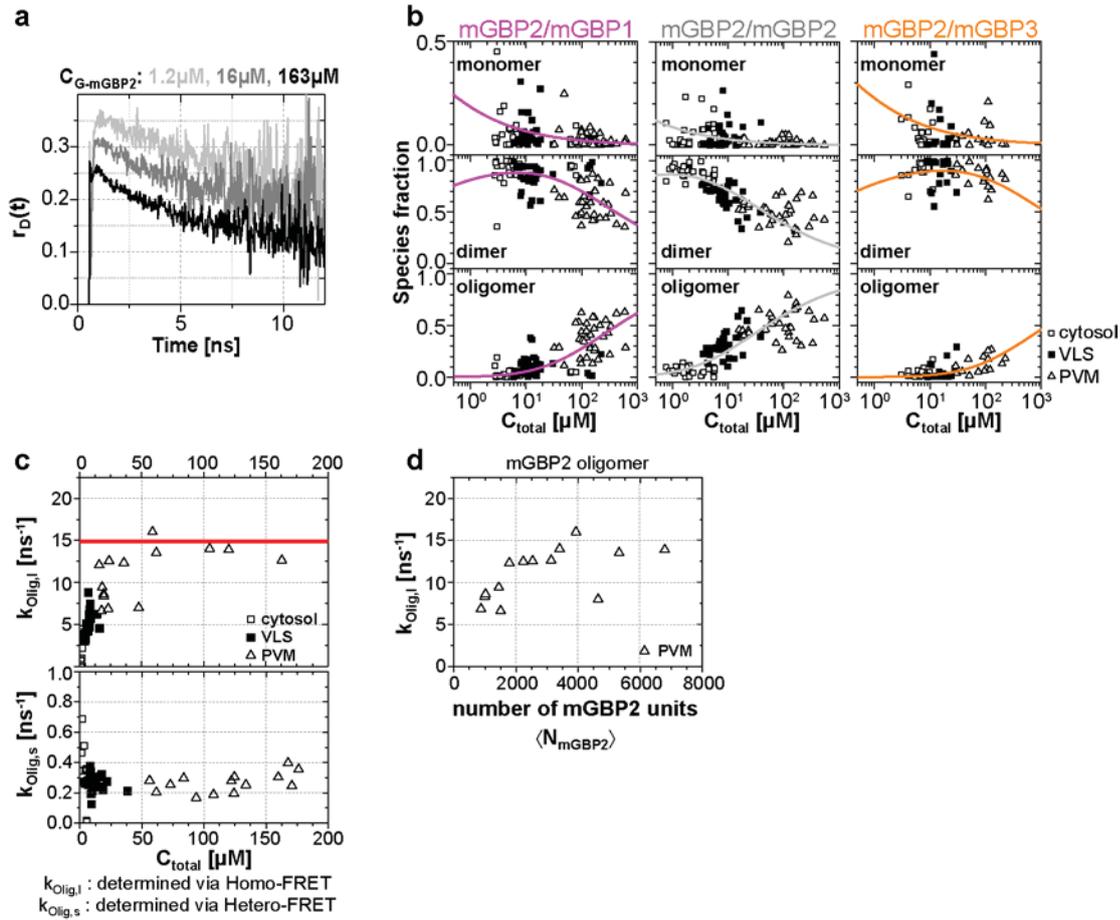


893 **FIGURE 6. Intracellular homo- and hetero-multimerization of mGBPs at the PVM of *T. gondii*.**  
 894  
 895 All cells were pre-treated with IFN $\gamma$  for 16h prior investigation (A) Left panels. GFP fluorescence  
 896 intensity images of living G-mGBP2 or G-mGBP2/mCh-mGBP(1,2,3,5,6) MEFs infected with *T.*  
 897 *gondii* highlighted with selections of pixels within different intracellular localizations. Right panels.  
 898 Two MFIS 2D-histograms of GFP  $\langle \tau_D \rangle_f$  on y axes, GFP/mCherry  $F_G/F_R$  and photon number per pixel  
 899 ( $N$ ) on x axes. The pixel populations locating in cytosol ( $N < 1000$ : red island) and at the PVM ( $N >$   
 900  $1000$ : green island) were separated according to photon numbers. (B) Upper panel. For individual G-

## Supplement C

901 mGBP2, G-mGBP2/mCh-mGBP2 or G-mGBP2/mCh-mGBP6 MEFs pixel averages of  $r_D$  in the cytosol  
902 and at the PVM were plotted against  $\langle \tau_D \rangle_f$  or  $C_{G\text{-mGBP2}}$ . Lower panel. Averages of  $\langle r_D \rangle_{\text{loc}}$  were plotted  
903 against  $\langle \tau_D \rangle_f$  and  $C_{G\text{-mGBP2}}$ . Please refer to Fig 4C for further information on the legend and overlaid  
904 curves. (C, D)  $\varepsilon_{\text{mix}}(t)$  and  $\varepsilon_{(D,A)}(t)$  diagrams of a representative *T. gondii* infected G-mGBP2/mCh-  
905 mGBP2 MEF (C) and G-mGBP2/mCh-mGBP6 MEF (D). The drop in  $\varepsilon_{\text{mix}}(t)$  curves, as marked by the  
906 arrows, represents  $x_{FRET}$  at the PVM (blue) and in the cytosol (red). The dashed curves representing the  
907  $\varepsilon_{(D,A)}(t)$  diagrams of G-mGBP2/mCh-mGBP2 interactions in the cytosol (red) and VLS (green) in  
908 uninfected cells are inserted for comparison from Fig. 4D. In (C),  $k_{FRET}$  at the PVM is  $0.24 \text{ ns}^{-1}$ .  
909

## Supplement C

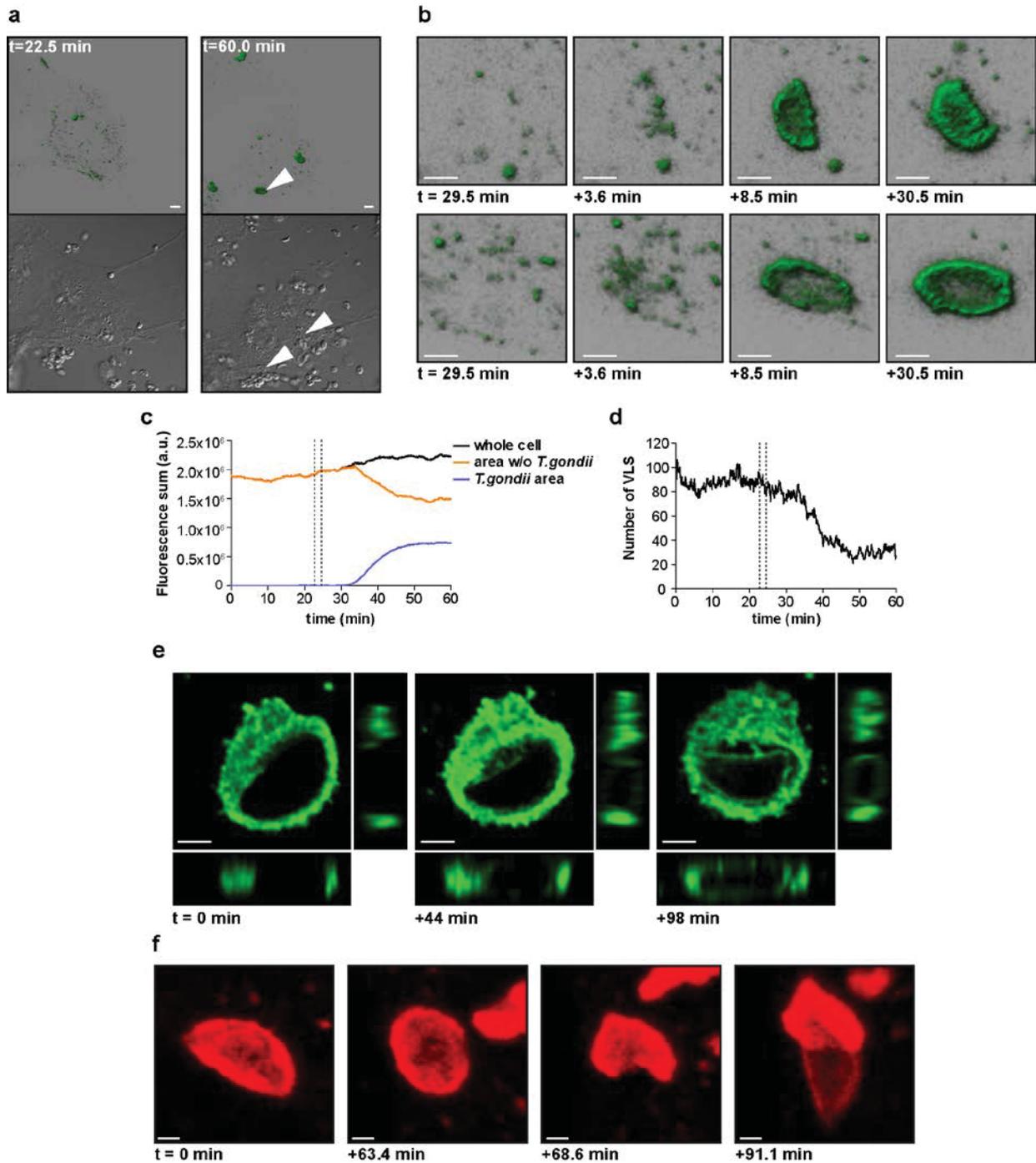


910

911 **FIGURE 7. Species-resolved analysis of mGBP2 homo- and hetero-complexes** (A) G-mGBP2  
 912 MEFs with higher concentration exhibited larger quasi instantaneous drop of  $r_D(t)$  from its initial value  
 913 of  $\sim 0.35$ , which proves the appearance of a very fast depolarization process due to homo-FRET in  
 914 mGBP2 oligomers. (B) Concentration dependence of the three mGBP species (monomer, dimer and  
 915 oligomer) obtained by the global pattern fit (eqs. 6 and 7) of  $r_{\text{mix}}(t)$  and  $\varepsilon_{\text{mix}}(t)$  for two localizations VLS  
 916 and PVM. The line depicts the fit (methods sections 10 and 13) to the corresponding binding  
 917 equilibrium with  $K_{D,\text{dim}}$ , and  $K_{D,\text{app-oligo}}$  (values are given in the main text). (C) Concentration  
 918 dependence of FRET rate constants for mGBP2 oligomers which formally differentiated as small  
 919 ( $k_{\text{Olig},s}$ ) and large ( $k_{\text{Olig},l}$ ). (D)  $k_{\text{Olig},l}$  versus the number of monomer units in mGBP2 multimers at the  
 920 PVM determined by scanning FIDA (see methods section 15).

921

## Supplement C

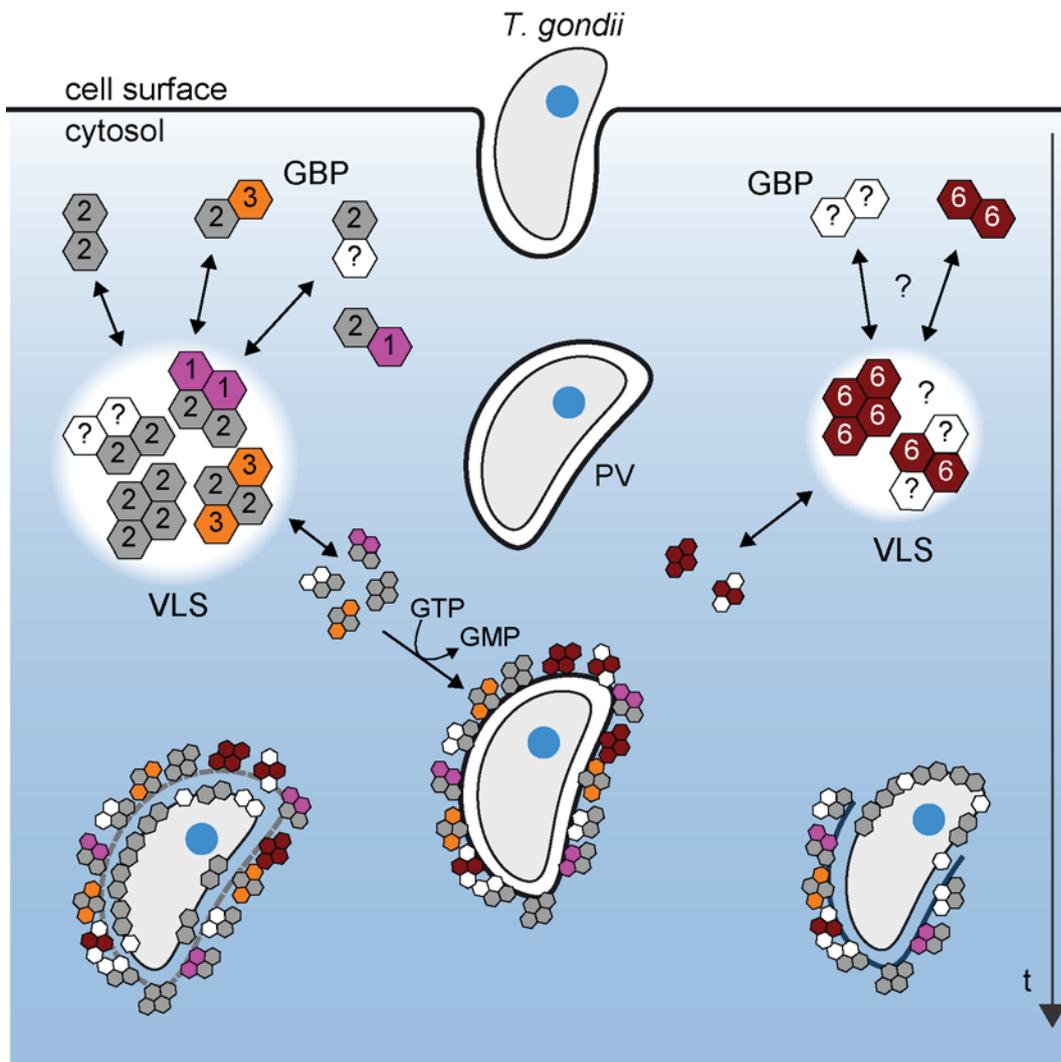


922  
 923 **FIGURE 8. Live-cell imaging of mGBP2 in *T.gondii* infection.** (A) G-mGBP2 MEFs were treated o/n  
 924 with IFN $\gamma$  and infected with *T.gondii* ME49. Living cells were observed by confocal microscopy at  
 925 37°C and a z-stack was recorded every 5-10 s. 4D data were processed and rendered in normal shading  
 926 mode (upper panels) and the DIC images are displayed (lower panels) for the indicated time points. One  
 927 out of at least 3 similar experiments is shown. Bar = 5 $\mu$ m. (B) Magnification from Video S1 and Figure  
 928 7A of G-mGBP2 accumulation around two *T.gondii* parasites at time points indicated. Bar = 2  $\mu$ m. (C)  
 929 Quantification of the total fluorescence intensity over the indicated voxels from Video S1. Vertical lines

## Supplement C

930 indicate the time points of *T. gondii* infection of MEFs. One representative analysis out of at least 3  
931 similar experiments is shown. (D) Number of cytosolic VLS with at least approx. 0.25  $\mu\text{m}$  diameter  
932 from Video S1 over time. Fluorescence signals close to the *T. gondii* area were excluded from the  
933 analysis. Vertical lines indicate the time points of *T. gondii* infection of MEFs. One representative  
934 analysis out of at least 3 similar experiments is shown (E) XY, XZ, and YZ projections of G-mGBP2  
935 around one *T. gondii* PVM are shown for the indicated time points. Bar = 2  $\mu\text{m}$ . (F) Maximum intensity  
936 projections of mCh-mGBP2 around one *T. gondii* are shown for the indicated time points. Bar = 1  $\mu\text{m}$ .

937

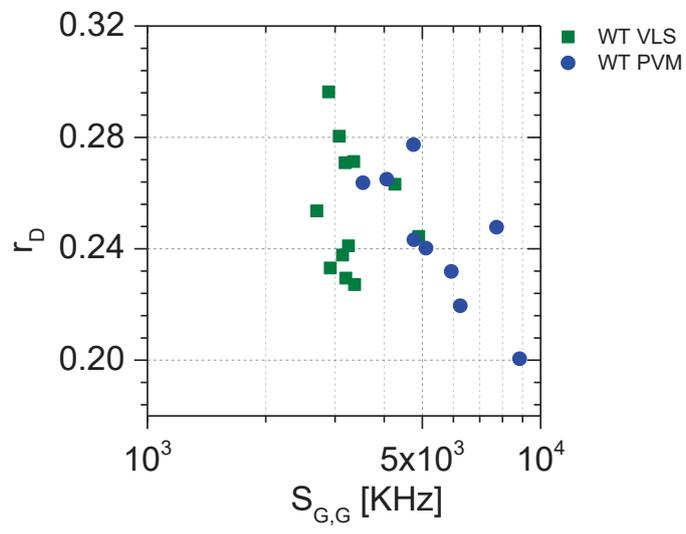


938

939 **FIGURE 9. Schematic view of mGBP dynamics and multimerization in *T.gondii* infected cells.** For  
 940 details see Results and Discussion

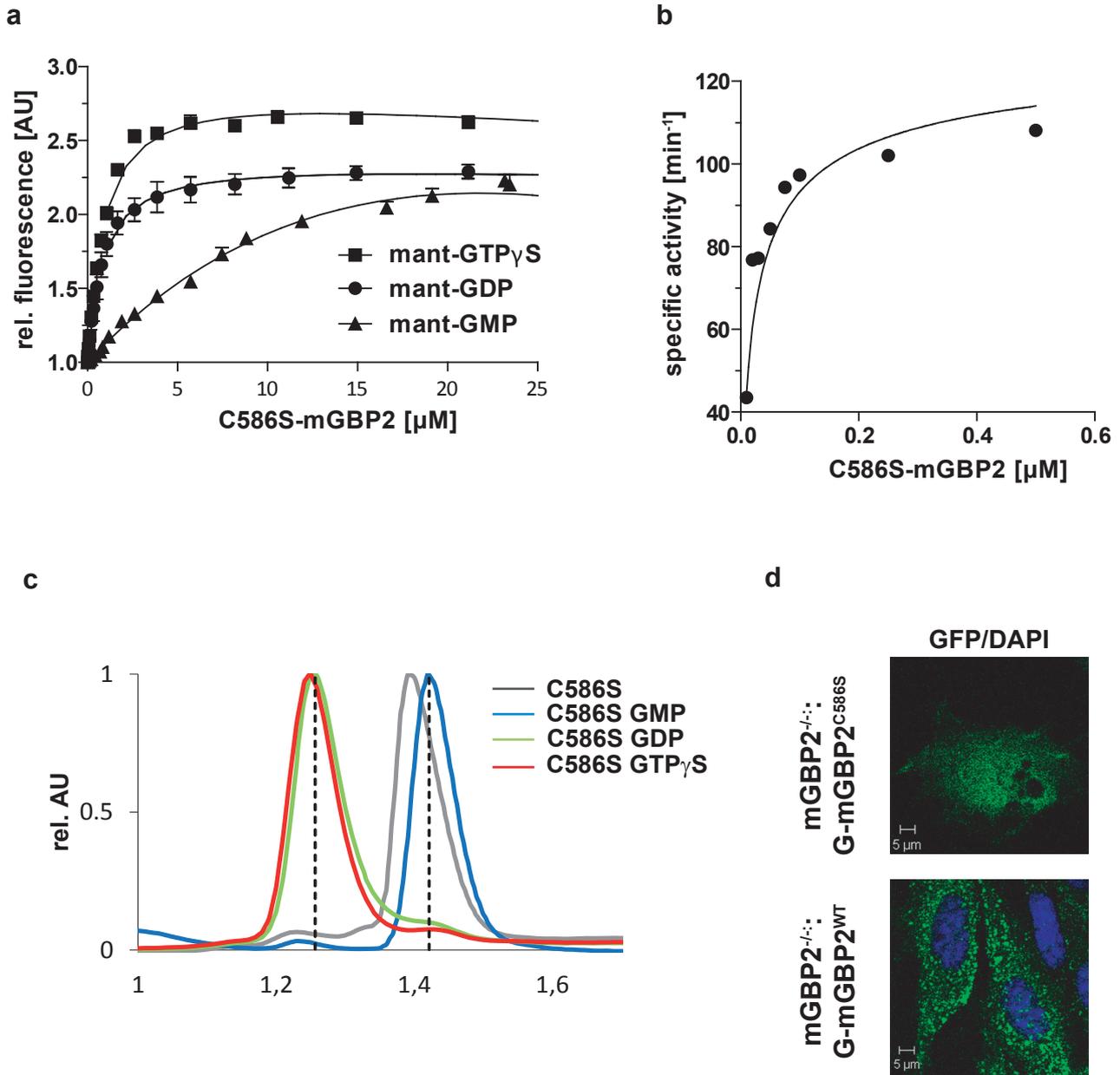
### Supplement C

Supplementary Figure S1



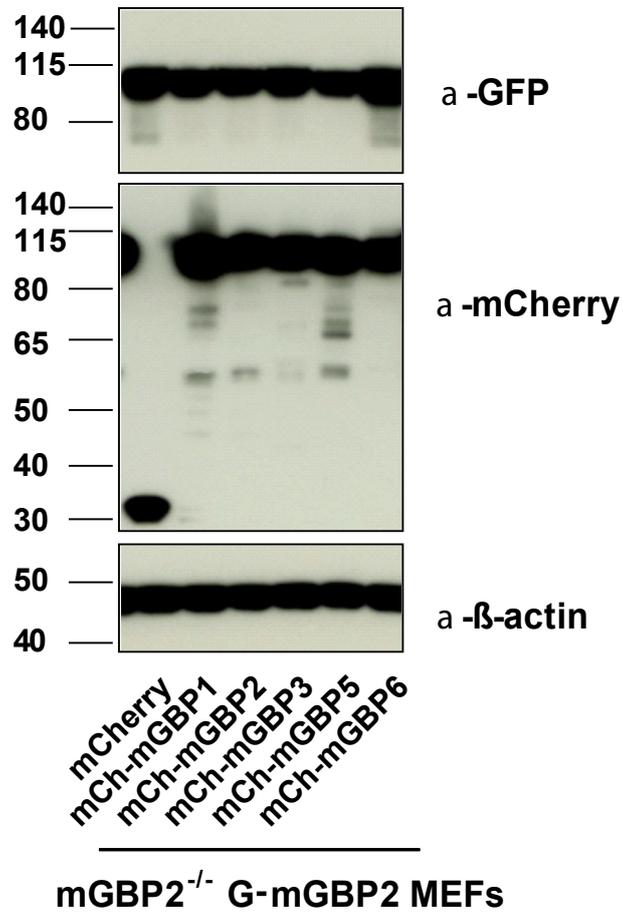
Supplement C

Supplementary Figure S2

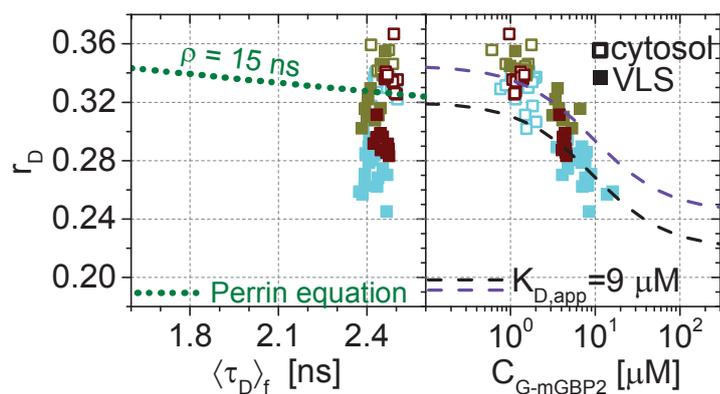


Supplement C

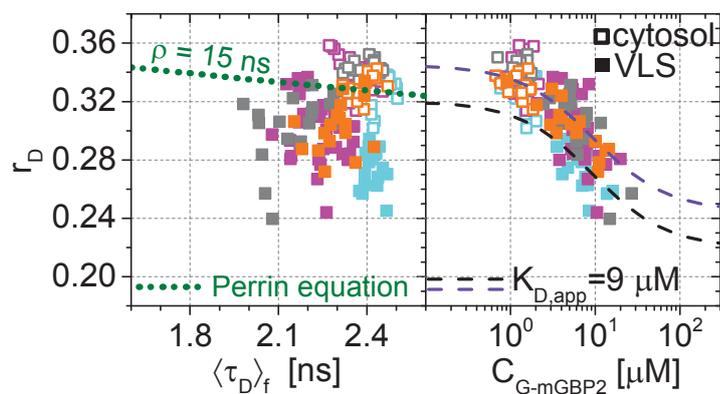
Supplementary Figure S3



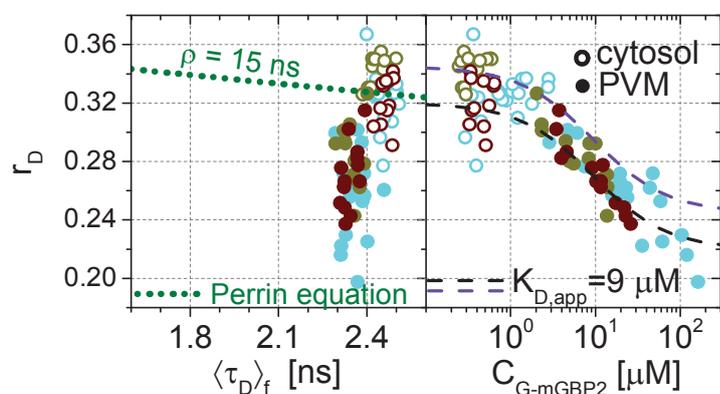
a



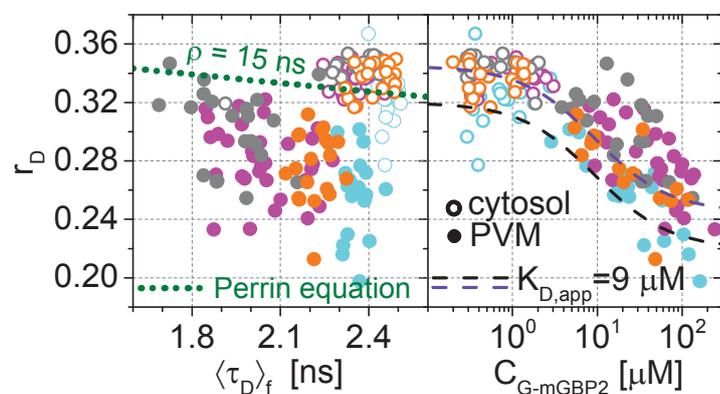
b



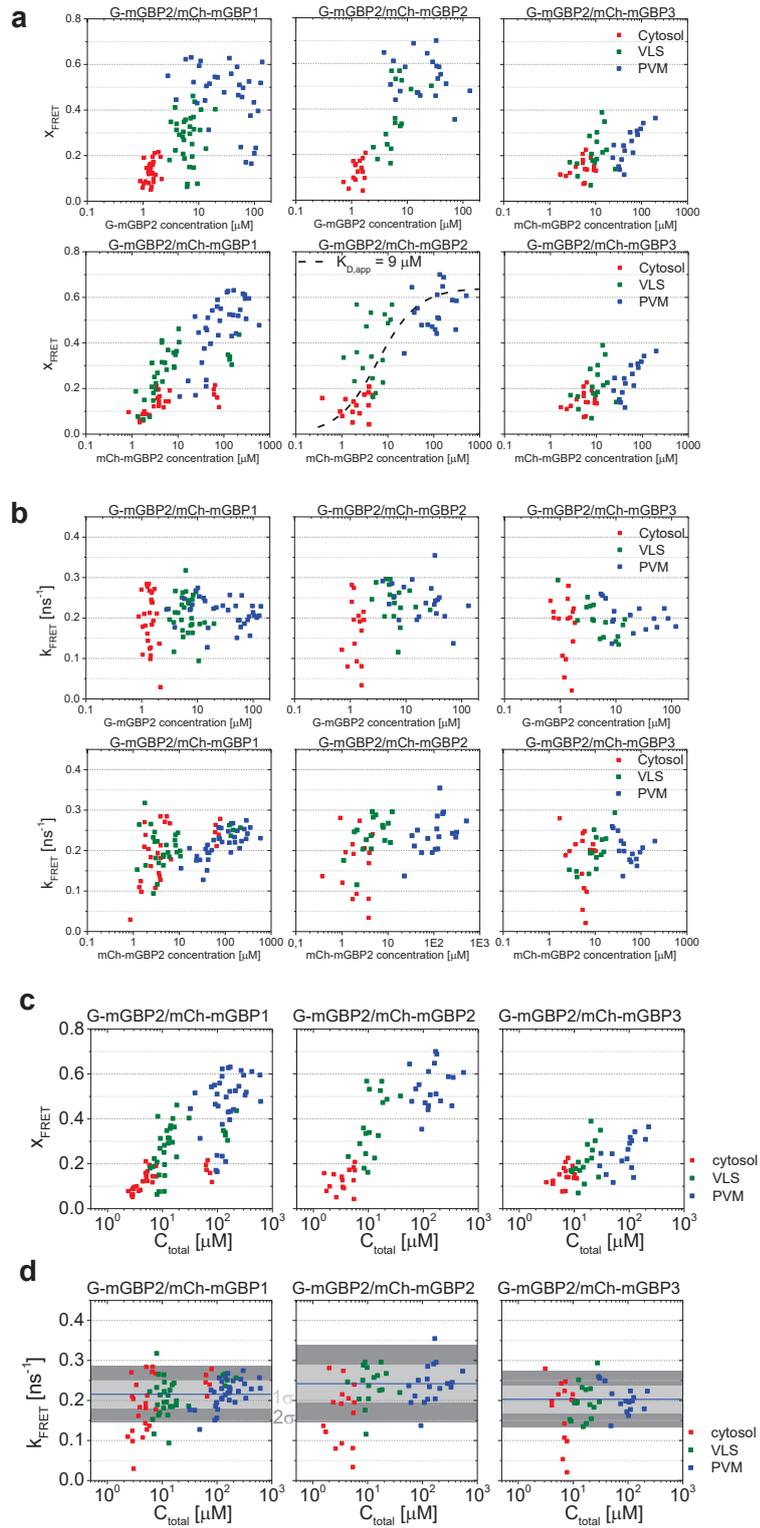
c



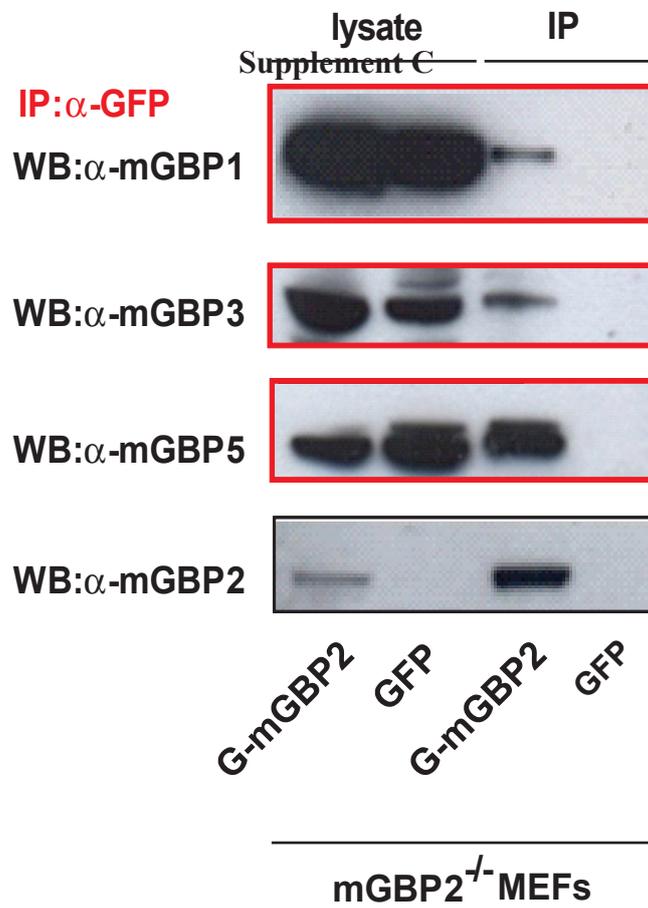
d



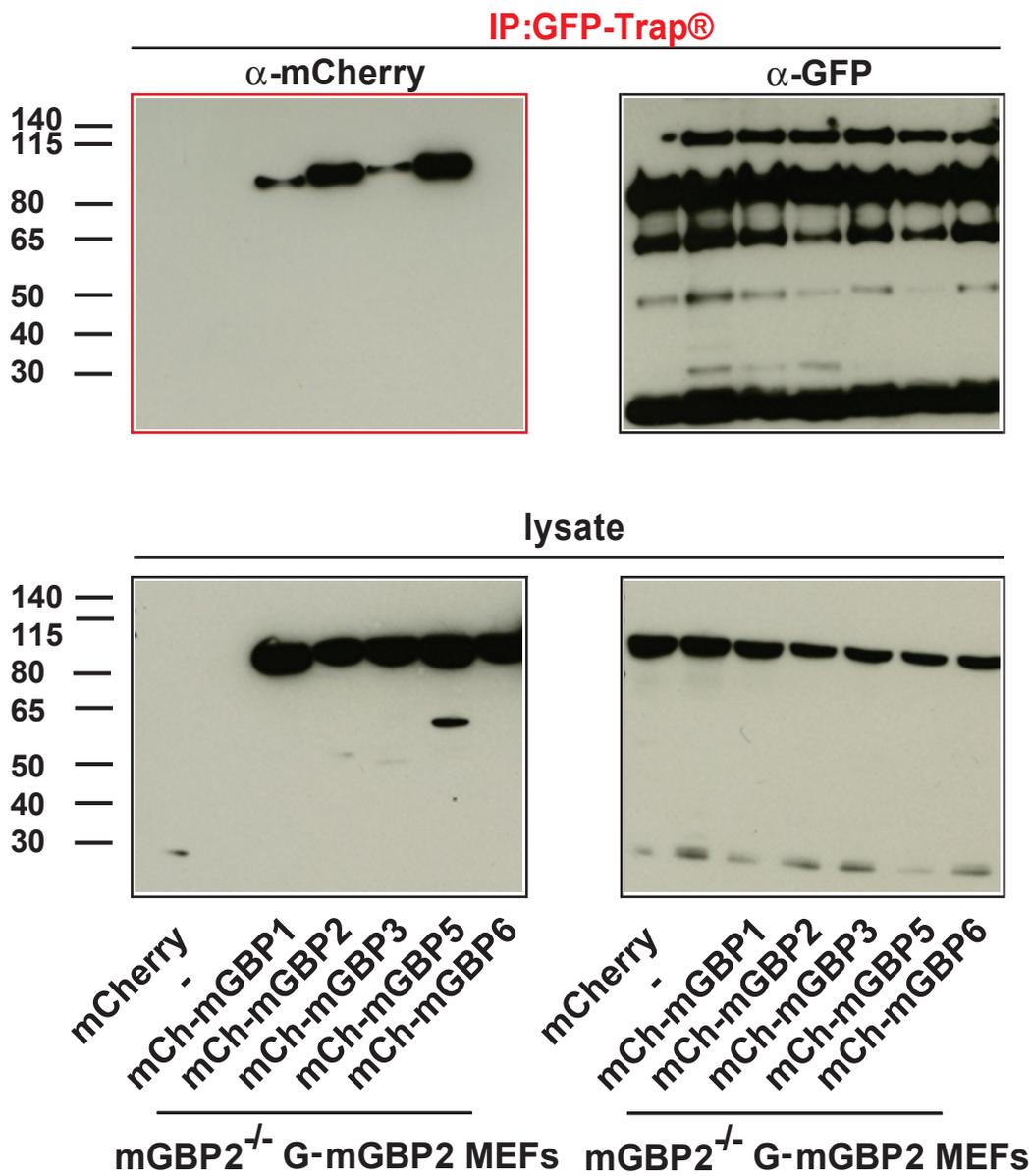
Supplement C



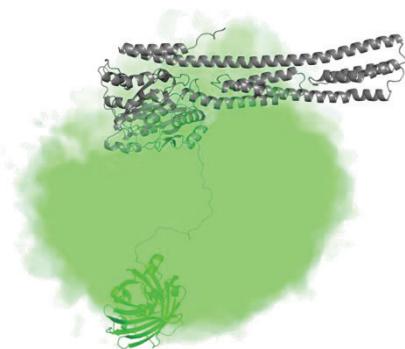
a



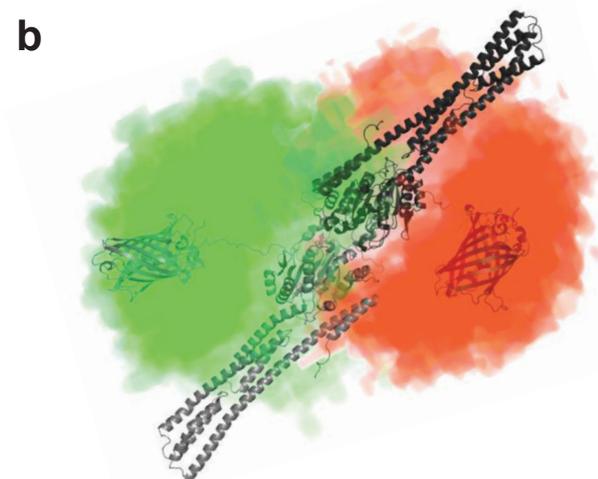
b



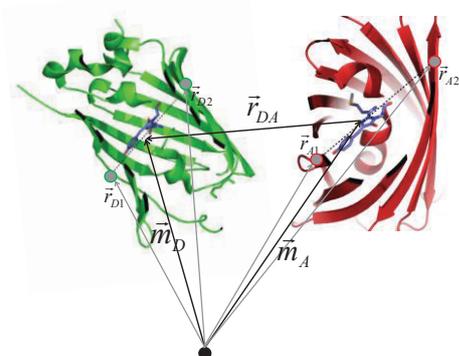
a



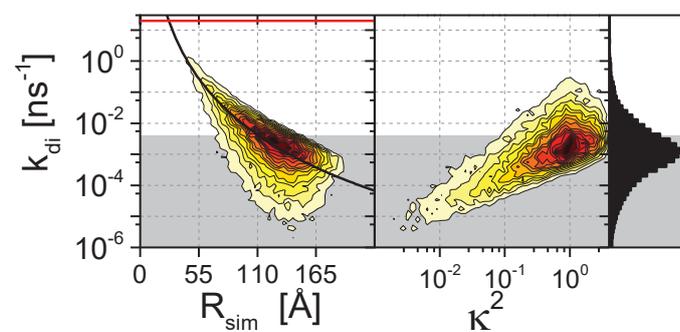
b



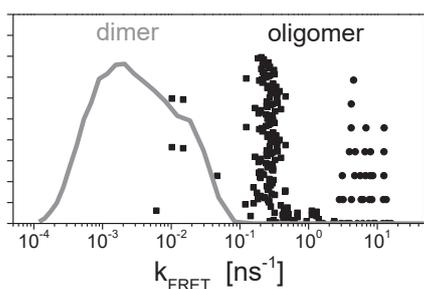
c



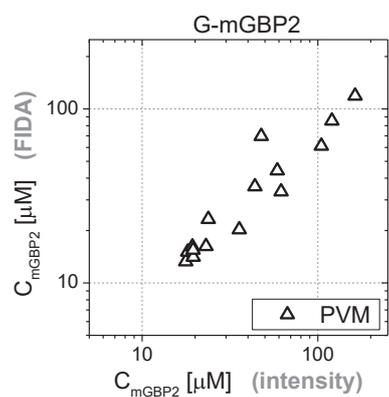
d



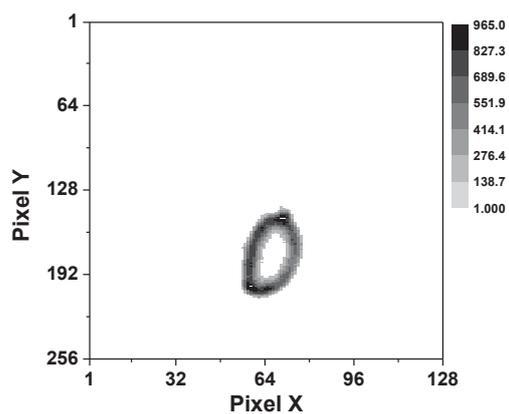
e



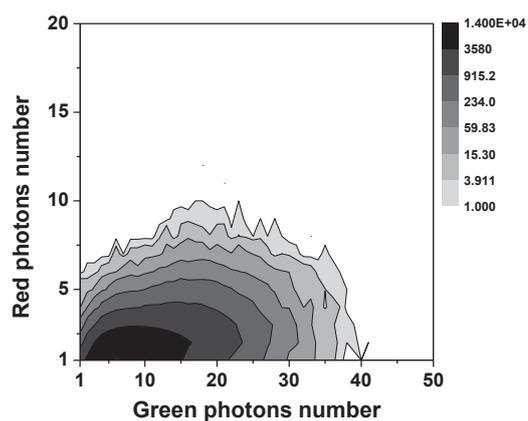
f



g



h



## Supplement C

**FIGURE S1. Spectroscopic characterization of G-mGBP2 WT in VLS in non-infected cells and at the PVM in *T. gondii* infected cells via homo-FRET assay.** Average values of GFP fluorescence anisotropy ( $r_D$ ) and signal intensity ( $S_{G,G}$ ) over single-cell measurements are plotted, in which  $S_{G,G}$  values are proportional to protein concentration. A much wider distribution of  $S_{G,G}$  can be observed for G-mGBP2 localizing at the PVM (blue circles) comparing to the  $S_{G,G}$  values for G-mGBP2 localizing in the VLS (green squares).

**FIGURE S2. Biochemical properties and intracellular localization of the C586S mutant of mGBP2.** (A) Nucleotide binding. A solution containing 0.5  $\mu$ M mant-GTP $\gamma$ S, mant-GDP and mant-GMP was titrated with C586S mutant of mGBP2. The fluorescence was excited at 355 nm and detected at 448 nm. The values were normalized to the fluorescence of the nucleotide alone. Dissociation constants are calculated from the fit of the binding curves as in (Kravets et al., 2012). The results averaged over two to four experiments each are given in the Table S2. (B) GTP-hydrolysis. Concentration-dependent GTP-hydrolysis catalyzed by the C586S mutant was measured with a fixed concentration of GTP (1 mM) at 37°C. The initial rates were measured (< 30% GTP hydrolyzed) from the linear parts of time-course experiments and normalized to the protein concentrations used (specific activity). The specific activities were then plotted against protein concentrations. The data were fitted to a model describing the interaction of two molecules of mGBP2, yielding  $K_D$  ( $\mu$ M) and the maximal specific activity  $K_{max}$  ( $\text{min}^{-1}$ ). The maximum specific GTPase activity, the dimer dissociation constant and the amount of GMP production are summarized in the Table S2. (C) Nucleotide-dependent multimerization. Size-exclusion chromatography of the C586S mutant of mGBP2 bound to GTP $\gamma$ S, GDP, GMP and in the nucleotide free state at 4°C. Elution of all proteins was followed using absorbance by 280 nm. The protein size was estimated by appropriate standard proteins and the absorbance values were normalized to the peaks of the curves. (D) Intracellular localization of WT and C586S-mGBPs was analyzed by transduction of the GFP fusion constructs in mGBP2<sup>-/-</sup> MEFs. Cells were stimulated with IFN $\gamma$  for 16 h. Glass slides were analyzed by confocal microscopy. Bars, 5  $\mu$ m.

## Supplement C

**FIGURE S3. Expression analysis of coexpressed mGBP proteins.** Expression levels of mGBPs in postnuclear supernatants of mGBP2<sup>-/-</sup> MEFs reconstituted with G-mGBP2 and coexpressing one of the mCh-mGBPs (mGBP1, mGBP2, mGBP3, mGBP5, mGBP6) were analyzed by Western Blotting. mCherry protein expressing cells served as controls. Cells were stimulated with IFN $\gamma$  for 16 h. Blots were stained with the  $\alpha$ -mCherry antibody.

**FIGURE S4. Intracellular homo- and hetero-multimerization of mGBPs in cells.** (A) For single IFN $\gamma$  stimulated mGBP2<sup>-/-</sup> MEFs expressing G-mGBP2 alone or coexpressing G-mGBP2/mCh-mGBP5, and G-mGBP2/mCh-mGBP6, average values of  $r_D$  in the cytosol (empty) and in the VLS (solid) were plotted against  $\langle \tau_D \rangle_f$  or G-mGBP2 concentrations ( $C_{G\text{-mGBP}2}$ ). See the legend of Figure 4C for the description of the overlay curves in both panels. (B) Corresponding plots as in (A) for single cells expressing G-mGBP2 alone or coexpressing G-mGBP2/mCh-mGBP1, G-mGBP2/mCh-mGBP2 and G-mGBP2/mCh-mGBP3. (C) Corresponding plots as in (A) for *T. gondii* infected cells. (D) Corresponding plots as in (B) for *T. gondii* infected cells.

**FIGURE S5. Quantitative MFIS-FRET analysis of mGBP2 hetero-multimerization in living IFN $\gamma$  stimulated cells.** (A) All the experiments on G-mGBP2/mCh-mGBP1, G-mGBP2/mCh-mGBP2 and G-mGBP2/mCh-mGBP3 interactions were formally analyzed according to eqs. 1-5. Fit results of species fraction of FRET-active complex ( $x_{FRET}$ ) is plotted against G-mGBP2 and mCh-mGBPs concentrations determined in cytosol (red), in VLS (green) and at PVM (blue). The overlaid fuction curve plotting  $x_{FRET} = S \cdot C_{mCh\text{-mGBP}2} / (C_{mCh\text{-mGBP}2} + K_{D,app})$  assumes a mGBP2 Langmuir binding model with apparent dissociation constant,  $K_{D,app} = 9 \mu\text{M}$ , the same value as applied in Figure 4C and 6B. The scaling factor  $S = 0.64$  was adjusted according to the saturation level of  $x_{FRET}$ . (B) For the same experiments as in (A), FRET rate constants ( $k_{FRET}$ ) are plotted versus G-mGBP2 and mCh-mGBPs concentrations. (C)  $x_{FRET}$  in (A) is plotted versus total protein concentration. (D)  $k_{FRET}$  in (B) is plotted versus total protein concentration.

**FIGURE S6. Immunoprecipitation analysis of mGBP proteins.** (A) mGBP2<sup>-/-</sup> MEFs reconstituted with G-mGBP2 or GFP were stimulated with IFN $\gamma$  for 16 h, subsequently lysed and postnuclear

## Supplement C

supernatants were incubated o/n with G-sepharose beads and the  $\alpha$ -GFP antibody at 4°C. IP probes were subjected to Western Blotting. Blots were stained with the  $\alpha$ -mGBP2,  $\alpha$ -mGBP1,  $\alpha$ -mGBP3,  $\alpha$ -mGBP5 antibodies. (B) Postnuclear supernatants of mGBP2<sup>-/-</sup> MEFs reconstituted with G-mGBP2 and coexpressing mCherry protein or one of the mCherry fused mGBPs (mGBP1, mGBP2, mGBP3, mGBP5, mGBP6) were incubated o/n with GFP-Trap® beads at 4°C. IP probes were subjected to Western Blotting. Blots were stained with the  $\alpha$ -GFP and  $\alpha$ -mCherry antibodies.

**FIGURE S7. Sample mGBP2 dimer conformations by MC molecular simulation.** (A) Conformational space sampled by the MC simulations of free mGBP2 is illustrated by the density of the GFP-chromophore, one conformation is shown using cartoon representation. (B) Structural properties of a predicted mGBP2 dimer based on the crystal structure of the hGBP1 dimer (PDB-ID 2BC9). The characteristic FRET features of the dimer with flexibly linked fluorescent proteins can be predicted by calculating inter fluorophore distances from the space that is sterically accessible to the fluorescent proteins. The accessible space of attached fluorescent proteins (green (GFP) and red (mCherry) is depicted as fuzzy cloud;  $\geq 60\%$  of all D-A configurations are FRET-inactive due to their large distances between the fluorophores, methods section 11). (C) Illustration of FRET parameter calculation on each sampled G-mGBP2/mCh-mGBP2 dimer conformation in the MC simulation. Vectors and coordinates in the figure are listed in Table S3. GFP: green, mCherry: red. (D) Donor-acceptor orientation factor ( $\kappa^2$ ), spatial distance ( $R_{sim}$ ) and FRET rate ( $k_{di}$ ) were computed for each sampled mGBP2 dimer conformation, and their relation is plotted in the histogram. In the left panel, the overlay curve in black assumes that the Förster radius between GFP and mCherry is 52 Å, unquenched GFP fluorescence lifetime is 2.6 ns and  $\langle \kappa^2 \rangle$  is 2/3. The red line indicates the maximum resolvable FRET rate constant for our detection system ( $20 \text{ ns}^{-1}$ ). The area shade in grey indicates the irresolvable low FRET rate constant ( $E < 1\%$ ,  $k_{di} < 0.004 \text{ ns}^{-1}$ ), in which the conformations constitute  $\sim 73\%$  of the whole population. (E) Distribution of FRET rate constants ( $k_{FRET}$ ) for mGBP2 dimer (gray curve) and oligomer species (black symbols). Small (black squares) and large (black dots) oligomers as formally differentiated in the pattern-based MFIS-FRET analysis show generally higher  $k_{FRET}$  than that of mGBP2 dimer estimated by the MC simulation. (F) mGBP2 concentration determined by 2D FIDA analysis is plotted versus that

## Supplement C

directly derived from G-mGBP2 fluorescence intensity. (G) A typical image showing the pixels at the PVM area which were analysed by scanning FIDA. The grey scale indicates acquired photon count per pixel. (H) The corresponding 2D FIDA matrix analysing the fluorescence intensity in the green and red detection channel of (G) (the details of FIDA are given in methods section 15).

**Video S1.** mGBP2<sup>-/-</sup> MEFs transduced with G-mGBP2 were treated o/n with IFN $\gamma$  and infected with *T. gondii*. The living cells were observed with a confocal microscope at 37°C and a z-stack was recorded every 5-10 s. 4D data were processed and rendered in normal shading mode. Bar = 5  $\mu$ m.

**Video S2.** mGBP2<sup>-/-</sup> MEFs transduced with G-mGBP2 were treated o/n with IFN $\gamma$  and infected with *T. gondii*. The living cells were observed with a confocal microscope at 37°C and a z-stack was recorded every 5-10 s. 4D data were processed and rendered as maximum intensity projection. Bar = 2  $\mu$ m.

**Video S3.** mGBP2<sup>-/-</sup> MEFs transduced with mCh-mGBP2 were treated o/n with IFN $\gamma$  and infected with *T. gondii*. The living cells were observed with a confocal microscope at 37°C and a z-stack was recorded every 5-10 s. 4D data were processed and rendered as maximum intensity projection. Bar = 1  $\mu$ m.

## Supplement C

### Supplementary Tables

**Supplementary Table 1.** Dissociation constants  $K_D$  of mant-nucleotides for mGBP2 WT and C586S mutant determined by fluorescence titrations and GTPase activity parameters obtained by protein concentration-dependent hydrolysis.

	Nucleotide binding			GTP-hydrolysis		
	<u>nt-GTP<math>\gamma</math>S</u>	<u>mant-GDP</u>	<u>mant-GMP</u>	$K_{max}$ (min <sup>-1</sup> )	Dimer $K_D$ ( $\mu$ M)	GMP (%)
	$K_D$ ( $\mu$ M)	$K_D$ ( $\mu$ M)	$K_D$ ( $\mu$ M)			
WT	0.45	0.54	14.4	102	0.029	74
C586S	0.50	0.45	15.5	133	0.026	72

The % GMP indicates the relative amount of the two products, GDP and GMP

## Supplement C

**Supplementary Table 2.** Amino-acid sequence settings in the MC molecular simulation. The residues used to define the dipole of the chromophoric groups are indicated.

GFP	kept rigid	MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVP WPTLVTTLTLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVK FEGDTLVNRIELKGI <del>DF</del> KEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHN IEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALS <del>KDP</del> NEKRDHMLLEFVTAA
	flexible	GITLGMDELYKSGLRSELNFEFPGASEIHMSEP
mGBP2	kept rigid	MCLIEN <del>TEA</del> Q <del>L</del> VINQ <del>EAL</del> RILSAITQPVVVVAIVGLYRTGKSYLMNKLAGKRTGFSL GSTVQSHTKGIWMWCVPHPKKAGQTLVLLDTEGLE <del>DVE</del> KGDNDQNCWIFALAVLLSS TFIYNSIGTINQQAMDQLHYVTELTDLIKSKSSPDQSGVDDSANFVGGFFPTFVWTLR DFSLELEVNGKPVTSDEYLEHSLTLKKGADKKT <del>KS</del> FNEPRLCIRKFFPKRKC <del>FI</del> FDR PAQRKQLSKLETLREEELCGEFVEQVAEFTSYILSYSSVKTLCGGIIVNGPRLKSLV QTYVGAISNGSLPCMESAVLTLAQIENSAAVQKAITHYEEQMNQKIQMPTETLQELL DLHRPIESEAEIEVFLKNSFKDVDQKFQTELG <del>NLL</del> VAKRDAFIKKNMDVSSARCSDLL EDIFGPLEEEV <del>KL</del> GTFSKPGGYLFLQMRQELEKKYNQAPGKGLQAEAMLKNYFDSK ADV <del>VET</del> L <del>L</del> QTDQSLTEAAKEVEEERTKAEAAEAANRELEKKQKEFELMMQQKEKSYQ EHVKKLTEKMKDEQKQLLAEQENIIAAKLREQEKFLKEGFENESKKLIREIDTLKQN KSSGKCTIL
mCherry	kept rigid	MVSKGEEDNMAI <del>I</del> KEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGG PLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQ DSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLK LKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDY <del>TIV</del>
	flexible	EQYERAEGRHSTGGMDELYKEFPGASEIHMSEP
mGBP2	kept rigid	MCLIEN <del>TEA</del> Q <del>L</del> VINQ <del>EAL</del> RILSAITQPVVVVAIVGLYRTGKSYLMNKLAGKRTGFSL GSTVQSHTKGIWMWCVPHPKKAGQTLVLLDTEGLE <del>DVE</del> KGDNDQNCWIFALAVLLSS TFIYNSIGTINQQAMDQLHYVTELTDLIKSKSSPDQSGVDDSANFVGGFFPTFVWTLR DFSLELEVNGKPVTSDEYLEHSLTLKKGADKKT <del>KS</del> FNEPRLCIRKFFPKRKC <del>FI</del> FDR PAQRKQLSKLETLREEELCGEFVEQVAEFTSYILSYSSVKTLCGGIIVNGPRLKSLV QTYVGAISNGSLPCMESAVLTLAQIENSAAVQKAITHYEEQMNQKIQMPTETLQELL DLHRPIESEAEIEVFLKNSFKDVDQKFQTELG <del>NLL</del> VAKRDAFIKKNMDVSSARCSDLL EDIFGPLEEEV <del>KL</del> GTFSKPGGYLFLQMRQELEKKYNQAPGKGLQAEAMLKNYFDSK ADV <del>VET</del> L <del>L</del> QTDQSLTEAAKEVEEERTKAEAAEAANRELEKKQKEFELMMQQKEKSYQ EHVKKLTEKMKDEQKQLLAEQENIIAAKLREQEKFLKEGFENESKKLIREIDTLKQN KSSGKCTIL

## Supplement C

**Supplementary Table 3.** Calculations of donor-acceptor distances ( $R_{sim}$ ) and orientation factors ( $\kappa^2$ ) from each sampled conformation from MC molecular simulation of G-mGBP2/mCh-mGBP2 dimer in steps. See Experimental procedures and Figure S7 for details.

	Donor (GFP)	Acceptor (mCherry)
Coordinates of the two chosen $C_\alpha$ -atoms	$\vec{r}_{D1}$ and $\vec{r}_{D2}$	$\vec{r}_{A1}$ and $\vec{r}_{A2}$
Distance between the two $C_\alpha$ -atoms	$R_{D,21} = \ \vec{r}_{D2} - \vec{r}_{D1}\ _2$	$R_{A,21} = \ \vec{r}_{A2} - \vec{r}_{A1}\ _2$
Unit vector connecting the two $C_\alpha$ -atoms	$\hat{\mu}_D = \frac{\vec{r}_{D2} - \vec{r}_{D1}}{R_{D,21}}$	$\hat{\mu}_A = \frac{\vec{r}_{A2} - \vec{r}_{A1}}{R_{A,21}}$
Coordinates of the middle point of the connecting vector	$\vec{m}_D = \vec{r}_{D1} + \frac{R_{D,21}}{2} \hat{\mu}_D$	$\vec{m}_A = \vec{r}_{A1} + \frac{R_{A,21}}{2} \hat{\mu}_A$
Calculation of donor-acceptor distance and orientation factor	$\vec{r}_{DA} = \vec{m}_D - \vec{m}_A \quad R_{DA} = \ \vec{m}_D - \vec{m}_A\ _2 \quad \hat{\mu}_{DA} = \frac{\vec{r}_{DA}}{R_{DA}}$ $\kappa^2 = \left( \langle \hat{\mu}_A, \hat{\mu}_D \rangle - 3 \cdot \langle \hat{\mu}_D, \hat{\mu}_{DA} \rangle \cdot \langle \hat{\mu}_A, \hat{\mu}_{DA} \rangle \right)^2$	