Multiparameter Fluorescence Spectroscopy and Microscopy of Biomolecular Systems

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Julian Koch

aus Bergisch-Gladbach

Düsseldorf, Oktober 2022

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

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

Berichterstatter:

1. Prof. Dr. Claus A. M. Seidel

2. Jun.-Prof. Dr. Alexej Kedrov

Tag der mündlichen Prüfung:

24. April 2023

Summary

This work aims to characterize four different biomolecular systems of increasing complexity with fluorescence spectroscopy and imaging as well as biochemical methods *in vitro* and in cells, achieving temporal-spatial information on a molecular and cellular level. Proteins, the most complex biomolecules, maintain their functionality via specific three-dimensional structure and dynamic movements. Therefore, understanding the structure and dynamics of proteins will usually yield functional knowledge as well, which is the key for many major contributions in life sciences. Fluorescence spectroscopy is a highly viable tool to accurately resolve structures and dynamics on a molecular level. Coupled with complementary methods like fluorescence imaging and biochemical basics, we were able to achieve a comprehensive readout over a magnitude of spatial and temporal scales. The first project aimed to characterize an unassigned quenching state found in single labeled T4-lysozyme (T4L). By using truncated T4L variants to suppress internal dynamics, we contextualized the unknown state within the known dynamics of T4L. In the second project, we followed domain movements upon dimerization of human guanylate binding protein 1 (hGBP1) on a millisecond timescale with a stopped-flow coupled ensemble time correlated photon counting setup (SF-eTCSPC). Changes in Förster resonance energy transfer (FRET) efficiency enabled us to monitor the transition of the molecule from a monomer into a (semi-)elongated dimer. The third project aimed to characterize protein-protein interaction and oligomerization behavior of murine guanylate-binding protein 7 (mGBP7) with two other mGBPs, mGBP2 and mGBP3. We used multi-parameter image fluorescence spectroscopy (MFIS), fluorescence recovery after photobleaching (FRAP) and image analysis tools to show and explain the dynamic and specific assembly of mGBP7 with mGBP3 in a co-localized/mixed phase and with mGBP2 in a spatially separated phase. Lastly, the lysine demethylase 6A (KDM6A) was complementary characterized regarding demethylase activity, interaction ability, cellular localization, protein stability of KDM6A variants and their relation to malignant cellular phenotypes. Here, we used a primarily biochemical toolset in synergy with fluorescence spectroscopy and fluorescence imaging. We showed that truncated variants of KDM6A cause significant levels of cellular damage and apoptosis, while substitution variants only show mild effects. Ultimately, this work gives a comprehensive insight into the strengths and limitations of our methods and delivers valuable readout to further investigate and understand the four characterized protein systems.

Zusammenfassung

In dieser Arbeit wurden vier unterschiedliche biomolekulare Systeme steigender Komplexität mit Fluoreszenzspektroskopie, Fluoreszenzmikroskopie und biochemischen Methoden charakterisiert. Dabei wurden zeitlich und räumlich aufgelöste Ergebnisse auf molekularen und zellulären Ebenen erzielt. Proteine, komplexe Biomoleküle, üben ihre Funktion über spezifische dreidimensionale Strukturen und Dynamiken aus. Kennt man die Struktur und die Bewegungsabläufe eines Proteins, kann man funktionelle Zusammenhänge verstehen – das ist der Schlüssel für viele beutende Entdeckungen innerhalb der Lebenswissenschaften. Fluoreszenzspektroskopie ist dabei ein wichtiges Werkzeug, um Strukturen und Dynamiken auf molekularem Level mit hoher Präzision und Richtigkeit aufzulösen. Zusammen mit ergänzenden Methoden aus der Fluoreszenzmikroskopie und biochemischen Methoden können somit Ergebnisse innerhalb einer großen Reichweite in zeitlicher und räumlicher Auflösung erzielt werden. Im ersten Projekt wurde ein fotophysikalischer Prozess unbekannter Herkunft in Fluoreszenzkorrelationsspektroskopie-Messungen von T4-Lysozym (T4L) untersucht. Mithilfe von verkürzten T4L-Proteinen konnten wir den unbekannten Prozess innerhalb der bekannten Dynamiken von T4L einordnen. Im zweiten Projekt konnten Proteindomänenbewegungen im Dimerisierungsprozesses des humanen Guanylat-Bindeprotein 1 (hGBP1) in einem Millisekunden-Zeitfenster dargestellt werden. Dazu wurde eine flussunterbrechende Analyse gekoppelt an eine zeitkorrelierte Einzelphotonenzählung (Ensemble) genutzt (SF-eTCSPC). Änderungen in der Effizienz des Förster-Resonanz Energietransfers ermöglichte uns, die Konformationsänderung von einem Monomer zu einem halbausgestreckten Dimer zu beobachten. Im dritten Projekt wurden die Protein-Protein-Interaktionen und Oligomerisierung des murinen Guanylat-Bindeproteins 7 (mGBP7) mit zwei weiteren mGBPs, mGBP2 und mGBP3 charakterisiert. Durch Einsatz von bildgebender Mehrparameter-Fluoreszenzspektroskopie (MFIS), Fluoreszenzregeneration nach Fotobleichung (FRAP) und Bildanalysen konnte gezeigt werden, dass mGBP7 und mGBP3 in einer kolokalisierten, gemischten Phase vorliegen, während mGBP7 und mGBP2 räumlich voneinander getrennt sind. Abschließend wurde die Lysin-Demethylase 6A (KDM6A) umfassend im Hinblick auf Demethylaseaktivität, Interaktion, Lokalisation, Stabilität und den resultierenden zellulären Phänotypen untersucht. Dafür wurde ein hauptsächlich biochemisches Methodenset in Kombination mit Fluoreszenzmikroskopie und Fluoreszenzspektroskopie verwendet. Wir konnten zeigen, dass verkürzte Varianten von KDM6A zellulären Schaden und Apoptose induzierten, während punktmutierte Varianten lediglich einen milden Effekt auf die Zellen hatten. Insgesamt gibt diese Arbeit einen umfassenden Einblick in die Stärken und Herausforderungen unserer Methoden und liefert viele Einblicke sowie neue Grundlagen für die weitere Arbeit mit den vier Systemen.

Table of Contents

List of figures

List of tables

Introduction

Tabulated thesis digest:

Chapter:	Chapter 1	Chapter 2	Chapter 3	Chapter 4
Bio-molecule:	Bacteriophage T4 Lysozyme (T4L)	human Guanylate-Binding Protein 1 (hGBP1)	murine Guanylate-Binding Protein 7 (mGBP7)	human Histone Lysine Demethylase (KDM6A/UTX)
Aims:	Characterization of truncated and full-length T4L to compare dye environment and assigned a slow (~260 µs) photophysical process in T4L variants.	Resolve domain re-arrangement of dimerized hGBP1 bound to GTPγS on a millisecond timescale.	Determine affinity between mGBP7 and other mGBPs, characterization of oligomerization for two exemplary mGBP combinations.	Analyzing demethylase activity, interaction ability, cellular localization, protein stability of KDM6A variants and observe the relation to cellular phenotypes.
Methods:	Comparison of dynamics found in full-length T4L with N-terminally truncated variants by FCS.	Stopped-flow measurements coupled with TCSPC for ensemble FRET studies.	MFIS-FRET, SR/confocal imaging and analysis, FRAP, bioinformatics.	ELISA, IP/WB, FACS, ICC, MFIS-FRET, SR/confocal imaging & image analysis, molecular biology & design, database analysis.
Outcome:	Resolution of multiple photophysical states. Comparison of full-length versus truncated T4L. Attempt to assign unknown state to internal dynamics.	Structural movement: Identification of the transition times from monomer into a consecutive dimer by fitting photon traces with two kinetic models.	Dimerization affinity determined phase behavior in a cellular environment. Diss. constants yielded (7/7, 7/3, 7/2 with K_D = 0.7, 2.5, >550 µM) are decided by LG-domain similarity.	Truncated variants of KDM6A cause significant levels of cellular damage and apoptosis, while substitution variants only show mild effects.

1. Exciting biomolecules: An introduction to structure and dynamics of biomolecules and implications on fluorescent labeling

Over time and with evolution towards higher organisms exhibiting even more specific tissues, many proteins became exceedingly complex biomolecules [1]. Per definition, biomolecules are mainly organic molecules and macro-molecules with a biological function [2] and can be categorized proteins, lipids, nucleic acids, carbohydrates, metabolites and natural products. Biomolecules are building the fundament of any living organism [2]. They are essential to biological processes such as division, development and morphogenesis [3-6].

Among the biomolecules, proteins are considered the most complex [7]. Driven by the nearly infinite combinations possible by combining all 21 canonical amino acids and rare amino acids [8], the number of proteins translated from the human genome is estimated to exceed 20,000 individual proteins [9]. Furthermore, the function and structure of proteins can be tuned by pH [10], co-factors [11], salt concentration [12], temperature [13], redox environment [14] and post-translational modifications [15]. Triggered by hydrophobicity [16], electrostatic interaction [17] and oftentimes through auxiliary folding by chaperones [18], most proteins assume a highly specific, but not necessarily rigid three-dimensional assembly. The structural hierarchy [2] starts at the primary structure, which is the one-dimensional sequence of amino acids. The secondary structure describes the arrangement of the amino acids into α-helices and β-sheets, while the tertiary structure describes the intra-domain arrangement. Finally, the quaternary structure specifies the way in which proteins interact in multi-protein arrangements. The biomolecular structure usually resembles the functionality of a protein [2]. Sophisticated protein structure can point towards a complex functionality [19], but there exist equally intricate, low-structured, intrinsically disordered proteins (IDPs) [20, 21]. Consequentially, both persistence and intrinsic transition of structural elements eventually add up and determine a protein's functional complexity [22]. Mapping both dynamic and rigid elements is the key to fully understand the function [23, 24].

Functionally, many proteins have implications as enzymes. Enzymes are biocatalyst, speeding up chemical reactions or pushing an equilibrium [25]. The catalytic functions are often connected to a distinct domain movement around the active center or even further away [26], thus acting as 'energetic counterweight' [27]. While those movements can rearrange the whole protein in particular cases [28, 29], sometimes only small angle twists or residual deposition are enough to suffice the protein's catalytic cycle [30]. Ultimately, the catalytic function can also contribute to a protein's complexity [31] and super-fast catalytic cycles or transition states push methods to their time-resolution limits [32, 33].
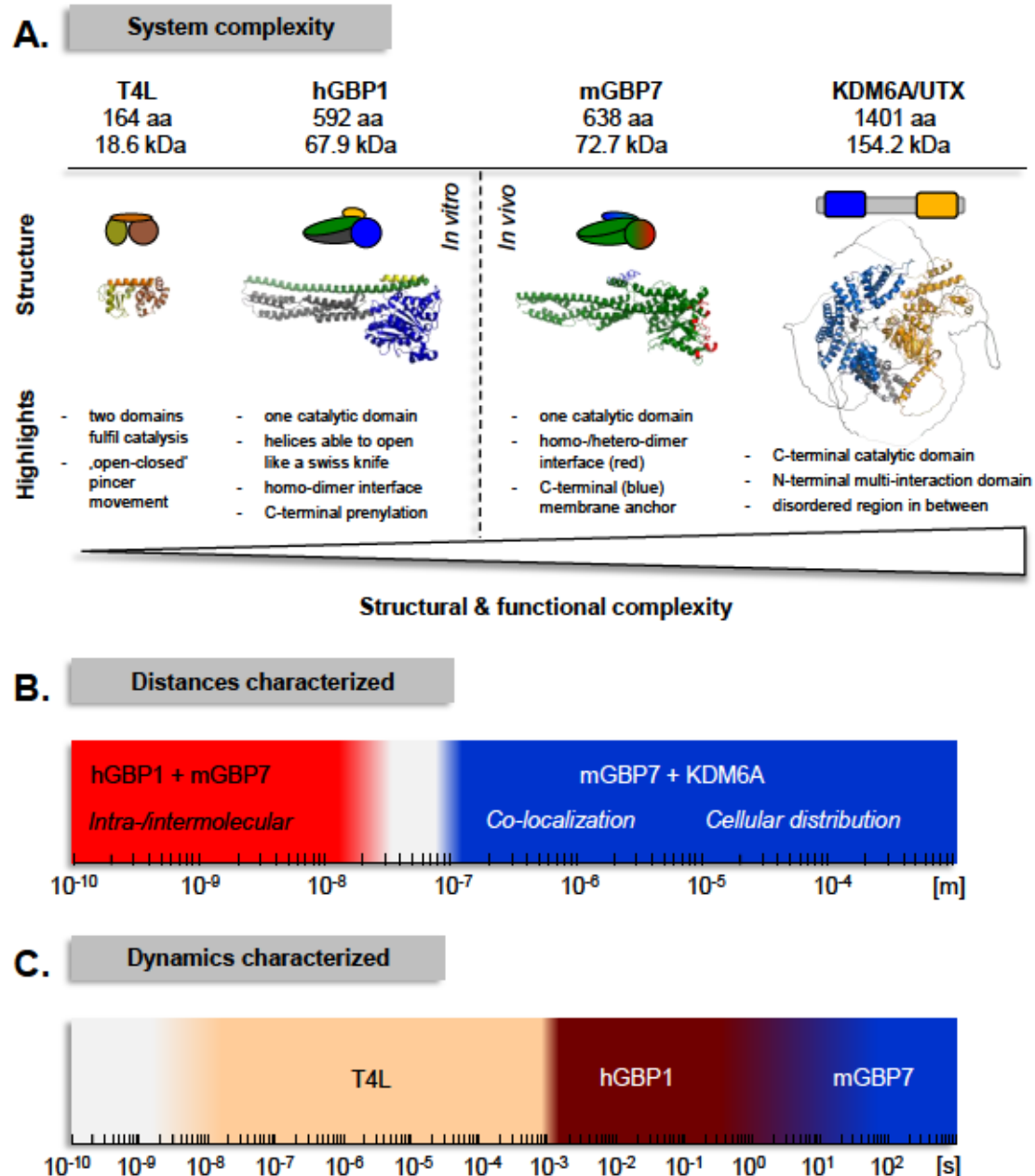
Figure 1: System complexity, dynamics and distances characterized in this work. A. Structural and functional complexity encountered in this work. T4L achieves catalysis with a two-domain minor pincer movement. hGBP1 operates with one catalytic domain (right side of depicted molecule) for GTP-binding and hydrolysis which will favor a homodimer organized via the catalytic domain mGBP7 proposedly behaves similar to hGBP1 and is able to form hetero-dimers with other mGBPs in vivo. KDM6A is a highly complex three-domain protein with a catalytic domain, an interaction domain and an intrinsically disordered domain in between. **B.** We characterized intra- and intermolecular distances in protein complexes in Ångström range. We resolved cellular colocalization and distribution (also ratio) of proteins in the nanometer-micrometer range. **C.** In this work, we characterized the fast conformation dynamics and diffusion of T4L on a nanosecond-microsecond timescale. We further characterized domain movement dynamics of hGBP1 on a millisecond timescale. In cells, we characterized the diffusion of mGBPs into previously bleached particles on a timescale of seconds.

3

All systems used in this work (**Fig. 1A**) exert an enzymatic function: The bacteriophage T4 Lysozyme (T4L) is hydrolyzing glycosidic bonds in bacterial cell walls by using a pincer movement between the domains during the catalytic cycle [34]. Both hGBP1 and mGBP7 hydrolyze GTP to change their domain arrangement and assemble in dimers and subsequent oligomers, which are able to directly attack intracellular parasites [35, 36]. KDM6A can specifically bind to Histone H3K27me3/2 residues and demethylate them, using α-ketoglutarate and Fe(II) as co-substrates and Zn(II) as co-factor [37].

Apart from the catalytic function, many proteins serve as scaffold protein to enable interaction with other proteins in order to trigger pathways and/or assemble into large complexes [38-40]. In our work (**Fig. 1A**), both hGBP1 and mGBP7 contain interaction interfaces to form dimers, a crucial step for the subsequent formation of oligomers and their anti-parasitic function [41, 42]. The scaffolding can also occur in conjunction with the catalytic function [29, 43]. The tetratricopeptide repeat region (TPR) found in KDM6A can interact with complexes which are directly connected to the enzymatic function of KDM6A [44]. Other structure proteins enable organisms to assume and hold shape, by forming or producing molecular meshes or grids [45, 46]. The functionality of antibodies or toxins is also determined by the structure fitting to the target sites [47, 48].

Fluorescence-based techniques can resolve most of the features described above. In order to be spectroscopically visible, proteins need to be labeled with suitable reporters. This is either an artificial fluorophore, attached to the biomolecule in a labeling procedure [49], or a fluorescent protein, usually cloned into the protein's expression gene [50]. Protein tags and fluorescent antibodies enabling specific labeling in cells [51, 52], albeit the spatial resolution is mostly limited to imaging resolution due to their large intrinsic size. It should be considered that any modification might significantly alter the proteins behavior in relation to the unmodified entity [53].

The <u>biomolecular systems characterized in this work</u> (**Fig. 1**) are classified into four projects, shown below. Please refer to the individual chapters for detailed introduction of the system and the respective results.

In **Chapter 1**, we aimed to complement and conclude *in vitro* experiments characterizing movement between the two sub-domains of the bacteriophage T4 Lysozyme (T4L) during the catalytic cycle.

In **Chapter 2**, we aimed to characterize the substantial change in relative domain positioning of the human guanylate-binding protein 1 (hGBP1) *in vitro* upon substrate binding and subsequent dimerization on a magnitude of seconds.

In **Chapter 3**, we aimed to understand the assembly of murine guanylate binding proteins (mGBPs) into dense protein phases known as vesicle-like structures (VLS) *in vivo*.

In **Chapter 4**, we aimed to comprehensively characterize the histone H3K27-specific demethylase 6A (KDM6A or UTX) and its cancer-derived/inspired variants *in vitro and in vivo* with respect to localization, activity, interaction and cellular phenotype observed after expression in urothelial cancer cell lines (UCCs).

2. Illuminating methods: An introduction to fluorescence-based spectroscopy and imaging

In modern day life sciences, especially in medical research and the rapidly advancing biotech sector, knowing the functionality of a protein and using it to one's objective is the key to success [54-56]. Consequentially, a central paradigm of protein structure research is the hunt for the structure-function relationship, aiming to expand the understanding of a protein's functionality via the knowledge of its structure. While the primary structure of a protein is easily extractable from genomic information [57], the correct topology, arrangements and even functionalities are harder to assess [58, 59]. Recently though, advances in deep learning deliver tremendous results on structure prediction from sequence [60, 61]. Quite often, the underlying secondary/tertiary structure elements and even quaternary assemblies are predictable by ubiquitous principles in consensus motifs [62, 63] and structure homology [64, 65]. These methods largely depend on *a posteriori* knowledge, even the recent deep learning algorithms (so far) used empirical resources to train the AI [60]. Hence, hard data gained from various spectroscopic methods builds up the fundaments of most databases on protein structure [66, 67]. By design, many of these classic spectroscopy techniques yield highly resolved protein structure as a rigid crystal but struggle in the resolution of dynamic areas [68, 69]. All the while, knowledge on protein dynamics became more and more relevant, especially since the dynamic parts of the protein are often the functional ones [70]. Both transition times and spatial resolution of flexible regions are of high interest [32]. Although spectroscopic methods are very different in sample requirements and technical limits, modern scientific communities now bring forward interdisciplinary studies, addressing both dynamic and rigid structural elements simultaneously [71-73]. Many of the spectroscopic methods have been interconnected and offer convertible formats to allow a global analysis of spectroscopy data in one effort [74-76].

As one of the fundamental pillars of spectroscopy, fluorescence-based spectroscopy and imaging techniques excel by concurrently delivering multi-magnitude temporal resolution and spatial information [77]. Spatial information can be accurately determined on an Ångström

level due to the direct distance dependency of Förster resonance energy transfer (FRET) [78]. Methodically, all fluorescence-based techniques exploit certain information gained from analysis of the photons emitted by the fluorophore upon controlled excitation with a monochromatic light source [78]. The yielded emission intensity can be used to localize the fluorophore's stationary position in space [79] or to follow the diffusion of molecules [80].

In fluorescence correlation spectroscopy (FCS), fluorescently labeled molecules translationally diffuse through the confocal volume (**Fig. 2B**) which is determined by the point spread function (PSF). Low concentration (low nM to pM) of molecules allows only a few molecules present in the confocal volume at the same time, causing the actual fluctuation in fluorescence to be visible and not being averaged over a large ensemble. The measured fluorescence signal can be self-correlated over an effective time-range between of sub-nanosecond to milliseconds (**Fig. 2E**). The resulting autocorrelation function $G(\tau)$ is the correlation of fluorescence intensity with itself, shifted by the lag-time τ. Whenever an event causes the fluorescent signal to fluctuate, the time-point and amplitude of this event causes the autocorrelation function to abbreviate from the baseline. In this work, we used FCS for our T4L-study to resolve quenching and diffusion dynamic (**Fig. 1B**).

When working with multiple distinguishable fluorophores, the contributions of resonance energy transfer can be assessed by intensity ratios [81] and converted into distance at Ångström level (**Fig 2D**). If the device is able to clock the time-period needed for the photon to be emitted and detected after excitation, this time-resolved spectroscopy can further reveal tempo-spatial features [82] (**Fig. 2E**). Finally, the detection of polarized emission by usage of multiple perpendicularly placed detector units enables to analyze an additional parameter, the anisotropy [76]. The resulting multiparameter fluorescence detection (MFD) can greatly enhance experimental options and data quality [83]. In the last decade, FRET-based and fluorescent spectroscopy steadily developed into a reliable high-accuracy toolset for a temporal-spatial readout to resolve protein dynamics and structure [84]. This cumulated into the latest comparative blind study, where 20 labs independently managed to determine FRET-efficiencies on DNA-duplexes with a standard deviation of ΔE between ± 0.02 and ± 0.05 [85]. Results concerning dynamics are equally impressive [32, 86] and the growing FRET-community is harmonizing approaches, aiming to enhance data accessibility, compatibility and set a gold standard for future FRET-based spectroscopy [84]. For the computational sciences, FRET-assisted comparative modelling has developed into an acclaimed tool to synergize *in silico* with experimental data [77, 87]. This cumulated in the recent addition of FRET-assisted comparative modelling to the PDB-Dev [88], a prototype archiving system for structural models obtained using integrative or hybrid modelling.

Technically, the effective resolution of FRET-techniques is determined by model accuracy, depending on data quality [89] and knowledge of fluorescent dye distributions and orientations [90] as well as specific dye photophysics [91]. All the established FRET-based methods are all able to achieve high-accuracy readout by design. Achieving very high accuracy can already be challenging in a controlled environment, making the comparative results of the first FRET-study [85] even more impressive. Correction factors need to be determined for our methods, even when working with relatively stable and uniform DNA-duplex samples [92, 93]. Calculating FRET-efficiency from fluorescence intensity ratios and fluorescence lifetimes, we need to consider correction factors [81]: i) spectral crosstalk α, ii) direct acceptor excitation δ and iii) the excitation/detection correction factor γ, with the latter one being identified as the main contribute for deviation in the study. The correction factors can be determined by using MFD combined with pulsed interleaved excitation (PIE) [94].

If we increase the system complexity (**Fig. 1**), a couple of additional factors come into play: i) labeling and sample preparation [53], ii) chemical environment of the fluorescent dye [95, 96], iii) internal dynamics and unknown states [32] and iv) growing user bias concerning sample preparation, measurements, data sampling and analysis [84, 97]. Most experiments are carried out *in vitro* at super low single-molecule concentrations (an immense advantage of fluorescence spectroscopy) so that labeling conditions allow the fluorophore to be attached in a well-controlled environment. While labeling, it is crucial that the protein retains its native state [53]. One needs to be sure that experimental conditions are not irreversibly changing the protein and that the fluorescent entity is not interfering with the protein structure. *Vice versa*, one needs to be sure that the protein will not quench the fluorophore, change the quantum yield and eventually meddle with the γ-factor estimation [91]. Consequentially, appropriate control experiments should always be in place [94]. The immense sensitivity of our methods will punish any aberrancy in fluorescence as a visible state. This can distort the data and lead to misinterpretation if the systemic knowledge is limited. Unknown dynamics found in biomolecules can complicate the data but can also be a very rewarding as novel results if resolved correctly [32]. Prior information, especially on basic structure or end-states, is highly valuable for any-well planned FRET-experiment. Consequentially, FRET-based spectroscopy will often excel by further advancing knowledge into a direction where other methods are out of bounds [86, 98]. We used a FRET-readout in an eTCSPC experiment on our hGBP1-system (**Fig. 1B/C**), measuring distances on an inter- and intramolecular magnitude (**Fig. 2D**) with multiple FRET-pairs. By combination with a stopped-flow system, we could follow the dynamic change of the distances on a millisecond timescale (**Fig. 2E**).

Changing from *in vitro* experiments to *in vivo* experiments brings a new set of challenges, namely i) frequent use of terminal fluorescent proteins or tags instead of fluorophores positioned on demand [99], ii) cellular background fluorescence [100] and iii) transient protein expression and cellular heterogeneity. Since labeling conditions in cells are hard to control by design, fluorescence proteins and tags are often the choice for the respective measurements [101]. Fluorescent proteins and most tags are challenging to use as spatial indicators since the positional distribution is wider than for dye fluorophores [99, 102] and the hydrodynamic radius of the reporter proteins already take up the high-efficiency section within the FRET distance range [103]. Photophysics of fluorescent proteins are usually complex [104, 105] and depend heavily on the cellular environment [106], systematic errors are possible. Our spectroscopic methods always retain their high accuracy - but the complex system will confine them to lower accuracy quite often. Still, qualitative interaction, complex assembly ratios and binding affinities can be estimated very well and give unique and valuable information about cellular processes [35, 107]. In both studies carried out in live cells, mGBP7 and KDM6A, we used an MFIS-FRET readout to obtain information about the interaction to a labeled binding partner. We achieved information about spatial proximity (**Fig. 2A**) and could consequentially determine binding affinities in live cells.

As a synergistic method to fluorescence spectroscopy [74], fluorescent imaging is traditionally applied to achieve resolution above the Abbe limit (> 200 nm) [108] (**Fig. 2D**). Recently, with the expansion of light microscopy below the classic resolution limit, super-resolution (SR) fluorescence microscopy techniques are also aiming to achieve molecular resolution on a sub-nanometer level [109, 110] (**Fig. 2D**). Still, fluorescence microscopy is preferably used as a tool to evaluate where and in which ratios proteins assemble in a cellular context. In this work, we used both SR and confocal microscopy (**Fig. 2C**) to i) resolve colocalization between protein species at pixel-resolution and ii) to quantify cellular distribution and to iii) observe cellular phenotypes to categorize them. Compared to fluorescence spectroscopy, fluorescence microscopy is also able to resolve cellular dynamics [111] but on a macroscopic time-scale, following cellular movement or particle diffusion (**Fig. 2E**). We used this technique, fluorescence recovery after photobleaching (FRAP), to follow the diffusion of mGBP7 molecules and their partners in large oligomers (**Fig. 1C**).

A. FRET

high-FRET

mid-FRET

low-FRET

FRET efficiency

1.0

0.5

0

R_0

R [Å]

S_1 — FRET → S_1

S_0

S_0

Donor

Acceptor

B. FCS

z

x

y

PSF

C. Fluorescence imaging/FRAP

1) Laser
2) Mirrors
3) and 4) Ex./Em. filter
5) Dichroic filter
6) Objective
7) Sample
8) Detector

1)
2)
3)
4)
5)
6)
7)
8)

Pre-bleach | Pulse | Recovery — FRAP

Objective

Ex./Em. pattern

Nucleus

Cell

Slide

Confocal imaging

D. Spatial resolution

FRET: eTCSPC/MFIS

Super-resolution Imaging

Confocal Imaging

10^{-10} 10^{-9} 10^{-8} 10^{-7} 10^{-6} 10^{-5} 10^{-4} [m]

E. Temporal resolution

FRET: eTCSPC/MFIS

FCS

Stopped flow + eTCSPC

Imaging (FRAP)

10^{-10} 10^{-9} 10^{-8} 10^{-7} 10^{-6} 10^{-5} 10^{-4} 10^{-3} 10^{-2} 10^{-1} 10^{0} 10^{1} 10^{2} [s]

Figure 2: Principles of fluorescence, FRET and resolution of methods used in this work. A. FRET-efficiency following an inverse sixth-power distance dependency. The Förster radius, R_0, defines the individual point at which two fluorophores are at a FRET-efficiency of 0.5.. The additional Jablonski-diagram depicts excitation of fluorophores and FRET between donor and acceptor. **B.** Fluorescent molecules diffusing through the confocal volume, determined by the PSF. **C.** Application of fluorescent imaging. The Laser (1) beam is directed (2) into an objective (6) through an excitation filter (3). The fluorescent signal emitted from the sample (7) will re-enter the objective, pass the dichroic filter (5) and an emission filter (4) and enter the detector (8). **D.** In this work, we spatially resolved molecular distances within FRET range (<100 Å). Cellular protein distributions in SR-imaging range (~100 nm) and the wider confocal range (>200 nm) was utilized to resolve intracellular distances. **E.** In this work, we temporally resolved the fluorescence decay occurring within nanosecond-range after excitation. We further resolve fast (minor) conformational dynamics and rotation within the nanosecond range, triplet states and diffusion in the microsecond range and slow (major) conformational dynamics in the millisecond range. In cellular systems, we resolve recovery after photobleaching and particle diffusion in the range of seconds.

The <u>fluorescence-based methods used in this work</u> **(Fig. 2)** are distributed over the course of the four main projects. Please refer to the individual chapters for detailed methodical description and data.

In **Chapter 1**, we mainly used fluorescence correlation spectroscopy (FCS) to compare diffusional, conformational and photophysical dynamics.

In **Chapter 2**, we mainly used ensemble time-correlated single photon counting (eTCSPC) with a readout based on Förster resonance energy transfer (FRET) and complemented with a macro-time extension (stopped-flow) to follow kinetics in the magnitude of milliseconds-seconds.

In **Chapter 3**, we mainly used multiparameter fluorescence image spectroscopy (MFIS) and imaging methods, such as confocal microscopy, structured illumination microscopy (SIM) and fluorescence recovery after photobleaching (FRAP) to characterize affinities and phase properties.

In **Chapter 4**, we used both confocal and SR-imaging, utilized auxiliary MFIS measurements and complemented the results with a vast amount of biochemical methods.

Introduction references:

1. Lynch, M. Simple evolutionary pathways to complex proteins. *Protein Science*, 2005, **14**, 2217-2225. DOI: 10.1110/ps.041171805.

2. Berg, J. M., Tymoczko, J. L., Stryer, L. Biochemistry. 2012, Basingstoke: W.H. Freeman.

3. Storck, E. M., Ozbalci, C., Eggert, U. S. Lipid cell biology: a focus on lipids in cell division. *Annual Review of Biochemistry*, 2018, **87**, 839-869. DOI: 10.1146/annurev-biochem-062917-012448.

4. Inaki, M., Liu, J. Y., Matsuno, K. Cell chirality: its origin and roles in left-right asymmetric development. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 2016, **371**. DOI: 10.1098/rstb.2015.0403.

5. Kanada, M., Bachmann, M. H., Hardy, J. W., Frimannson, D. O., Bronsart, L., Wang, A., et al. Differential fates of biomolecules delivered to target cells via extracellular vesicles. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, **112**, E1433-E1442. DOI: 10.1073/pnas.1418401112.

6. Saito, H., Fujita, Y., Kashida, S., Hayashi, K., Inoue, T. Synthetic human cell fate regulation by protein-driven RNA switches. *Nature Communications*, 2011, **2**. DOI: 10.1038/ncomms1157.

7. Boettcher, T. From Molecules to Life: Quantifying the Complexity of Chemical and Biological Systems in the Universe. *Journal of Molecular Evolution*, 2018, **86**, 1-10. DOI: 10.1007/s00239-017-9824-6.

8. Otaki, J. M., Ienaka, S., Gotoh, T., Yamamoto, H. Availability of short amino acid sequences in proteins. *Protein Science*, 2005, **14**, 617-625. DOI: 10.1110/ps.041092605.

9. Salzberg, S. L. Open questions: how many genes do we have? *Bmc Biology*, 2018, **16**. DOI: 10.1186/s12915-018-0564-x.

10. Valery, C., Deville-Foillard, S., Lefebvre, C., Taberner, N., Legrand, P., Meneau, F., et al. Atomic view of the histidine environment stabilizing higher-pH conformations of pH-dependent proteins. *Nature Communications*, 2015, **6**. DOI: 10.1038/ncomms8771.

11. Hemschemeier, A., Happe, T. The plasticity of redox cofactors: from metalloenzymes to redox-active DNA. *Nature Reviews Chemistry*, 2018, **2**, 231-243. DOI: 10.1038/s41570-018-0029-3.

12. Zheng, W. J., Glenn, P. Probing the folded state and mechanical unfolding pathways of T4 lysozyme using all-atom and coarse-grained molecular simulation. *Journal of Chemical Physics*, 2015, **142**. DOI: 10.1063/1.4905606.

13. Rothschild, L. J., Mancinelli, R. L. Life in extreme environments. *Nature*, 2001, **409**, 1092-1101. DOI: 10.1038/35059215.

14. Habich, M., Salscheider, S. L., Murschall, L. M., Hoehne, M. N., Fischer, M., Schorn, F., et al. Vectorial import via a metastable disulfide-linked complex allows for a quality control step and import by the mitochondrial disulfide relay. *Cell Reports*, 2019, **26**, 759-+. DOI: 10.1016/j.celrep.2018.12.092.

15. Hofweber, M., Dormann, D. Friend or foe: post-translational modifications as regulators of phase separation and RNP granule dynamics. *Journal of Biological Chemistry*, 2019, **294**, 7137-7150. DOI: 10.1074/jbc.TM118.001189.

16. Meyer, E. E., Rosenberg, K. J., Israelachvili, J. Recent progress in understanding hydrophobic interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 2006, **103**, 15739-15746. DOI: 10.1073/pnas.0606422103.

17. Zhou, H. X., Pang, X. D. Electrostatic interactions in protein structure, folding, binding and condensation. *Chemical Reviews*, 2018, **118**, 1691-1741. DOI: 10.1021/acs.chemrev.7b00305.

18. Hartl, F. U., Bracher, A., Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature*, 2011, **475**, 324-332. DOI: 10.1038/nature10317.

19. Lehn, J. M. Toward self-organization and complex matter. *Science*, 2002, **295**, 2400-2403. DOI: 10.1126/science.1071063.

20. Romero, P., Obradovic, Z., Li, X. H., Garner, E. C., Brown, C. J., Dunker, A. K. Sequence complexity of disordered protein. *Proteins: Structure Function and Bioinformatics*, 2001, **42**, 38-48. DOI: 10.1002/1097-0134(20010101)42:1<38::Aid-prot50>3.0.Co;2-3.

21. Banani, S. F., Lee, H. O., Hyman, A. A., Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology*, 2017, **18**, 285-298. DOI: 10.1038/nrm.2017.7.

22. Lazar, T., Schad, E., Szabo, B., Horvath, T., Meszaros, A., Tompa, P., et al. Intrinsic protein disorder in histone lysine methylation. *Biology Direct*, 2016, **11**. DOI: 10.1186/s13062-016-0129-2.

23. Schlessinger, A., Rost, B. Protein flexibility and rigidity predicted from sequence. *Proteins-Structure, Function and Bioinformatics*, 2005, **61**, 115-126. DOI: 10.1002/prot.20587.

24. Gaspar, M. E., Csermely, P. Rigidity and flexibility of biological networks. *Briefings in Functional Genomics*, 2012, **11**, 443-456. DOI: 10.1093/bfgp/els023.

25. Alberty, R. A. Biochemical Thermodynamics and Rapid-Equilibrium Enzyme Kinetics. *Journal of Physical Chemistry B*, 2010, **114**, 17003-17012. DOI: 10.1021/jp107337g.

26.   Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., et al. Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, 2005, **438**, 117-121. DOI: 10.1038/nature04105.

27.   Müller, C. W., Schlauderer, G. J., Reinstein, J., Schulz, G. E. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure*, 1996, **4**, 147-156. DOI: 10.1016/s0969-2126(96)00018-4.

28.   Kraft, L., Sprenger, G. A., Lindqvist, Y. Conformational changes during the catalytic cycle of gluconate kinase as revealed by X-ray crystallography. *Journal of Molecular Biology*, 2002, **318**, 1057-1069. DOI: 10.1016/s0022-2836(02)00215-2.

29.   Peulen, T. O., Hengstenberg, C. S., Biehl, R., Dimura, M., Lorenz, C., Valeri, A., et al. Integrative dynamic structural biology unveils conformers essential for the oligomerization of a large GTPase. *arXiv: Biological Physics*, 2020. DOI: arXiv:2004.04229.

30.   Tousignant, A., Pelletier, J. N. Protein motions promote catalysis. *Chemistry & Biology*, 2004, **11**, 1037-1042. DOI: 10.1016/j.chembiol.2004.06.007.

31.   Yoshida, M., Muneyuki, E., Hisabori, T. ATP synthase - a marvellous rotary engine of the cell. *Nature Reviews Molecular Cell Biology*, 2001, **2**, 669-677. DOI: 10.1038/35089509.

32.   Sanabria, H., Rodnin, D., Hemmen, K., Peulen, T. O., Felekyan, S., Fleissner, M. R., et al. Resolving dynamics and function of transient states in single enzyme molecules. *Nature Communications*, 2020, **11**. DOI: 10.1038/s41467-020-14886-w.

33.   Gopich, I. V., Szabo, A. Theory of the statistics of kinetic transitions with application to single-molecule enzyme catalysis. *Journal of Chemical Physics*, 2006, **124**. DOI: 10.1063/1.2180770.

34.   Yirdaw, R. B., McHaourab, H. S. Direct observation of T4 lysozyme hinge-bending motion by fluorescence correlation spectroscopy. *Biophysical Journal*, 2012, **103**, 1525-1536. DOI: 10.1016/j.bpj.2012.07.053.

35.   Kravets, E., Degrandi, D., Ma, Q. J., Peulen, T. O., Klumpers, V., Felekyan, S., et al. Guanylate binding proteins directly attack Toxoplasma gondii via supramolecular complexes. *Elife*, 2016, **5**. DOI: 10.7554/eLife.11479.

36.   Tretina, K., Park, E. S., Maminska, A., MacMicking, J. D. Interferon-induced guanylate-binding proteins: guardians of host defense in health and disease. *Journal of Experimental Medicine*, 2019, **216**, 482-500. DOI: 10.1084/jem.20182031.

37.   Agger, K., Cloos, P. A. C., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., et al. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature*, 2007, **449**, 731-U10. DOI: 10.1038/nature06145.

38. Miller, T., Krogan, N. J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., et al. COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proceedings of the National Academy of Sciences of the United States of America*, 2001, **98**, 12902-12907. DOI: 10.1073/pnas.231473398.

39. Smith, T. F., Gaitatzes, C., Saxena, K., Neer, E. J. The WD repeat: a common architecture for diverse functions. *Trends in Biochemical Sciences*, 1999, **24**, 181-185. DOI: 10.1016/s0968-0004(99)01384-5.

40. Blatch, G. L., Lassle, M. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays*, 1999, **21**, 932-939. DOI: 10.1002/(sici)1521-1878(199911)21:11<932::Aid-bies5>3.3.Co;2-e.

41. Legewie, L., Loschwitz, J., Steffens, N., Prescher, M., Wang, X., Smits, S. H. J., et al. Biochemical and structural characterization of murine GBP7, a guanylate binding protein with an elongated C-terminal tail. *Biochemical Journal*, 2019, **476**, 3161-3182. DOI: 10.1042/bcj20190364.

42. Ghosh, A., Praefcke, G. J. K., Renault, L., Wittinghofer, A., Herrmann, C. How guanylate-binding proteins achieve assembly-stimulated processive cleavage of GTP to GMP. *Nature*, 2006, **440**, 101-104. DOI: 10.1038/nature04510.

43. Völpel, T., Syguda, A., Britzen-Laurent, N., Kunzelmann, S., Ludemann, M. B., Dovengerds, C., et al. Mechanism of GTPase-activity-induced self-assembly of human guanylate binding protein 1. *Journal of Molecular Biology*, 2010, **400**, 63-70. DOI: 10.1016/j.jmb.2010.04.053.

44. Kato, H., Asamitsu, K., Sun, W. D., Kitajima, S., Yoshizawa-Sugata, N., Okamoto, T., et al. Cancer-derived UTX TPR mutations G137V and D336G impair interaction with MLL3/4 complexes and affect UTX subcellular localization. *Oncogene*, 2020, **39**, 3322-3335. DOI: 10.1038/s41388-020-1218-3.

45. Discher, D. E., Janmey, P., Wang, Y. L. Tissue cells feel and respond to the stiffness of their substrate. *Science*, 2005, **310**, 1139-1143. DOI: 10.1126/science.1116995.

46. Schleifer, K. H., Kandler, O. Peptidoglycan types of bacterial cell-walls and their taxonomic implications. *Bacteriological Reviews*, 1972, **36**, 407-477. DOI: 10.1128/mmbr.36.4.407-477.1972.

47. Cassel, D., Pfeuffer, T. Mechanism of cholera toxin action - covalent modification of guanyl nucleotide-binding protein of adenylate-cyclase system. *Proceedings of the National Academy of Sciences of the United States of America*, 1978, **75**, 2669-2673. DOI: 10.1073/pnas.75.6.2669.

48. Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., Hendrickson, W. A. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*, 1998, **393**, 648-659. DOI: 10.1038/31405.

49. Hawe, A., Sutter, M., Jiskoot, W. Extrinsic fluorescent dyes as tools for protein characterization. *Pharmaceutical Research*, 2008, **25**, 1487-1499. DOI: 10.1007/s11095-007-9516-9.

50. Giepmans, B. N. G., Adams, S. R., Ellisman, M. H., Tsien, R. Y. The fluorescent toolbox for assessing protein location and function. *Science*, 2006, **312**, 217-224. DOI: 10.1126/science.1124618.

51. Gautier, A., Juillerat, A., Heinis, C., Correa, I. R., Kindermann, M., Beaufils, F., et al. An engineered protein tag for multiprotein labeling in living cells. *Chemistry & Biology*, 2008, **15**, 128-136. DOI: 10.1016/j.chembiol.2008.01.007.

52. Murrey, H. E., Judkins, J. C., Ende, C. W. A., Ballard, T. E., Fang, Y. Z., Riccardi, K., et al. Systematic evaluation of bioorthogonal reactions in live cells with clickable halo tag ligands: implications for intracellular imaging. *Journal of the American Chemical Society*, 2015, **137**, 11461-11475. DOI: 10.1021/jacs.5b06847.

53. Quinn, M. K., Gnan, N., James, S., Ninarello, A., Sciortino, F., Zaccarelli, E., et al. How fluorescent labelling alters the solution behaviour of proteins. *Physical Chemistry Chemical Physics*, 2015, **17**, 31177-31187. DOI: 10.1039/c5cp04463d.

54. Yan, R. H., Zhang, Y. Y., Li, Y. N., Xia, L., Guo, Y. Y., Zhou, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*, 2020, **367**, 1444-+. DOI: 10.1126/science.abb2762.

55. Patterson, G. H., Lippincott-Schwartz, J. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science*, 2002, **297**, 1873-1877. DOI: 10.1126/science.1074952.

56. Cong, L., Ran, F. A., Cox, D., Lin, S. L., Barretto, R., Habib, N., et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 2013, **339**, 819-823. DOI: 10.1126/science.1231143.

57. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. The sequence of the human genome. *Science*, 2001, **291**, 1304-+. DOI: 10.1126/science.1058040.

58. Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C., Doyle, F., et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 2012, **489**, 57-74. DOI: 10.1038/nature11247.

59. Kuhlman, B., Bradley, P. Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology*, 2019, **20**, 681-697. DOI: 10.1038/s41580-019-0163-x.

60. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. DOI: 10.1038/s41586-021-03819-2.

61. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 2021, **373**, 871-876. DOI: 10.1126/science.abj8754.

62. Miyashita, T., Reed, J. C. Tumor-suppressor p53 is a direct transcriptional activator of the human BAX gene. *Cell*, 1995, **80**, 293-299. DOI: 10.1016/0092-8674(95)90412-3.

63. Dingwall, C., Laskey, R. A. Nucelar targeting sequences - a consensus. *Trends in Biochemical Sciences*, 1991, **16**, 478-481. DOI: 10.1016/0968-0004(91)90184-w.

64. Arnold, K., Bordoli, L., Kopp, J., Schwede, T. The SWISS-model workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 2006, **22**, 195-201. DOI: 10.1093/bioinformatics/bti770.

65. Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research*, 2013, **41**. DOI: 10.1093/nar/gkt263.

66. Berman, H., Henrick, K., Nakamura, H., Markley, J. L. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Research*, 2007, **35**, D301-D303. DOI: 10.1093/nar/gkl971.

67. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. The Protein Data Bank. *Nucleic Acids Research*, 2000, **28**, 235-242. DOI: 10.1093/nar/28.1.235.

68. Krissinel, E., Henrick, K. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology*, 2007, **372**, 774-797. DOI: 10.1016/j.jmb.2007.05.022.

69. Li, X. M., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods*, 2013, **10**, 584-+. DOI: 10.1038/nmeth.2472.

70. Karplus, M., McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 2002, **9**, 646-652. DOI: 10.1038/nsb0902-646.

71. Möckel, C., Kubiak, J., Schillinger, O., Kühnemuth, R., Della Corte, D., Schröder, G. F., et al. Integrated NMR, fluorescence, and molecular dynamics benchmark study of protein mechanics and hydrodynamics. *Journal of Physical Chemistry B*, 2019, **123**, 1453-1480. DOI: 10.1021/acs.jpcb.8b08903.

72. Viegas, A., Dollinger, P., Verma, N., Kubiak, J., Viennet, T., Seidel, C. A. M., et al. Structural and dynamic insights revealing how lipase binding domain MD1 of Pseudomonas aeruginosa foldase affects lipase activation. *Scientific Reports*, 2020, **10**. DOI: 10.1038/s41598-020-60093-4.

73. Stella, S., Mesa, P., Thomsen, J., Paul, B., Alcon, P., Jensen, S. B., et al. Conformational activation promotes CRISPR-Cas12a catalysis and resetting of the endonuclease activity. *Cell*, 2018, **175**, 1856-+. DOI: 10.1016/j.cell.2018.10.045.

74. Wallrabe, H., Periasamy, A. Imaging protein molecules using FRET and FLIM microscopy. *Current Opinion in Biotechnology*, 2005, **16**, 19-27. DOI: 10.1016/j.copbio.2004.12.002.

75. Budde, J. H., Van der Voort, N. T. M., Felekyan, S., Folz, J., Kühnemuth, R., Lauterjung, P., et al. FRET nanoscopy enables seamless imaging of molecular assemblies with sub-nanometer resolution. *arXiv: Physics.optics*, 2021. DOI: arXiv:2108.00024.

76. Weidtkamp-Peters, S., Felekyan, S., Bleckmann, A., Simon, R., Becker, W., Kühnemuth, R., et al. Multiparameter fluorescence image spectroscopy to study molecular interactions. *Photochemical & Photobiological Sciences*, 2009, **8**, 470-480. DOI: 10.1039/b903245m.

77. Dimura, M., Peulen, T. O., Hanke, C. A., Prakash, A., Gohlke, H., Seidel, C. A. M. Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. *Current Opinion in Structural Biology*, 2016, **40**, 163-185. DOI: 10.1016/j.sbi.2016.11.012.

78. Lakowicz, J. R. Principles of fluorescence spectroscopy. 2006, New York: Springer.

79. Testa, I., Wurm, C. A., Medda, R., Rothermel, E., von Middendorf, C., Folling, J., et al. Multicolor fluorescence nanoscopy in fixed and living cells by exciting conventional fluorophores with a single wavelength. *Biophysical Journal*, 2010, **99**, 2686-2694. DOI: 10.1016/j.bpj.2010.08.012.

80. Eggeling, C., Ringemann, C., Medda, R., Schwarzmann, G., Sandhoff, K., Polyakova, S., et al. Direct observation of the nanoscale dynamics of membrane lipids in a living cell. *Nature*, 2009, **457**, 1159-U121. DOI: 10.1038/nature07596.

81. Sisamakis, E., Valeri, A., Kalinin, S., Rothwell, P. J., Seidel, C. A. M. Accurate single molecule FRET studies using multiparameter fluorescence detection. *Methods in Enzymology*, 2010, **475**, 455-514. DOI: 10.1016/s0076-6879(10)75018-7.

82. Felekyan, S., Kühnemuth, R., Kudryavtsev, V., Sandhagen, C., Becker, W., Seidel, C. A. M. Full correlation from picoseconds to seconds by time-resolved and time-correlated single photon detection. *Review of Scientific Instruments*, 2005, **76**. DOI: 10.1063/1.1946088.

83. Kühnemuth, R., Seidel, C. A. M. Principles of single molecule multiparameter fluorescence spectroscopy. *Single Molecules*, 2001, **2**, 251-254. DOI: 10.1002/1438-5171(200112)2:4<251::AID-SIMO251>3.0.CO;2-T.

84. Lerner, E., Barth, A., Hendrix, J., Ambrose, B., Birkedal, V., Blanchard, S. C., et al. FRET-based dynamic structural biology: challenges, perspectives and an appeal for open-science practices. *Elife*, 2021, **10**. DOI: 10.7554/eLife.60416.

85. Hellenkamp, B., Schmid, S., Doroshenko, O., Opanasyuk, O., Kühnemuth, R., Adariani, S. R., et al. Precision and accuracy of single-molecule FRET measurements-a multi-laboratory benchmark study. *Nature Methods*, 2018, **15**, 669-+. DOI: 10.1038/s41592-018-0085-0.

86. Kilic, S., Felekyan, S., Doroshenko, O., Boichenko, I., Dimura, M., Vardanyan, H., et al. Single-molecule FRET reveals multiscale chromatin dynamics modulated by HP1 alpha. *Nature Communications*, 2018, **9**. DOI: 10.1038/s41467-017-02619-5.

87. Dimura, M., Peulen, T. O., Sanabria, H., Rodnin, D., Hemmen, K., Hanke, C. A., et al. Automated and optimally FRET-assisted structural modeling. *Nature Communications*, 2020, **11**. DOI: 10.1038/s41467-020-19023-1.

88. Vallat, B., Webb, B., Westbrook, J. D., Sali, A., Berman, H. M. Development of a prototype system for archiving integrative/hybrid structure models of biological macromolecules. *Structure*, 2018, **26**, 894-+. DOI: 10.1016/j.str.2018.03.011.

89. Woehler, A., Wlodarczyk, J., Neher, E. Signal/noise analysis of FRET-based sensors. *Biophysical Journal*, 2010, **99**, 2344-2354. DOI: 10.1016/j.bpj.2010.07.053.

90. Kalinin, S., Peulen, T. O., Sindbert, S., Rothwell, P. J., Berger, S., Restle, T., et al. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nature Methods*, 2012, **9**, 1218-U129. DOI: 10.1038/nmeth.2222.

91. Kalinin, S., Sisamakis, 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. *Journal of Physical Chemistry B*, 2010, **114**, 6197-6206. DOI: 10.1021/jp100025v.

92. Wozniak, A. K., Schröder, G. F., Grubmüller, H., Seidel, C. A. M., Oesterhelt, F. Single-molecule FRET measures bends and kinks in DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105**, 18337-18342. DOI: 10.1073/pnas.0800977105.

93. Sindbert, S., Kalinin, S., Hien, N., Kienzler, A., Clima, L., Bannwarth, W., et al. Accurate distance determination of nucleic acids via Förster resonance energy tansfer: implications of dye linker length and rigidity. *Journal of the American Chemical Society*, 2011, **133**, 2463-2480. DOI: 10.1021/ja105725e.

94. Kudryavtsev, V., Sikor, M., Kalinin, S., Mokranjac, D., Seidel, C. A. M., Lamb, D. C. Combining MFD and PIE for accurate single-pair Förster resonance energy transfer measurements. *Chemphyschem*, 2012, **13**, 1060-1078. DOI: 10.1002/cphc.201100822.

95. Mansoor, S. E., McHaourab, H. S., Farrens, D. L. Mapping proximity within proteins using fluorescence spectroscopy. A study of T4 lysozyme showing that tryptophan residues quench bimane fluorescence. *Biochemistry*, 2002, **41**, 2475-2484. DOI: 10.1021/bi011198i.

96. Jameson, L. P., Smith, N. W., Annunziata, O., Dzyuba, S. V. Interaction of BODIPY dyes with bovine serum albumin: a case study on the aggregation of a click-BODIPY dye. *Physical Chemistry Chemical Physics*, 2016, **18**, 14182-14185. DOI: 10.1039/c6cp00420b.

97. Thomsen, J., Sletfjerding, M. B., Jensen, S. B., Stella, S., Paul, B., Malle, M. G., et al. DeepFRET, a software for rapid and automated single-molecule FRET data classification using deep learning. *Elife*, 2020, **9**. DOI: 10.7554/eLife.60404.

98. Lehmann, K., Felekyan, S., Kühnemuth, R., Dimura, M., Toth, K., Seidel, C. A. M., et al. Dynamics of the nucleosomal histone H3 N-terminal tail revealed by high precision single-molecule FRET. *Nucleic Acids Research*, 2020, **48**, 1551-1571. DOI: 10.1093/nar/gkz1186.

99. Cranfill, P. J., Sell, B. R., Baird, M. A., Allen, J. R., Lavagnino, Z., de Gruiter, H. M., et al. Quantitative assessment of fluorescent proteins. *Nature Methods*, 2016, **13**, 557-+. DOI: 10.1038/nmeth.3891.

100. Ecker, R. C., de Martin, R., Steiner, G. E., Schmid, J. A. Application of spectral imaging microscopy in cytomics and fluorescence resonance energy transfer (FRET) analysis. *Cytometry Part A*, 2004, **59A**, 172-181. DOI: 10.1002/cyto.a.20053.

101. Albertazzi, L., Arosio, D., Marchetti, L., Ricci, F., Beltram, F. Quantitative FRET Analysis With the eGFP-mCherry Fluorescent Protein Pair. *Photochemistry and Photobiology*, 2009, **85**, 287-297. DOI: 10.1111/j.1751-1097.2008.00435.x.

102. Day, R. N., Davidson, M. W. Fluorescent proteins for FRET microscopy: monitoring protein interactions in living cells. *Bioessays*, 2012, **34**, 341-350. DOI: 10.1002/bies.201100098.

103. Mastop, M., Bindels, D. S., Shaner, N. C., Postma, M., Gadella, T. W. J., Goedhart, J. Characterization of a spectrally diverse set of fluorescent proteins as FRET acceptors for mTurquoise2. *Scientific Reports*, 2017, **7**. DOI: 10.1038/s41598-017-12212-x.

104. Ghosh, A., Isbaner, S., Veiga-Gutierrez, M., Gregor, I., Enderlein, J., Karedla, N. Quantifying microsecond transition times using fluorescence lifetime correlation spectroscopy. *Journal of Physical Chemistry Letters*, 2017, **8**, 6022-6028. DOI: 10.1021/acs.jpclett.7b02707.

105. Cotlet, M., Hofkens, J., Maus, M., Gensch, T., Van der Auweraer, M., Michiels, J., et al. Excited-state dynamics in the enhanced green fluorescent protein mutant probed

by picosecond time-resolved single photon counting spectroscopy. *Journal of Physical Chemistry B*, 2001, **105**, 4999-5006. DOI: 10.1021/jp003813i.

106. Kirk, W., Allen, T., Atanasova, E., Wessels, W., Yao, J., Prendergast, F. Photophysics of eGFP (E222H) mutant, with comparisons to model chromophores: excited state pK's, progressions, quenching and exciton interaction. *Journal of Fluorescence*, 2017, **27**, 895-919. DOI: 10.1007/s10895-017-2025-2.

107. Greife, A., Felekyan, S., Ma, Q. J., Gertzen, C. G. W., Spomer, L., Dimura, M., et al. Structural assemblies of the di- and oligomeric G-protein coupled receptor TGR5 in live cells: an MFIS-FRET and integrative modelling study. *Scientific Reports*, 2016, **6**. DOI: 10.1038/srep36792.

108. Hell, S. W. Far-field optical nanoscopy. *Science*, 2007, **316**, 1153-1158. DOI: 10.1126/science.1137395.

109. Balzarotti, F., Eilers, Y., Gwosch, K. C., Gynna, A. H., Westphal, V., Stefani, F. D., et al. Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. *Science*, 2017, **355**, 606-612. DOI: 10.1126/science.aak9913.

110. Gwosch, K. C., Pape, J. K., Balzarotti, F., Hoess, P., Ellenberg, J., Ries, J., et al. MINFLUX nanoscopy delivers 3D multicolor nanometer resolution in cells. *Nature Methods*, 2020, **17**, 217-+. DOI: 10.1038/s41592-019-0688-0.

111. Li, P. L., Banjade, S., Cheng, H. C., Kim, S., Chen, B., Guo, L., et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature*, 2012, **483**, 336-U129. DOI: 10.1038/nature10879.

Chapter 1: Expression and experimental evaluation of a stable T4 lysozyme fragment as a platform for dye quenching studies in a low-dynamic environment

Chapter digest

Aims	Methods	Outcome
Characterization of a slow (~260 µs) photophysical process in T4L variants.	FCS with full-length (FL) T4L compared with N-terminally truncated (CT) variants.	FL-T4L and CT-T4L perform

Contributions

dN-T4L protein expression and troubleshooting: *Julian Koch, Aiswaria Prakash*. CT-T4L and full-length T4L protein expression, labeling, measurement and analysis: *Julian Koch, Alexander Larbig*. Initial full-length T4L protein expression, labeling, measurement and analysis: *Katharina Hemmen, Aiswaria Prakash, Laura Steffens*. Project supervision: *Claus A. M. Seidel*.

Abstract

T4 lysozyme (T4L), expressed by bacteriophage T4, is a protein that hydrolyzes 1,4-beta linkages between N-acetyl-D-glucosamine and N-acetylmuramic acid of peptidoglycans in prokaryotic cell walls, initiating their lysis. Due to an excellent characterization, T4L is used as a model protein in protein stability and was established as a scaffold platform to test environment-specific fluorescent dye behavior in a rigid system. Using Multiparameter Fluorescence Detection (MFD), we observed quenching of unknown origin in full-length T4L. This quenching was initially attributed to the known intermolecular cycling between three conformational states (C1-C3) in T4L, involving both N- and C-terminus. We therefore generated an N-terminally truncated construct, delta-N (dN-T4L), to avoid any effects caused by internal dynamics. Initially encountering stability issues, we retained the stabilizing helix connecting N- and C-terminus. With the resulting C-terminal variant, CT-T4L, we managed to express multiple stable N-terminally truncated T4L species. We directly compared this truncated protein to the full-length T4L species in fluorescence correlation spectroscopy (FCS). We confirmed three labeled variants as strongly quenched (R119C, D127C, N132C), one variant as hardly quenched (P86C) and one variant (I150C) inept to (good) labeling. We were unable to attribute the observed slow quenching to internal dynamics. We conclude that, although the variants used in this study are suitable for quenching study, they are not sufficient to elucidate the dynamics of all the different conformational states of T4L.

1. Introduction

The well-characterized [1] T4 lysozyme is a routinely used protein to probe for and verify models on protein stability depending on temperature [2] and pH [3], energy landscapes [4], experimental [5] and modeled [6] structure-thermodynamics relationship, ligand binding [7] and protein dynamics [8]. Since fluorescent spectroscopy is a powerful tool for the elucidation of protein dynamics, there are multiple contributing studies available within the scope of T4L characterization [9-12]. Functionally, T4L is a monomeric protein that hydrolyzes peptidoglycan, effectively breaking the β-(1,4)-glycosidic bonds between its fundamental components [13]. As gram-positive bacteria use a cell wall consisting of peptidoglycan as fundamental structural component [14], T4L has a profound implication as antibiotic component [15]. The structure-function relationship of T4L is determined by two main conformers [16, 17]: The more open state 'C1' to initiate substrate binding and a closed, substrate-bound conformation 'C2' were the actual hydrolysis takes place. Recently, an additional and transient third conformer 'C3' was identified, being involved into an active release of the product [18]. Consequentially, T4L cycles between those states during catalysis and even in a substrate free state, although C1 emerges as the dominant conformer [18]. The state interchange is characterized by a hinge-movement [19] between the two domains of T4L (see **Fig. 1A/B**), the N-terminal subdomain (NTsD) and the C-terminal subdomain (CTsD), topologically clasped around the active center which will slightly open and close [20]. The NTsD (olive) and CTsD (brown) will move towards each other in a pincer movement when cycling between C1, C2 and C3. Substrate binding occurs in the pocket between both domains, the NTsD binds the glycoside part while the CTsD largely binds the tetra-peptide of peptidoglycan [16]. The linker helix (orange) acts as the hinge between both domains, it bends and kinks back and forth slightly while cycling through the conformers.

Although the internal dynamics found in the three major states of T4L might complicate the readout of spectroscopic methods aimed at a rigid, non-dynamic target, T4L has received major interest as a model system for quenching studies, in particular on tryptophan as a natural photophysical quencher [11, 21, 22]. Tryptophan as an UV-emitter [23] is already a tempting target but comes with the disadvantage of a strong dependence on the chemical environment as the indol-moiety is often embedded or topologically close to other residues and as such heavily exposed [24] to variable dipol-dipol coupling, resulting in a complex multi-exponential fluorescence lifetime decay. A more expedient application for tryptophan is the role as a quencher for a fluorescent dye in a Förster Resonance Energy Transfer (FRET) or Photoinduced Electron Transfer (PET) experiment [25, 26].
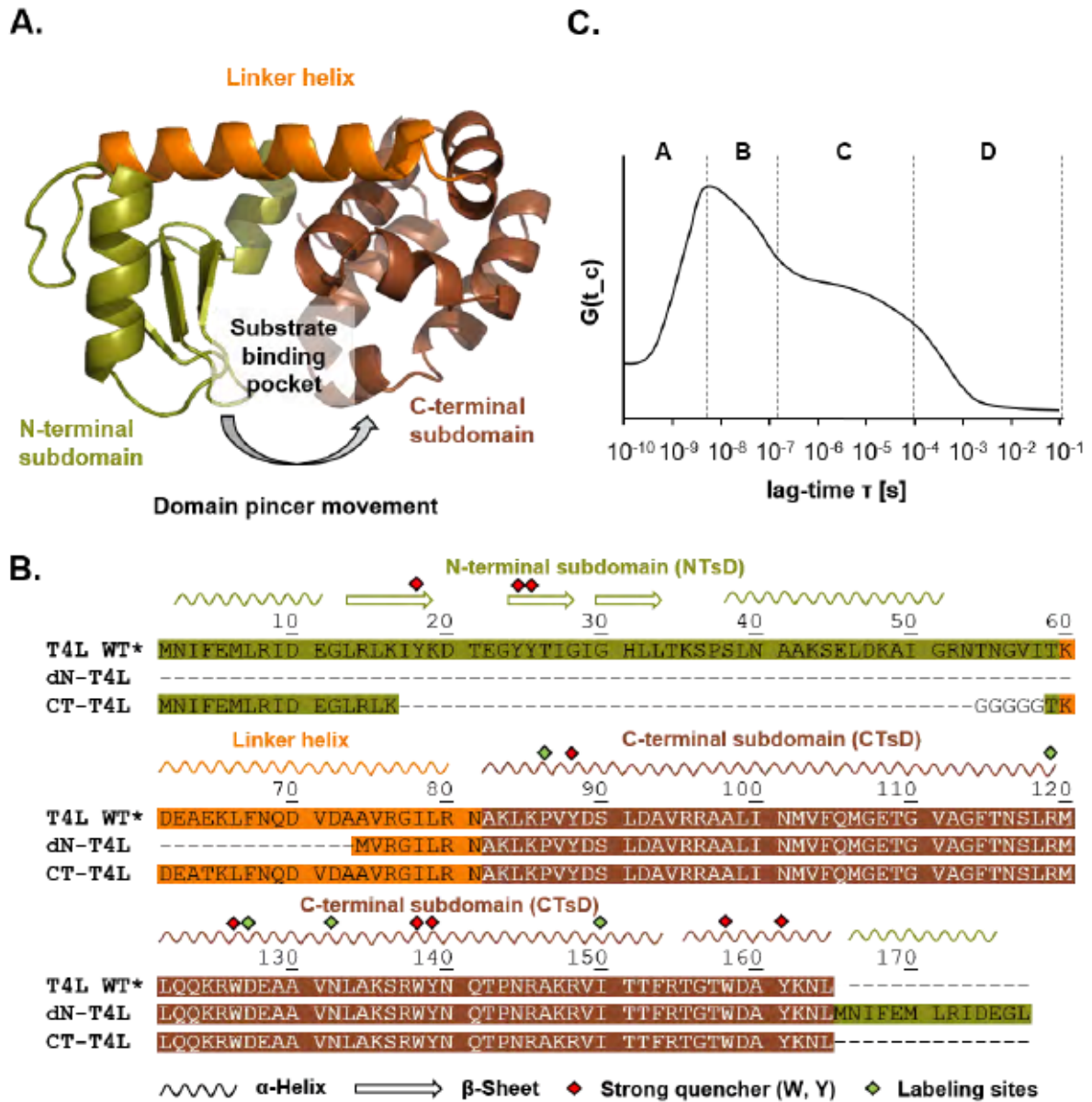
Figure 1: Structure and sequence of T4I and truncated variants, principle of FCS. A. Crystal structure (PDB: 148L) of T4L representing the open state (C1). **B.** Sequence alignment of full-length T4L, dN-T4L and CT-T4L. Domains are color coded according to crystal structure. Structural elements are highlighted above sequence. **C.** FCS, explanation of the autocorrelation function. **Segment A** in the representation of G(τ) contains the photon antibunching, which centers around the average lifetime of the fluorescence dye – usually a sub-nanosecond up to a couple of nanoseconds [27]. PET also takes place in this time-window [26, 28]. **Segment B** represents the time-scale of fast folding or very fast conformational dynamics [29] as well as rotational diffusion of molecules with a defined dipole in low-viscosity medium [30]. **Segment C** is dominated by intersystem crossing from excited states into dark triplet states, depending on the laser power, dye characteristics and the presence of triplet quencher in the system [31]. Lastly, the **Segment D** mostly represents the translational diffusion of the fluorescent molecules, depending on the diffusion coefficient of the molecule [25].

To resolve these different quenching effects on their respective timescales, FCS (see **Fig. 1C**) is the method of choice. Overall, there is a high experimental potential for controlled and well-designed dye-quencher systems using the fluorescence spectroscopy techniques FCS and FRET. Ambiguously, fluorescent dye quenching by tryptophan can also heavily disturb FRET experiments by strong static quenching, worst-case resulting in a multi-exponential decay of the fluorescent dye before FRET even occurred. It is therefore highly beneficial to understand fluorescent dye quenching by in detail. Appropriately, T4L has three natural tryptophan residues in the CTsD at residues W126, W138 and W158, which have all been characterized towards their chemical environment and resulting photophysical behavior [11, 21, 22, 24]. In addition, the weaker quenchers tyrosine, methionine, arginine and histidine in T4L further pronounce the effect. This well-characterized and unique quenching environment motivated our working group to establish T4L as a scaffold platform to test for dye quenching in a controlled environment. Prior to this consecutive project, five single-cysteine variants (T4L P86C, R119C, D127C, N132C and I150C) based on the WT* construct C54T/C97A were designed, labeled with Alexa Fluor 488 and analyzed [32]. While results were promising in underlining the strong influence of tryptophan as quencher, we noticed quenching on a timescale in the higher microsecond range in FCS measurements. If such a quenching effect appears it is expected to originate from domain movements, possibly an interchange between the three main conformers of T4L. Intriguingly, the rate constant for the 'C3'-'C2' interchange has been determined with $k_{c3\rightarrow c2} = 3.7$ ms^{-1} [18], fitting very well into the approx. average $\tau = 260$ μs observed for the quenched state in the FCS measurements [32]. Therefore, we designed a set of truncated T4L variants, which partially lack the N-terminal domain (**Fig. 1B**), to get a 'motionless' prospect on the dye quenching and elucidate the nature of the slow-quenched state. Both variants loose weaker N-terminal quenchers (Y18, Y24, Y25, R52, additionally M1, M6, R8, R14 while they are moved to the C-terminus in dN-T4L, altering their environment) but maintain all three tryptophan residues in the C-terminal part. These truncated variants were measured back-to-back with the full-length variants by FCS to compare fit result and conclude if the truncation impacts the dye environment. Furthermore, we wanted to find the origin of the unknown quenching state and try to attribute it to internal dynamics.

2. Results

We first established dN-T4L variants (see **Fig. 1B** for sequence, **Fig. 2A** for structure) based on Llinás et al. [33]. Here, the first 12 amino acids of T4L are attached to the C-terminus. The N- and C-terminus of T4L are in close proximity. The rest of the NTsD (olive) is missing as well as 68% of the linker helix (orange). The CTsD (brown) was fully incorporated into this

construct. After multiple expressions, it became evident that the dN-T4L variants (based on the T4L WT*) P86C, R119C, D127C, N132C and I150C precipitated as inclusion bodies [34]. Dealing with inclusion bodies usually requires a solution step, where the aggregates are dissolved in chaotropic agents like urea or guanidinium chloride (GdmCl) and a refolding step, where the monomeric, unfolded protein is returned to its native state [35, 36] (**Fig. 2B**).
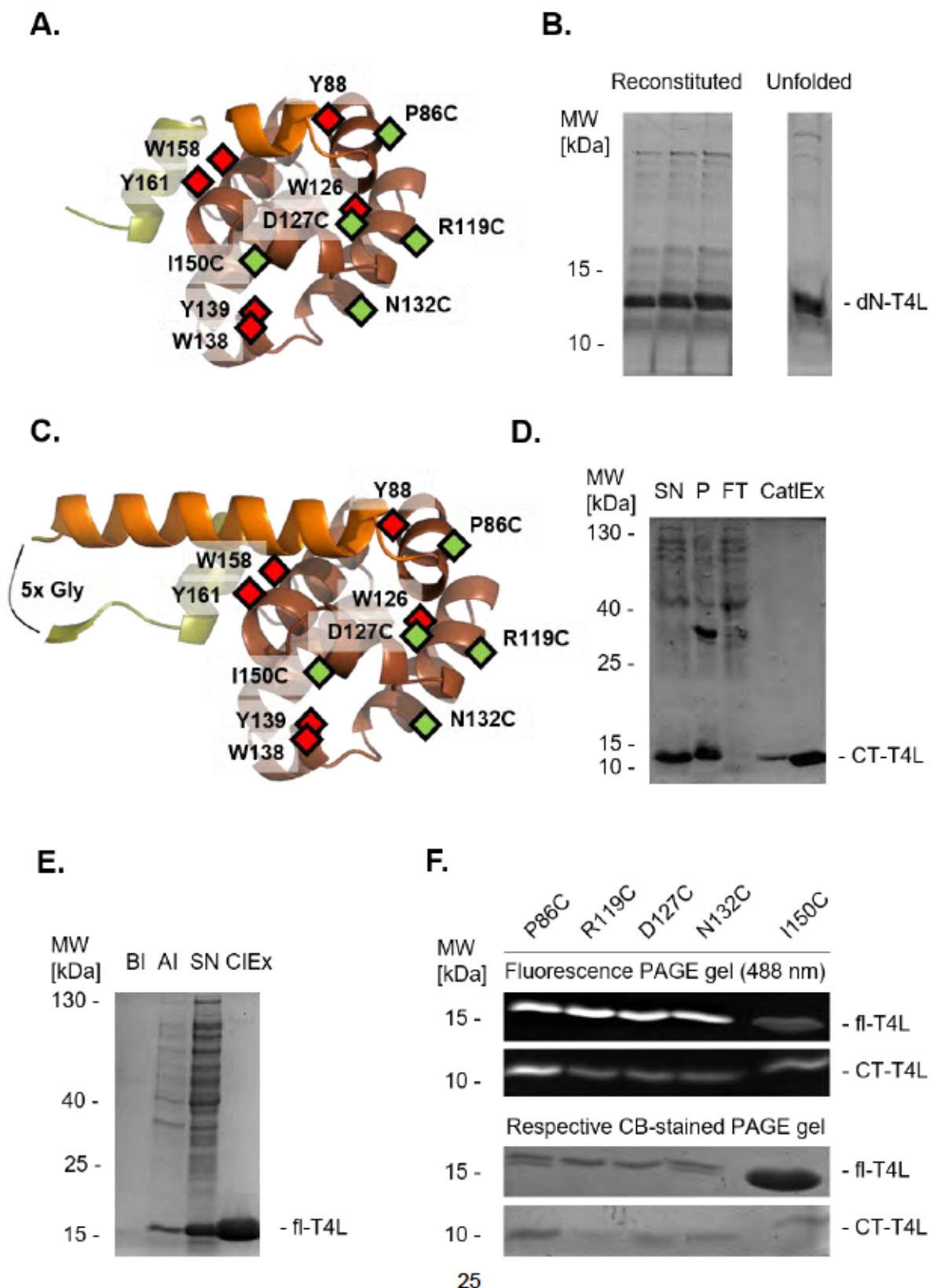
Figure 2: Structure and purification of truncated T4L variants. A. Simplified crystal structure of dN-T4L based on and modified from PDB: 148L. 2D-depiction of quencher in red and labeling positions in green. **B.** Purification SDS-PAGE gel of dN-T4L P86C with (unfolded) and without (reconstituted) GdmCl. GdmCl in concentrations >1 M smeared the gel band. **C.** Simplified crystal structure of CT-T4L based on and modified from PDB: 148L. The first 12 amino acids are linked via a 5xGly region to the full-length linker helix+1 (not displayed). **D.** Purification gel of CT-T4L showing supernatant (SN) and pellet (P) fractions before cation exchange (CatIEx), flow-through (FT) is gathered after applying SN to the CatIEx-column. **E.** Exemplary purification gel of full-length T4L N132C variant. BI: Before induction; AI: After induction; SN: Supernatant (before CatIEx); CIE: Cation exchange chromatography (main fraction), clean lane. **F.** All variants used were checked for integrity via fluorescence at approx. 488 nm excitation and 515 nm emission and coomassie blue (CB).

The concurrent design of CT-T4L (see **Fig. 1B** for sequence, **Fig. 2C** for structure) was based on Kato et al. [37]. In contrast to the dN-T4L fragment (**Fig. 2A**), the CT-T4L was soluble after expression and the variants P86C, R119C, D127C, N132C and I150C (based on WT*) were eventually purified using the standard full-length T4L protocol [18] (**Fig. 2D**). The full-length T4L variants were re-expressed and purified according to the same protocol (**Fig. 2E**). After labeling with Alexa 488-maleimide, all variants were checked for degradation (**Fig. 2F**) before performing FCS measurements. All variants except the full-length I150C variant show a good correlation in band density between fluorescence gel and CB stain. The full-length T4L I150C shows a much higher density in CB staining, underlining the limited degree of labeling (DOL) found for the I150C variants. The DOL for CT-T4L I150C was improved by a massive excess of dye, which proved disadvantageous for FCS measurements due to residual free dye which could not be removed by desalting/buffer exchange columns, but needed a size exclusion chromatography treatment on a larger column (e.g. Superdex 10/300).

FCS curves (**Fig. 3**) were fitted with three fit models to allow an unbiased, naïve approach on how many terms are needed to describe the system. We fitted either one (1-BT, orange), two (2-BT, brown) or three (3-BT, olive) (see **5. Methods**) bunching terms to account for all possible photophysical processes within the system. The respective fit with highest goodness of fit was displayed over data. Additionally, CT-T4L I150C was fitted with two diffusion/one bunching (2-diff., 1-BT, bright green) and two diffusion/two bunching (2-diff., 2-BT, bright blue) (see **5. Methods**) to account for possible dye diffusion. Weighted residuals are displayed on top. Confocal volume was assumed as three-dimensional Gauss-shaped, determined by dye reference measurements and kept constant for each respective set. Diffusion time of the free Alex Fluor 488 dye in our confocal volume was determined with ~260 µs, depending on the fit used (see **Chapter 1-Table supplement 1-4**). For free Alexa Fluor 488 (**Fig. 3F**) and both FL- and CT-P86C (**Fig. 3A**) variants, almost no static quenching

was present. One bunching term (BT) was therefore sufficient to describe the data, since triplet quenching (C-term) was the only contribution for this variant and for the free dye. Note that slight aggregation was visible for both FL- and CT-P86C variants at approx. 0.5-10 ms. For the R119C (**Fig. 3B**), D127C (**Fig. 3C**) and N132C (**Fig. 3D**) variants as well as fl-T4L I150C (**Fig. 3E**), two BT were needed to describe the data, both for triplet quenching (C-term) and static quenching (B-term). Only the CT-T4L 150C (**Fig. 3E/F**) variant needed a third bunching term (A-term) or alternatively a second diffusion term to achieve the best goodness of fit. The diffusion component is able to replace the slow quenching (A-term) found in the CT-I150C at 260 µs, yielding the same goodness of fit for two diffusion/two bunching (**Fig. 3F**) or one diffusion/three bunching (**Fig. 3E**). In general, the amount of static quenching found for each variant is concurrent to earlier measurements [32] and does not majorly differ between full-length and truncated variants see (**Chapter 1-Table supplement 1-4**). There are minor changes in triplet time and amplitude as well as static quenching between the FL- and CT-T4L variants which will be discussed accordingly. To resolve the unknown quenching attributed to internal dynamics originally found at 260 µs, a state at 260 µs (t_A) was forced in the 3-BT fit for all variants (as specified in **Table 1**). This made it possible to observe which data has an increased amplitude for the quenching described by A (data cells highlighted by thick line in **Table 1**). Only the CT-T4L I150C has a considerable amplitude (0.089) for the forced 260 µs A-term.
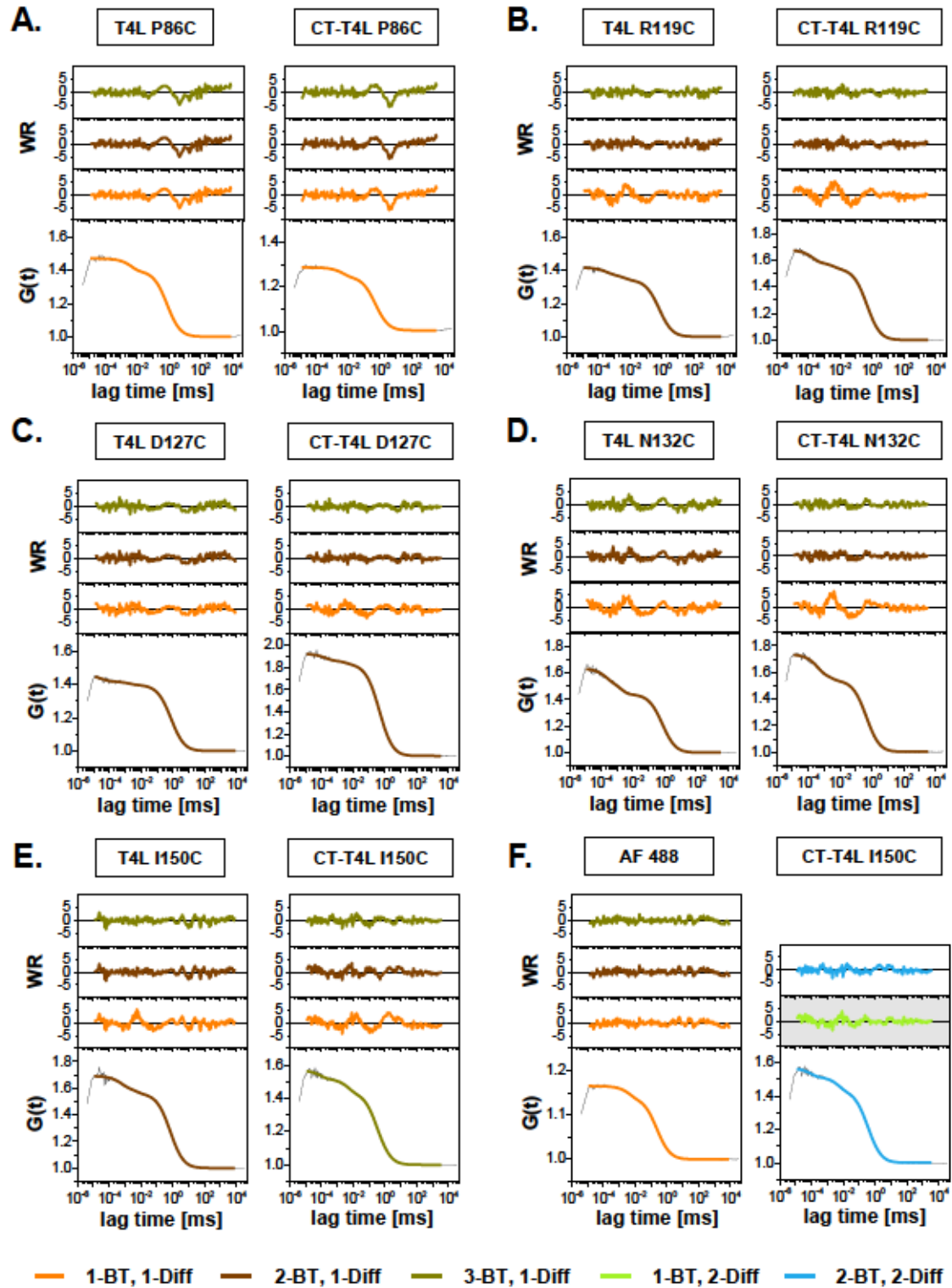
Figure 3: FCS data fitted with different fit functions (BT = bunching term, Diff = diffusion term).
A. T4L P86C and CT-T4L P86C, best chi^2_{red} if fitted with one bunching term (1-BT) (triplet). Adding a third state at 260 µs (3-BT) does not improve goodness of fit. **B.** T4L R119C and CT-T4L R119C, best chi^2_{red} if fitted with two bunching terms (2-BT) (one triplet, one fast quenching). Again, adding a third state at 260 µs (3-BT) does not improve goodness of fit. **C.** T4L D127C and CT-T4L D127C, best chi^2_{red} if fitted with two bunching terms (2-BT) (one triplet, one static quenching).

Adding a third state at 260 µs (3-BT) does not improve goodness of fit. **D.** T4L N132C and CT-T4L N132C, best chi²$_{red}$ if fitted with two bunching terms (2-BT) (one triplet, one static quenching). Again, adding a third state at 260 µs (3-BT) does not improve goodness of fit. **E.** T4L I150, best chi²$_{red}$ if fitted with two bunching terms (2-BT) (one triplet, one static quenching). Adding a third state at 260 µs (3-BT) does not improve goodness of fit. CT-T4L I150, best chi²$_{red}$ if fitted with three bunching terms (one triplet, one static quenching, one slow quenching state at 260 µs). **F.** Alexa Fluor 488 (Carboxylic acid), best chi²$_{red}$ if fitted with one bunching term (triplet) Adding a second and third state does not improve goodness of fit any further. CT-T4L I150, fitted with two diffusing components and one or two bunching terms (one triplet, one static quenching).

Table 1: FCS fit-parameter. BT = Bunching term. Fit results for Fit 28 (**see 5. Methods,** *Eq. (3)*) with the additional bunching-term fixed to 260 µs, forcing the proposed domain dynamic. Fixed parameters in italic. See **5. Methods,** *Eq. (3)* for explanation of fit parameters. See **Chapter 1-Table supplement 3** for non-rounded numbers.

		Full-length T4L					CT-T4L				
Species	A488	P86C	R119C	D127C	N132C	I150C	P86C	R119C	D127C	N132C	I150C
chi²$_{red}$:	0.736	2.512	1.064	1.280	1.828	1.138	2.725	0.945	0.834	0.919	0.950
Offset (DC):	1.000	1.000	1.000	1.000	1.001	1.000	1.003	1.001	1.005	1.003	1.000
Number of molecules (N):	6.056	2.122	2.408	2.258	1.541	1.447	3.515	1.477	1.075	1.363	1.756
Diffusion time (t_{trans}) [ms]:	0.299	0.801	0.694	0.742	0.728	0.719	0.513	0.501	0.472	0.471	0.424
Obs. Vol. ($\omega0/z0$) [fL]:	*3.690*	*3.690*	*3.690*	*3.690*	*3.690*	*3.690*	*3.690*	*3.690*	*3.690*	*3.690*	*3.450*
Ampl. of C-BT (triplet) (C):	0.147	0.151	0.081	0.031	0.198	0.076	0.132	0.069	0.035	0.089	0.136
Time of C-BT (t_C) [µs]:	4.420	4.421	4.232	2.553	1.451	4.500	4.787	3.636	3.123	2.270	4.950
Ampl. of B-BT (stat. quenching) (B):	-	0.003	0.091	0.068	0.121	0.109	0.006	0.133	0.068	0.174	0.089
Time of B-BT (t_B) [ns]:	-	137.2	342.0	95.02	94.26	541.5	703.6	184.0	135.9	339.1	88.65
Ampl. of A-BT (additional fast/slow) (A):	-	0.000	0.003	0.000	0.001	0.000	0.000	0.000	0.002	0.005	0.089
Ampl. of A-BT (t_A):	-	*0.260*	*0.260*	*0.260*	*0.260*	*0.260*	*0.260*	*0.260*	*0.260*	*0.260*	*0.260*

3. Discussion

Expression and purification of the truncated T4L variants

During protein purification, the formation of inclusion bodies can be triggered by high protein concentration levels as well as destabilizing the protein due to high expression temperature, lack of post-translational modifications (PTM) and loss of stability due to mutations [35, 38, 39]. Since the full-length T4L is a highly soluble protein [40] and our full-length variants never formed visible amount of inclusion bodies, we attributed the formation of inclusion bodies to destabilization caused by the missing N-terminal domain, the position of the N-terminal alpha helix (res. 1-12) switched to the C-terminus and the severely shortened linker helix. The stabilization issue of the dN-T4L construct was already noted within the original publication [33]. Although appearing to be well folded it was determined to be approx. 6.5x less stable compared to full-length WT* (2.1 kcal/mol vs 14.1 kcal/mol) due to hydrophobic parts being exposed to solvent [33]. Still, we managed to achieve good expression levels of the dN-T4L construct after testing different post-induction conditions, settling on lower temperature over night to minimize instability due to higher temperature. The purity achieved from our expression was decent using only molecular weight cut-off filter (MWCO) for a crude purification since the high amount of GdmCl prohibited the use of ion-exchange columns. The addition of GdmCl smeared the SDS-PAGE and masked UV-based concentration determination. It was more practical to work with reconstituted protein, although labeling with dyes is usually more efficient with unfolded proteins. A challenge was the current approach of refolding the protein correctly for our sensitive spectroscopic techniques. Depending on the shape of the folding energy landscape, it is possible that a considerable fraction of the protein gets stuck in one or multiple partially-refolded states [41, 42]. If this fraction exceeds our sensitive detection limit for sub-ensemble states, especially in SM-FRET experiments, partially refolded species might be misinterpreted as an actual state of the fully refolded T4L species. In contrast to unfolding experiments [1, 4, 37] which can usually be forced to an endpoint by maximizing chaotropic agent concentration and/or high temperatures [43], re-folding is more indifferent, hard to control and needs elaborate optimization [44, 45]. Since we were already aiming to characterize an unassigned and potentially elusive state, using a re-folding approach was risky. A robust control method like circular dichroism (CD) would likely not match the sensitivity of our spectroscopic methods to control our re-folding efforts. Consequentially, we decided to increase stability on a molecular level, avoiding inclusion bodies from the beginning and established the new construct CT-T4L [37] with an elongated linker helix and a more native position of the N-terminal alpha helix (res. 1-12). This scaffold represents an excellent tradeoff between low interdomain dynamics due to the perpetual absence of the NTsD and enough stability to effectively render the unfolding/re-folding procedure obsolete.

Fluorescence correlation spectroscopy

After purification and labeling, subsequent fitting of our FCS-data could confirm the results found by Stephan, L. (Thesis) [32] with FL-P86C showing the lowest amount of quenching and FL-N132C showing the highest amount of quenching. In the above-mentioned work, this finding was structurally explained and supported by MD simulations. It depends on dye position in respect to the quenchers and the tendency of charged residues to pull dyes towards quenchers. Our FCS measurements revealed a high similarity between the parameters of the CT and full-length variants for the same labeling positions. This finding indicates that the contribution to dye quenching originating from the NTsD is very low, as can be seen in **Fig. 1**. All tryptophanes and most of the tyrosine quenchers are positioned on the CTsD. In MD-simulations in Stephan, L. (Thesis) [32], three residues are identified as potential quenchers, W126, Y138 and M120. As all of them are positioned on the CTsD, we have already a good support for our findings. The minor differences in static quenching (as e.g. observed between FL- and CT-N132C with same amplitude of B but different t_B, see **Table 1**) could be due to small structural differences in the overall structure between FL- and CT-T4L caused by the lack of the NTsD. In general, the CTsD might be less stabilized in a sense of rigidity. In order to properly resolve the internal dynamics between CTsD and NTsD, we would likely need to design constructs of dyes/quenchers over both domains in order to display the domain movement.

The slow quenching (~260 µs) found in Stephan, L. (Thesis) [32] was not reproducible in both full-length and CT-T4L variants, even if a third quenching time at t_A = 260 µs was forced (3-BT fit). Almost all variants yielded sufficient goodness of fit when using a fit with two bunching terms, except for P86C and CT-T4L I150C. Both FL- and CT-P86C could be fitted with one bunching term since the amount of dye quenching by tryptophan is low enough. Additionally, both variants showed fluctuations in $G(t_c)$ at approx. 0.5-10 ms which is very likely due to aggregation of this variants, possibly due to the P86 position being an important structural element [46]. In the case of CT-T4L I150C, the labeling was challenging (see **Fig. 2F**). The full-length I150C already had a very low degree of labeling (DOL) und native conditions, yielding only a few percent labeled protein. The CT-I150C appeared to be even less accessible, prompting us to use a high dye excess to increase the DOL. Intriguingly, the CT-I150C was the only variant with a considerable amplitude when we forced the third time at t_A = 260 µs, the amplitude being at approx. 0.09 (see **Fig. 3E, Table 1**). This prompted us to evaluate the effect of a low DOL on our experiments. Low DOL would cause us to use a high excess of dye >10-fold. The present of free dye in our experiment would have an enormous impact on the FCS readout. The diffusion time of free Alexa Fluor 488 in our setup was determined with 235 µs, unluckily matching the transition time of the C2-C3 state

transition quite well. Earlier measurements only used a MWCO-column protocol to concentrate the labeled protein followed by additional washing steps to get rid of the excess dye (see e.g. Stephan, L. (Thesis) [32]). For all our measurements shown in this work, we purified the T4L variants with a size exclusion chromatography (SEC) column to minimize the amount of free dye. This resulted in an almost complete lack of the additional state in all variants except the CT-T4L I150C where we tried to enhance the DOL by working with a massive excess of dye. The CT-T4L I150C FCS measurement is the only dataset fitted well with a fit with two bunching terms and two diffusion terms, one for the free dye and one for the labeled protein (see **Fig. 3F**). We compared measurements of stocked MWCO-purified variants vs. SEC-purified variants and could reproduce the effect (data not shown). The SEC-purification step after labeling was thus to be highly important for all applications with high sensitivity to the presence of free dye and whenever large amounts of dye are used to compensate a low DOL.

4. Conclusion and outlook

Overall, we were able to reproduce the results found in earlier measurements of this system. Comparing the FL- and CT-T4L variants showed only minor differences, since the quenching is almost exclusively occuring within the CTsD and not between the domains

The unassigned state was shown to coincide with the presence of free dye, confusingly having by chance a similar diffusion time as the transition time of the T4L C2-C3 transition.

We conclude, that more optimized constructs (having quenchers/dyes on both domains, designed to detect inter-domain dynamics) will be suited to resolve the transitions into the elusive C3-state in a more explicit way.

5. Methods

Experimental Procedure:

dN-T4L, CT-T4L and fl-T4L were obtained readily cloned into pET-11a (BioCat, Heidelberg, Germany). Expression was done in BL 21 AI (F-ompT hsdS$_B$ (r$_B$- m$_B$-) gal dcm araB:T7RNAP-tetA) E. Coli bacteria (Thermo Fisher Scientific, Waltham, MA, USA). Chemically competent BL21 AI were transformed via heat-shock transformation with the respective plasmids, plated on LB-Agar plates with 50µg/ml kanamycin. After single colony picking and pre-culture, 1.5 l LB-medium was inoculated and grown at 37 °C until an optical density at 600 nm (OD$_{600}$) of 0.8. Expression was induced with a final concentration of 1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) and 4 g/l L-arabinose, bacteria were left expressing for 16 h at 30 °C. After expression, bacteria were harvested by centrifugation (10 min at 8,000 xg), resuspended in lysis buffer (10 mM HEPES, pH 8, 0.1 mM EDTA, 20 mM NaCl), disintegrated by sonication, incubated with 1 µg/ml DNase I for 1 h at 4 °C and segmented

by centrifugation (30 min at 24,000 xg, 4 °C). For purification from inclusion bodies (dN-T4L), the pellet was resuspended in solubilization buffer (50 mM Tris-HCl, pH 8, 1 mM EDTA, 1% (v/v) Triton X-100), sonicated and centrifuged (30 min at 24,000 xg, 4 °C). The supernatant was discarded and the step was repeated. Subsequently, the remaining pellet was resuspended in denaturing buffer (200 mM Tris-HCl, pH 7.5, 6 M GdmCl) and left stirring for 16 h at 4 °C. The suspension was again centrifuged (30 min at 24,000 xg, 4 °C) and the protein was crudely purified by using MWCO-columns (Merck, Darmstadt, Germany) with the respective cut-offs. For purification from supernatant (fl-T4L, CT-T4L), the supernatant was filtered and loaded on a Resource S cation exchange column (Cytiva, Marlborough, MA, USA). The protein was then eluted in an IEx-buffer (25 mM HEPES, pH 8, 2 mM DTT) with an increasing NaCl-concentration from 0-1 M NaCl over 100 ml. For labeling, the protein-buffer was exchanged to PBS (11.8 mM P_i, pH 7.4, 137 mM NaCl, 2.7 mM KCl) in NAP-5 columns (Cytiva), incubated with a 10x excess TCEP for 30 min at 4 °C and then incubated with a 2x excess of the maleimide-conjugated Alexa Fluor 488 dye (Thermo Fisher Sci.) for an empirically determined time, 15 min for all variants except I150C and 16 h for I150C. The dye/protein-mixture was then loaded on a SEC-column and eluted with PBS-buffer. Fraction purity and DOL was evaluated via UV/VIS-absorption and PAGE. FCS measurements were done in spectroscopically clean PBS buffer. To equalize triplet quenching, the buffer was equilibrated at air for 1 h to allow a homogenous oxygen concentration in the buffer. 1 mM BSA was added to the buffer to avoid protein sticking to the glass slide. Laser power at the objective was adjusted to 60 μW.

Experimental Setup:

Confocal setup for FCS was illuminated by a continuous wave diode laser 488 06-MLD (Cobolt, Solna, Sweden) through a water-immersion objective lens UPLSAPO 60x (Olympus, Hamburg, Germany) with a numerical aperture of 1.2. Detection side contained two avalanche photodiodes PDM 50CT (PicoQuant, Berlin, Germany) coupled to a photon counting module SPCM-AQR-14 (PerkinElmer, Waltham, MA, USA). A dichroic beamsplitter 488 DCLP (AHF, Tübingen, Germany) was used in the excitation path and a 520/30 bandpass filter (AHF) in the emission path.

Data analysis:

FCS data was analyzed with the software correlator and fitting routine KRISTINE, part of the MFD-software package available from our research group. The following five fits were used with fitting parameters iterating towards a minimal chi-square according to the Levenberg-Marquardt algorithm.

$$G(t_c) = DC + \frac{1}{N} \cdot \frac{1}{1+\frac{t_c}{|t_{trans}|}} \cdot \frac{1}{\left(1+\frac{t_c}{\left(\frac{\omega_0}{z_0}\right)^2 \cdot |t_{trans}|}\right)^{\frac{1}{2}}} \cdot \left(1 - |C| + |C| \cdot e^{-\frac{t_c}{|t_C|}}\right) \tag{1}$$

with DC = offset, N = total number of fluorophore emitters independently diffusing in the observation volume, t_{trans} = diffusion time, ω_0/z_0 = observation volume approximated by a 3D-Gaussian volume with $1/e^2$ radii in the lateral (ω_0) and axial direction (z_0), C = amplitude of triplet states, t_C = triplet transition time. Eq. 1 (Fit 24 in KRISTINE) represents the most appropriate FCS fit for any freely diffusing standard dye, with one bunching term (1-BT) being present for triplet-quenching.

$$G(t_c) = DC + \frac{1}{N} \cdot \frac{1}{1+\frac{t_c}{|t_{trans}|}} \cdot \frac{1}{\left(1+\frac{t_c}{\left(\frac{\omega_0}{z_0}\right)^2 \cdot |t_{trans}|}\right)^{\frac{1}{2}}} \cdot \left(1 - |C| + |C| \cdot e^{-\frac{t_c}{|t_C|}} - |B| + |B| \cdot e^{-\frac{t_c}{|t_B|}}\right) \tag{2}$$

with DC = offset, N = total number of fluorophore emitters independently diffusing in the observation volume, t_{trans} = diffusion time, ω_0/z_0 = observation volume approximated by a 3D-Gaussian volume with $1/e^2$ radii in the lateral (ω_0) and axial direction (z_0), C = amplitude of triplet states, t_C = triplet transition time, B = amplitude of static quenching, t_B = static quenching time. Eq. 2 (Fit 26 in KRISTINE) represents the most appropriate FCS fit for any fluorescently labeled protein where the dye is affected by static quenching. This is considered by a second bunching term (B-term) in addition to the first bunching term being present for triplet-quenching (C-term).

$$G(t_c) = DC + \frac{1}{N} \cdot \frac{1}{1+\frac{t_c}{|t_{trans}|}} \cdot \frac{1}{\left(1+\frac{t_c}{\left(\frac{\omega_0}{z_0}\right)^2 \cdot |t_{trans}|}\right)^{\frac{1}{2}}} \cdot \tag{3}$$

$$\left(1 - |C| + |C| \cdot e^{-\frac{t_c}{|t_C|}} - |B| + |B| \cdot e^{-\frac{t_c}{|t_B|}} - |A| + |A| \cdot e^{-\frac{t_c}{|t_A|}}\right)$$

with DC = offset, N = total number of fluorophore emitters independently diffusing in the observation volume, t_{trans} = diffusion time, ω_0/z_0 = observation volume approximated by a 3D-Gaussian volume with $1/e^2$ radii in the lateral (ω_0) and axial direction (z_0), C = amplitude of triplet states, t_C = triplet transition time, B = amplitude of static quenching, t_B = static quenching time, A = additional bunching term used to either fit fast processes (photon

antibunching, PET) or slower processes (e.g. relaxation time of internal dynamics), t_A = corresponding time. Eq. 3 (Fit 28 in KRISTINE) represents a more complex FCS fit for any fluorescently labeled protein where the dye is affected by static quenching and additional quenching on a separate timescale (faster or slower). This is considered by a third bunching term (A-term) in addition to the two bunching terms (C-term, B-term) being present for triplet-quenching and static quenching.

$$f(\tau) = DC + \frac{1}{N} \cdot \left(\frac{|R|}{1+\frac{t_c}{|t_{trans;1}|}} \cdot \frac{1}{\left(1+\frac{t_c}{\left(\frac{\omega_{0;1}}{z_{0;1}}\right)^2 \cdot |t_{trans;1}|}\right)^{\frac{1}{2}}} + |1-|R|| \frac{1}{1+\frac{t_c}{|t_{trans;2}|}} \cdot \frac{1}{\left(1+\frac{t_c}{\left(\frac{\omega_{0;2}}{z_{0;2}}\right)^2 \cdot |t_{trans;2}|}\right)^{\frac{1}{2}}} \right) \cdot$$

$$\left(1 - |C| + |C| \cdot e^{\frac{t_c}{|t_c|}} \right) \tag{4}$$

with DC = offset, N = total number of fluorophore emitters independently diffusing in the observation volume, t_{trans} = diffusion times, ω_0/z_0 = observation volumes approximated by a 3D-Gaussian volume with $1/e^2$ radii in the lateral (ω_0) and axial direction (z_0), C = amplitude of triplet states, t_C = triplet transition time, R = fraction of the first diffusion term, denoted with '1'. Eq. 4 (Fit 33 in KRISTINE) represents a more complex FCS fit for a mixture of two differently diffusing fluorophore emitters affected by triplet-quenching (C-term).

$$f(\tau) = DC + \frac{1}{N} \cdot \left(\frac{|R|}{1+\frac{t_c}{|t_{trans;1}|}} \cdot \frac{1}{\left(1+\frac{t_c}{\left(\frac{\omega_{0;1}}{z_{0;1}}\right)^2 \cdot |t_{trans;1}|}\right)^{\frac{1}{2}}} + |1-|R|| \frac{1}{1+\frac{t_c}{|t_{trans;2}|}} \cdot \frac{1}{\left(1+\frac{t_c}{\left(\frac{\omega_{0;2}}{z_{0;2}}\right)^2 \cdot |t_{trans;2}|}\right)^{\frac{1}{2}}} \right) \cdot$$

$$\left(1 - |C| + |C| \cdot e^{\frac{t_c}{|t_c|}} - |B| + |B| \cdot e^{\frac{t_c}{|t_B|}} \right) \tag{5}$$

with DC = offset, N = total number of fluorophore emitters independently diffusing in the observation volume, t_{trans} = diffusion times, ω_0/z_0 = observation volumes approximated by a 3D-Gaussian volume with $1/e^2$ radii in the lateral (ω_0) and axial direction (z_0), C = amplitude of triplet states, t_C = triplet transition time, R = fraction of the first diffusion term, denoted with '1'. Eq. 5 (Fit 34 in KRISTINE) represents a more complex FCS fit for a mixture of two differently diffusing fluorophore emitters affected by triplet-quenching (C-term) and static quenching (B-term).

6. Chapter 1 References

1.   Baase, W. A., Liu, L. J., Tronrud, D. E., Matthews, B. W. Lessons from the lysozyme of phage T4. *Protein Science*, 2010, **19**, 631-641. DOI: 10.1002/pro.344.

2.   Becktel, W. J., Baase, W. A. Thermal denaturation of bacteriophage-T4 lysozyme at neutral pH. *Biopolymers*, 1987, **26**, 619-623. DOI: 10.1002/bip.360260505.

3.   Matthews, B. W. Studies on protein stability with T4 lysozyme. *Advances in Protein Chemistry*, 1995, **46**, 249-278. DOI: 10.1016/s0065-3233(08)60337-x.

4.   Llinas, M., Gillespie, B., Dahlquist, F. W., Marqusee, S. The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nature Structural Biology*, 1999, **6**, 1072-1078. DOI: 10.1038/14956.

5.   Eriksson, A. E., Baase, W. A., Matthews, B. W. Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4-lysozyme have different structural and thermodynamic consequences. *Journal of Molecular Biology*, 1993, **229**, 747-769. DOI: 10.1006/jmbi.1993.1077.

6.   Deng, Y. Q., Roux, B. Calculation of standard binding free energies: aromatic molecules in the T4 lysozyme L99A mutant. *Journal of Chemical Theory and Computation*, 2006, **2**, 1255-1273. DOI: 10.1021/ct060037v.

7.   Mobley, D. L., Graves, A. P., Chodera, J. D., McReynolds, A. C., Shoichet, B. K., Dill, K. A. Predicting absolute ligand binding free energies to a simple model site. *Journal of Molecular Biology*, 2007, **371**, 1118-1134. DOI: 10.1016/j.jmb.2007.06.002.

8.   Morton, A., Matthews, B. W. Specificity of ligand-binding in a buried nonpolar cavity of T4 lysozyme - linkage of dynamics and structural plasticity. *Biochemistry*, 1995, **34**, 8576-8588. DOI: 10.1021/bi00027a007.

9.   Chen, Y., Hu, D. H., Vorpagel, E. R., Lu, H. P. Probing single-molecule T4 lysozyme conformational dynamics by intramolecular fluorescence energy transfer. *Journal of Physical Chemistry B*, 2003, **107**, 7947-7956. DOI: 10.1021/jp022406z.

10.   Yirdaw, R. B., McHaourab, H. S. Direct observation of T4 lysozyme hinge-bending motion by fluorescence correlation spectroscopy. *Biophysical Journal*, 2012, **103**, 1525-1536. DOI: 10.1016/j.bpj.2012.07.053.

11.   VanGilst, M., Hudson, B. S. Histidine-tryptophan interactions in T4 lysozyme: 'anomalous' pH dependence of fluorescence. *Biophysical Chemistry*, 1996, **63**, 17-25. DOI: 10.1016/s0301-4622(96)02181-3.

12.   Hu, D. H., Lu, H. P. Placing single-molecule T4 lysozyme enzymes on a bacterial cell surface: toward probing single-molecule enzymatic reaction in living cells. *Biophysical Journal*, 2004, **87**, 656-661. DOI: 10.1529/biophysj.104.042101.

13. Matthews, B. W., Remington, S. J., Grutter, M. G., Anderson, W. F. Relation between hen-egg white lysozyme and bacteriophage-T4 lysozyme - evolutionary implications. *Journal of Molecular Biology*, 1981, **147**, 545-558. DOI: 10.1016/0022-2836(81)90399-5.

14. Demchick, P., Koch, A. L. The permeability of the wall fabric of Escherichia coli and Bacillus subtilis. *Journal of Bacteriology*, 1996, **178**, 768-773. DOI: 10.1128/jb.178.3.768-773.1996.

15. Boyer, P. D. The enzymes. 3rd ed. 1971: Academic Press

16. Kuroki, R., Weaver, L. H., Matthews, B. W. A covalent enzyme-substrate intemediate with saccharide distortion in a mutant T4 lysozyme. *Science*, 1993, **262**, 2030-2033. DOI: 10.1126/science.8266098.

17. Goto, N. K., Skrynnikov, N. R., Dahlquist, F. W., Kay, L. E. What is the average conformation of bacteriophage T4 lysozyme in solution? A domain orientation study using dipolar couplings measured by solution NMR. *Journal of Molecular Biology*, 2001, **308**, 745-764. DOI: 10.1006/jmbi.2001.4614.

18. Sanabria, H., Rodnin, D., Hemmen, K., Peulen, T. O., Felekyan, S., Fleissner, M. R., et al. Resolving dynamics and function of transient states in single enzyme molecules. *Nature Communications*, 2020, **11**. DOI: 10.1038/s41467-020-14886-w.

19. McHaourab, H. S., Oh, K. J., Fang, C. J., Hubbell, W. L. Conformation of T4 lysozyme in solution. Hinge-bending motion and the substrate-induced conformational transition studied by site-directed spin labeling. *Biochemistry*, 1997, **36**, 307-316. DOI: 10.1021/bi962114m.

20. Zhang, X. J., Wozniak, J. A., Matthews, B. W. Protein flexibility and adabptability seen in 25 crystal forms of T4 lysozyme. *Journal of Molecular Biology*, 1995, **250**, 527-552. DOI: 10.1006/jmbi.1995.0396.

21. Mansoor, S. E., McHaourab, H. S., Farrens, D. L. Mapping proximity within proteins using fluorescence spectroscopy. A study of T4 lysozyme showing that tryptophan residues quench bimane fluorescence. *Biochemistry*, 2002, **41**, 2475-2484. DOI: 10.1021/bi011198i.

22. Van Gilst, M., Tang, C., Roth, A., Hudson, B. S. Quenching interactions and nonexponential decay: tryptophan 138 of bacteriophage T4 lysozyme. *Journal of Fluorescence*, 1994, **4**, 203-207. DOI: 10.1007/BF01878452.

23. Lakowicz, J. R., Maliwal, B. P., Cherek, H., Balter, A. Rotational freedom of tryptophan residues in proteins and peptides. *Biochemistry*, 1983, **22**, 1741-1752. DOI: 10.1021/bi00277a001.

24. Harris, D. L., Hudson, B. S. Photophysics of tryptophan in bacteriophage-T4 lysozymes. *Biochemistry*, 1990, **29**, 5276-5285. DOI: 10.1021/bi00474a009.

25. Luitz, M. P., Barth, A., Crevenna, A. H., Bomblies, R., Lamb, D. C., Zacharias, M. Covalent dye attachment influences the dynamics and conformational properties of flexible peptides. *Plos One*, 2017, **12**. DOI: 10.1371/journal.pone.0177139.

26. Sharma, A., Enderlein, J., Kumbhakar, M. Photon antibunching reveals static and dynamic quenching interaction of tryptophan with Atto-655. *Journal of Physical Chemistry Letters*, 2017, **8**, 5821-5826. DOI: 10.1021/acs.jpclett.7b02430.

27. Berezin, M. Y., Achilefu, S. Fluorescence lifetime measurements and biological imaging. *Chemical Reviews*, 2010, **110**, 2641-2684. DOI: 10.1021/cr900343z.

28. Knowles, K. E., Malicki, M., Weiss, E. A. Dual-time scale photoinduced electron transfer from PbS quantum dots to a molecular acceptor. *Journal of the American Chemical Society*, 2012, **134**, 12470-12473. DOI: 10.1021/ja3060222.

29. Nettels, D., Gopich, I. V., Hoffmann, A., Schuler, B. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, **104**, 2655-2660. DOI: 10.1073/pnas.0611093104.

30. Wöll, D. Fluorescence correlation spectroscopy studies of polymer systems. 1st ed. Springer Series on Fluorescence. Vol. 16. 2016: Springer.

31. Ha, T., Tinnefeld, P. Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. *Annual Review of Physical Chemistry*, 2012, **63**, 595-617. DOI: 10.1146/annurev-physchem-032210-103340.

32. Stephan, L., Influence of the protein environment on fluorophore properties, *Molecular Physical Chemistry*. 2017, HHU: Duesseldorf.

33. Llinas, M., Marqusee, S. Subdomain interactions as a determinant in the folding and stability of T4 lysozyme. *Protein Science*, 1998, **7**, 96-104. DOI: 10.1002/pro.5560070110.

34. Kane, J. F., Hartley, D. L. Formation of recombinant proteininclusion-bodies in Escheria coli. *Trends in Biotechnology*, 1988, **6**, 95-101. DOI: 10.1016/0167-7799(88)90065-0.

35. Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M., Panda, A. K. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microbial Cell Factories*, 2015, **14**. DOI: 10.1186/s12934-015-0222-8.

36. Yang, Z., Zhang, L. L., Zhang, Y., Zhang, T., Feng, Y. Y., Lu, X. X., et al. Highly efficient production of soluble proteins from insoluble inclusion bodies by a two-step-denaturing and refolding method. *Plos One*, 2011, **6**. DOI: 10.1371/journal.pone.0022981.

37. Kato, H., Feng, H. Q., Bai, Y. W. The folding pathway of T4 lysozyme: the high-resolution structure and folding of a hidden intermediate. *Journal of Molecular Biology*, 2007, **365**, 870-880. DOI: 10.1016/j.jmb.2006.10.047.

38. Thomas, J. G., Baneyx, F. Protein misfolding and inclusion body formation in recombinant Escherichia coli cells overexpressing heat-shock proteins. *Journal of Biological Chemistry*, 1996, **271**, 11141-11147. DOI: 10.1074/jbc.271.19.11141.

39. Tokmakov, A. A., Kurotani, A., Takagi, T., Toyama, M., Shirouzu, M., Fukami, Y., et al. Multiple post-translational modifications affect heterologous protein synthesis. *Journal of Biological Chemistry*, 2012, **287**, 27106-27116. DOI: 10.1074/jbc.M112.366351.

40. Thorsen, T. S., Matt, R., Weis, W. I., Kobilka, B. K. Modified T4 lysozyme fusion proteins facilitate G protein-coupled receptor crystallogenesis. *Structure*, 2014, **22**, 1657-1664. DOI: 10.1016/j.str.2014.08.022.

41. Chong, S. H., Im, H., Ham, S. Explicit characterization of the free energy landscape of pKID-KIX coupled folding and binding. *ACS Central Science*, 2019, **5**, 1342-1351. DOI: 10.1021/acscentsci.9b00200.

42. Bollen, Y. J. M., Kamphuis, M. B., van Mierlo, C. P. M. The folding energy landscape of apoflavodoxin is rugged: hydrogen exchange reveals nonproductive misfolded intermediates. *Proceedings of the National Academy of Sciences of the United States of America*, 2006, **103**, 4095-4100. DOI: 10.1073/pnas.0509133103.

43. Walters, J., Milam, S. L., Clark, A. C. Practical approaches to protein folding and assembly: spectroscopic statistics in thermodynamics and kinetics. *Methods in Enzymology*, 2009, **455**, 1-39. DOI: 10.1016/s0076-6879(08)04201-8.

44. Vallejo, L. F., Rinas, U. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microbial Cell Factories*, 2004, **3**. DOI: 10.1186/1475-2859-3-11.

45. Fersht, A. R., Daggett, V. Protein folding and unfolding at atomic resolution. *Cell*, 2002, **108**, 573-582. DOI: 10.1016/s0092-8674(02)00620-7.

46. Alber, T., Bell, J. A., Daopin, S., Nicholson, H., Wozniak, J. A., Cook, S., et al. Replacements of Pro86 in phage-T4 lysozyme extend an alpha-helix but do not alter protein stability. *Science*, 1988, **239**, 631-635. DOI: 10.1126/science.3277275.

Chapter 1 supplement

DATAFINDER

A) FCS data and fits are available at:
   :\User\KochJ\Thesis\T4L story\FCS and fits

B) Text and images used in this work are available at:
   :\User\KochJ\Thesis\T4L story

C) Original data used for this work available at:
   :\FCS\2021-1\February\T4L_AL,KO

Chapter 1-Table supplement 1

Fit 24 (1-BT), fixed parameter in *italic*

Species	A488	FL				
		P86C	**R119C**	**D127C**	**N132C**	**I150C**
chi²:	0.80968258	2.45518853	2.32571452	1.37375446	2.63801627	2.18614206
b0:	0.99991279	1.00026769	1.00017258	0.99966266	1.00064601	1.0005712
b1:	6.05554822	2.13129418	2.4945268	2.34827089	1.65820336	1.48044091
b2:	0.23490155	0.80203979	0.68310791	0.74174323	0.72518689	0.70977429
b3:	3.69090736	*3.69*	*3.69*	*3.69*	*3.69*	*3.69*
b4:	0.1449594	0.1503468	0.13968238	0.06275093	0.26601239	0.16218254
b5:	0.00434078	0.00447693	0.00197795	0.00102812	0.00105921	0.00167515
Species	-	CT				
		P86C	**R119C**	**D127C**	**N132C**	**I150C**
chi²:	-	2.66485705	3.59228098	1.60852387	3.44454503	3.00471144
b0:	-	1.00339897	1.00119386	1.00504648	1.00293483	1.00067106
b1:	-	3.52213781	1.5879945	1.11322571	1.40059084	1.89485154
b2:	-	0.5129526	0.49472511	0.46832095	0.46155212	0.35938823
b3:	-	*3.69*	*3.69*	*3.67*	*3.67*	*3.45*
b4:	-	0.13611994	0.13830162	0.0692765	0.23994681	0.16497693
b5:	-	0.00461285	0.00123217	0.00098715	0.00078713	0.00421087

Chapter 1-Table supplement 2

Fit 26 (2-BT), fixed parameter in *italic*

		FL				
Species	**A488**	**P86C**	**R119C**	**D127C**	**N132C**	**I150C**
chi²:	0.76850719	2.47678915	1.05698533	1.24472521	1.69195961	1.13383372
b0:	0.99990486	1.00026615	1.00009929	0.99964595	1.00057248	1.00043424
b1:	6.056904	2.12731179	2.40850172	2.18789112	1.57996802	1.44692808
b2:	0.24981167	0.80226287	0.6903474	0.74332719	0.73099757	0.7182335
b3:	3.636364	*3.69*	*3.69*	*3.69*	*3.69*	*3.69*
b4:	0.14669966	0.14979112	0.08134651	0.04492976	0.17246691	0.07427606
b5:	0.00443864	0.00449429	0.00427541	0.00161423	0.00172768	0.00447773
b6:	0.04483363	0.00220608	0.09107824	0.08249625	0.13018418	0.11038866
b7:	0.30080548	0.00020892	0.00034136	3.3558E-05	0.0001946	0.000557
		CT				
Species	-	**P86C**	**R119C**	**D127C**	**N132C**	**I150C**
chi²:	-	2.68349454	0.92265542	0.806551	0.89437239	1.41365901
b0:	-	1.0034018	1.00100992	1.00489724	1.00278121	1.00031527
b1:	-	3.52485101	1.47822437	1.08567149	1.36471238	1.88258321
b2:	-	0.51308872	0.5009438	0.4709931	0.46659155	0.37718976
b3:	-	*3.69*	*3.69*	*3.69*	*3.69*	*3.45*
b4:	-	0.13472979	0.06756951	0.0323251	0.08560032	0.1504262
b5:	-	0.00469564	0.00381218	0.00349726	0.00238205	0.00314615
b6:	-	0.00082465	0.13396957	0.06210195	0.17703541	0.03951995
b7:	-	0.00057931	0.0001904	0.00018656	0.00035163	0.04808755

Chapter 1-Table supplement 3

Fit 28 (3-BT), fixed parameter in *italic*

Species	A488	FL				
		P86C	**R119C**	**D127C**	**N132C**	**I150C**
chi²:	0.73570771	2.51196026	1.06429984	1.2804549	1.82778039	1.13788677
b0:	0.99988164	1.00014829	1.00007823	0.99968666	1.00059703	1.0004262
b1:	6.05590168	2.12221159	2.40848649	2.25763419	1.54136311	1.4471226
b2:	0.29946007	0.80121145	0.69421299	0.74214171	0.72842246	0.71866597
b3:	*3.69*	*3.69*	*3.69*	*3.69*	*3.69*	*3.69*
b4:	0.14673863	0.15056182	0.08106572	0.03138068	0.1981315	0.07618555
b5:	0.0044202	0.0044213	0.00423153	0.00255285	0.00145045	0.0045
b6:	0.06714207	0.00289181	0.09097636	0.06812928	0.1210975	0.10858549
b7:	0.12736747	0.00013716	0.00034196	9.5018E-05	9.4726E-05	0.00054147
b8:	0.12704334	1.8E-08	0.00314625	5.0538E-05	0.00062826	4.7601E-05
b9:	0.64011377	*0.26*	*0.26*	*0.26*	*0.26*	*0.26*
Species	-	CT				
		P86C	**R119C**	**D127C**	**N132C**	**I150C**
chi²:	-	2.72453492	0.94543721	0.83388013	0.91935866	0.94959722
b0:	-	1.0033975	1.00101506	1.00488295	1.00273632	1.0000864
b1:	-	3.51486171	1.47707419	1.07529803	1.36314755	1.75568955
b2:	-	0.51296536	0.50071574	0.47202232	0.47101953	0.4241694
b3:	-	*3.69*	*3.69*	*3.69*	*3.69*	*3.45*
b4:	-	0.13207987	0.06936562	0.03483247	0.08908338	0.13553533
b5:	-	0.00478676	0.00363596	0.00312283	0.00227	0.00495025
b6:	-	0.00607321	0.13265773	0.06802545	0.17401031	0.08942964
b7:	-	0.00070363	0.00018397	0.00013591	0.00033911	8.8652E-05
b8:	-	0.00037192	6.765E-06	0.0017359	0.00515348	0.08922678
b9:	-	*0.26*	*0.26*	*0.26*	*0.26*	*0.26*

Chapter 1-Table supplement 4

Fit 33 and 34 (1-BT/2-diff) (2-BT/2-diff), fixed parameter in _italic_

	Fit 33	Fit 33	Fit 34
Species	A488	CT I150C	CT I150C
chi²:	0.71849424	1.20533616	0.96203904
b0:	0.99991004	1.00006249	1.00033348
b1:	6.05661053	1.8862536	2.26049976
b2:	0.24892839	_0.26_	_0.26_
b3:	1.89883066	_3.45_	_3.45_
b4:	0.23908248	0.57449873	_0.45_
b5:	4.87597462	_3.45_	_3.45_
b6:	0.29671993	0.62719251	0.57482177
b7:	0.14612293	0.15761336	0.17055092
b8:	0.00441092	0.0035811	_0.005_
b9:	-	-	0.11797748
b10:	-	-	_8.8000E-05_

Chapter 2: Advanced fluorescence spectroscopy and stopped-flow form a perfect couple to elucidate dimer association dynamics and domain movements of the *human* guanylate binding protein hGBP1

Chapter digest

Aims	Methods	Outcome
Follow domain re-arrangement of dimerizing hGBP1 on a millisecond timescale after addition of GTPγS.	Ensemble FRET measurements on a stopped-flow coupled TCSPC.	Structural movement: Identification of the transitional timescale from monomer into consecutive dimer conformation.

Contributions

TCSPC/SF measurements: *Julian Koch, Thomas Peulen*. Data analysis/fitting: *Julian Koch, Oleg Opanasyuk, Julian Folz*. Mathematical models: *Oleg Opanasyuk*. TCPSC/SF setup implementation and testing: *Jan-Hendrik Budde*. hGBP1 expression and labeling: *Thomas Peulen, Carola Hengstenberg*. Project supervision: *Christian Herrmann, Claus A. M. Seidel*

Abstract

Human guanylate-binding protein 1 (hGBP1), a member of the dynamin-superfamily of large GTPases, can self-associate up to higher order oligomers in a nucleotide dependent manner, even forming ring-like and tubular structures. Therefore, hGBP1 is a prime example to study the relevance of structural dynamics as a prerequisite for the self-association of proteins. To follow the rearrangement of hGBP1 within microseconds to seconds after dimerization, we combined a time-correlated single photon counting (TCSPC) device for inter- and intramolecular Förster resonance energy transfer (FRET) experiments with a stopped-flow (SF) setup, using a network of FRET-pairs between the different domains of hGBP1. We followed FRET observables over time during the measurement, directly linking their change over time to domain movements and extracting rate constants. Under the given experimental conditions, the results obtained with our best fit indicate three major transitions, starting with a fast association of two hGBP1-monomers into an initial, 'flexible' dimer within a time-window of approx. t_0+360 ms. Next, after approx. t_0+830 ms, the hGBP1 dimer undergoes intermolecular re-arrangement of the C-terminal helix α12 and α13 into a 'bridged' dimer. Finally, we observe a slow ($t_0+6\text{-}8$ s) increase in distance between all intra-FRET pairs and a closing-in of the inter-FRET pair 344D + 481A, which is best explained by stretching out helix α12 and α13 into an 'elongated dimer'. Our findings prove the feasibility of the method and yield valuable information on understanding the transitional coordinate of hGBP1.

1. Introduction

The human guanylate binding protein hGBP1 is an interferon-gamma induced large GTPase and part of the immune response to intracellular parasites [1]. hGBP1 can dimerize and subsequently oligomerize to sustain cellular trafficking [3] and thus achieve its anti-parasitic functionality [4]. The dimer conformation needed for formation of a higher multimer is initiated by GTP-dependent engagement of the LG (Large GTPase) domains of two hGBP1 molecules (**Fig. 1A**, blue) [5], which arrange in a head-to-head conformation [6]. The dimer is now able to assume multiple conformations where the α12-helices (**Fig. 1A**, green) and the C-terminal α13-helices (**Fig. 1A**, yellow) are arranged in different positions around the core LG and middle domain [9], a multi-state we termed 'flexible dimer'. These internal re-arrangements happen in a microsecond regime, making them faster than the subsequent domain movements by orders of magnitude. From these initial fast-transitioning flexible dimer conformations, the α13-helices [7] can lock onto each other, pulling the α12-helices slightly away from the LG/middle into a planar arrangement [7,8] termed 'bridged dimer'. The LG-domain maintains permanent contact during these rearrangements. The bridged dimer represents the major species fraction of dimers in an equilibrium [10] before transitioning into elongated conformation and subsequent formation of oligomers happens under certain conditions. The (semi-)elongated dimer emerging from the bridged dimer is structurally elusive. Structures being half-kinked at the hinge around pos. 480 or a fully stretched-out form have been implied [20, 24]. Between the bridged dimer and the consecutive (semi-)elongated dimer seems to be a clear possibility for a transition state or transition ensemble. Here, the α12-helices would likely be in a slightly open or hinging move, increasing spatial distance in respect to each other and to the middle domain. The state could also represent a further advanced variation of the bridged dimer with locked α13-helices and the α12-helices being slightly decoupled [7]. Naturally, FRET-pairs used in our experiments were selected explicitly to show a change in FRET efficiency while transitioning between these states. Especially the residues pos. 525 and pos. 540 in relation to the others was of major interest since they best render the opening of the α12-helices away from the LG/middle domains.

The elaborate placement of fluorescent dyes at labeling sites into a FRET-network has been widely utilized for elucidation of more sophisticated protein structures [11] and even more so for dynamic movements [12], as it is exerted with hGBP1 in this work. The conformation and dynamics at the steady state monomer and the endpoint dimer of unfarnesylated hGBP1 was already comprehensibly resolved with FRET measurements [9], ultimately gaining a topological map of the different conformations possible, putting a frame around the start and the end of the transition coordinate. Since the transition times between the monomeric state and the endpoint dimer where unassessed, we aimed to resolve the domain rearrangement

and transition states on our macro-timescale, set within a millisecond time range. Although it is certainly possible to catch transient states down to a microsecond timescale [13, 14] and hence extract equilibrium constants [15] with single molecule FRET measurements, the proposed transient conformational states of hGBP1 are not easily eligible for the classical single molecule approaches. This limitation is mainly due to the necessity of adding an additional substrate (GTP or analogues) in order to trigger dimerization [7]. The fast dimerization reaction coordinate [16] makes it impossible to mix and start the experiment by hand and still detect transient states before the system enters an 'late' equilibrium, or respectively an endpoint state after a few seconds. Therefore, the combination of a detector unit for TCSPC, measuring ensemble FRET (eFRET), with a stopped-flow device (**Fig. 1B**) enables us to obtain the fluorescence decay points quickly after mixing hGBP1 with a GTP-analogue, GTPγS. When using eTCSPC to follow domain movements, a change in the relative distance of a FRET-pair in the sample due to internal dynamics will alter the energy-transfer rate k_{FRET} (see **6. Methods**) [17]. Consequentially, the fluorescence decays and the resulting FRET photons detected within the TCSPC 'microtime' (nanoseconds regime) will change over the duration of the whole measurement, the 'macrotime' (high milliseconds/seconds regime). With that type of data, it is possible to fit the signal fluctuation over the whole macrotime to derive time constants (**Fig. 1C**) and/or ananlyze the individual microtime-decays. The kinetic 2AB(B)C models used for macrotime fitting are based on an initial bimolecular assembly (state 2A), followed by compact, non-elongated dimers (state B) and are followed up by an elongated or semi-elongated dimer (state C) (**Fig. 1F**).

In theory, each recorded detection window could give an individually analyzable photon histogram per corresponding laser pulse, only limited by the nanosecond window the dye needs to decay from an excited to a steady state to be excitable again. In reality, dye photophysics, low concentration of labeled protein and detector capacity limit the number of photons collected per detection window [18, 19]. Consequentially, low amounts of photon-counts in the histogram would lead to poor statistics and a bad signal-to-noise ratio, resulting in unreliable fits. This forced us to bin all detection windows derived within 10 ms, a value which empirically provided enough photons per timepoint to produce a confident fit per time-bin. As a trade-off, time-resolution decreased. In the schematic examples shown in **Fig. 1D** and **Fig. 1E**, the inter-dye distance would increase for all intramolecular dye-pairs used in this experiment. Concurrently, the no-FRET fraction x_{noFRET} will increase over time while the FRET-fractions x_{FRET} will decrease. Based on available data [9] and knowledge of the endpoint equilibria, we theorized a working model as depicted in **Fig. 1F**, where we can follow the dimer transition over time. We then deployed **two kinetic models** to fit the data accordingly (see **2. Results** below).

A. Dye labelling positions

481A
525D/A
540D/A
344D/A
18D/A
254D/A
Dimer

— LG-domain
— Middle-domain
— α12-helix
— α13-helix

B. *TCSPC stopped-flow*

Detector G
Detector R
Mixing chamber
Stopped-flow device
PC
Laser 480
Laser 635

Binning (10 ms)
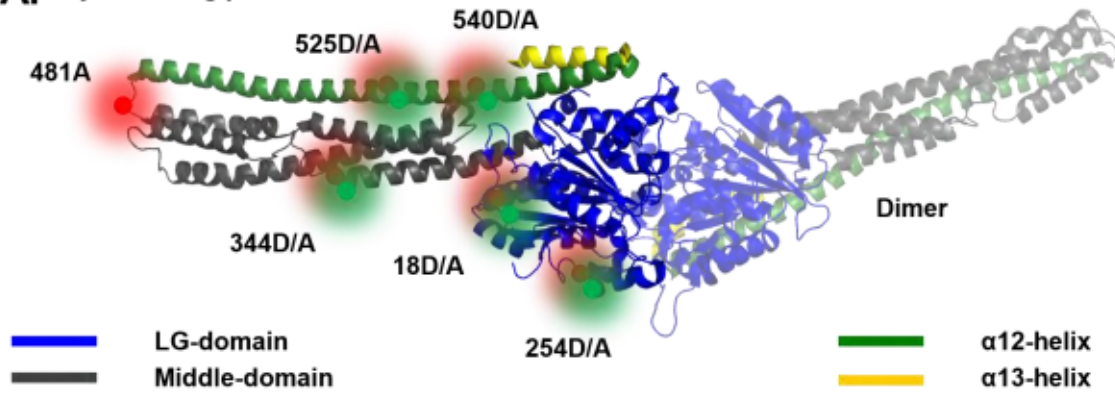
Macrotime fits → Microtime fits
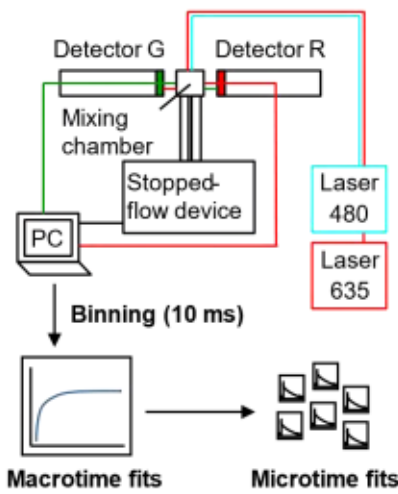
C. *Macrotime*: Fitting whole signal over time (scheme)

Change in FRET photon number

slower transition rates
faster transition rates

10^2 10^3 10^4 [ms]
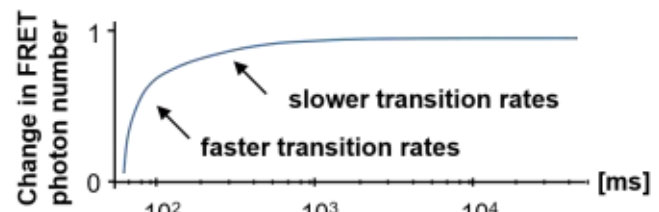
D. *Microtime*: Fitting signal in individual bins (scheme)

FRET: High   Mid   Low/no

Change in fraction x

$= x_{noFRET}$
$= x_{FRET}$

10^2 10^3 10^4 [ms]

E. *Distance distribution*: R_{DA} [Å]

Norm. frac.
0   100
0   100
0   100

Species frac.

FRET fraction II
FRET fraction I
No-FRET fraction

10^1 10^5 [ms]

F. **State:**

Simplified model: 3 states, 2 rates: 2ABC
Extended model: 4 states, 3 rates: 2ABBC

B

2x *A* $\xrightarrow{k_a}$ *B1* $\xrightarrow{k_b}$ *B2* $\xrightarrow{k_c}$ *C*

monomeric hGBP1   flexible dimer   bridged dimer   (semi-)elongated dimer

47
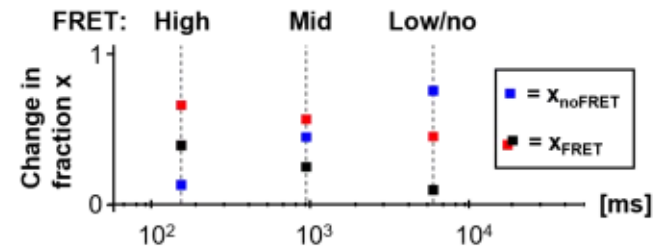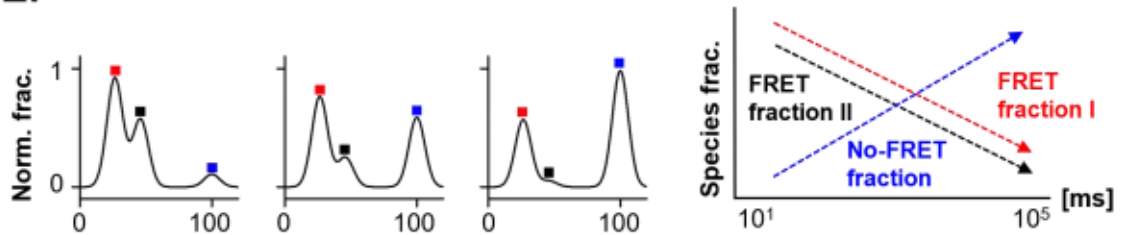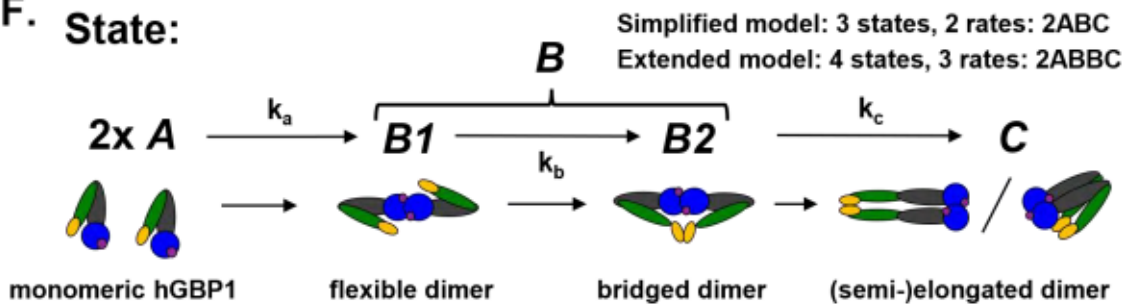
Figure 1: FRET-pairs, SF-FRET setup, readout and scheme of kinetic model. A. Color-coded domain architecture and schematic (no dye-linker included) FRET-network used in this study, based on structures derived from PDB-IDs 1f5n (whole molecule) aligned with 2b92 (dimer interface). **B.** Schematic setup of the SF-TCSPC device. **C.** Macrotime fitting was done applying kinetic models to the FRET-photon fluctuation observed over the course of the experiment, extracting transition rate constants. **D.** Microtime fitting yields FRET-fractions at any macrotime window (within a 10 ms bin). **E.** Distance distributions and x_{DO}/x_{FRET}-fractions change over the course of the macrotime to a no-FRET dominated state for all intramolecular FRET pairs used in this work. **F.** Scheme of kinetic model for the transition from a monomeric, GTP-unbound conformation (free hGBP1) into the consecutive dimeric state. Upon a GTP binding step (GTP in purple), two parts of the LG-domain, **state 2x A**, will form the dimerization surface. Dimer might now cycle between different non-elongated early dimer conformations, **state B**, before transitioning into the consecutive, (semi-)elongated dimers, **state C**. **State B** might be resolvable into two sub-states **B1** and **B2** using an extended model. Higher oligomer assembly is a topic of ongoing discussion and might depend on type of nucleotide used, farnesylation status and presence of pathogenic organisms and was not addressed in this study.

2. Results

Fit function of kinetic models

According to the working model shown in **Fig. 1F**, we developed two consecutive kinetic models to fit the data for the inter FRET (**Fig. 2**) and intra FRET (**Fig. 3**) pairs (see **Chapter 2-Written supplement 2** for full derivation of models). Both models are consecutive (meaning one state follows another $A \rightarrow B \rightarrow \cdots$) and start with a bimolecular assembly step $2A \rightarrow B$. The fraction of each state A, B, C changes over time while the rates are fitted as constant over time. We use two models to confirm if one additional state and one additional rate is needed to increase the goodness of fit. The simplified model (2ABC) with $2A \xrightarrow{k_a} B \xrightarrow{k_c} C$, assumes three states A, B, C and two rates k_a, k_c. The extended model (2ABBC) with $2A \xrightarrow{k_a} B_1 \xrightarrow{k_b} B_2 \xrightarrow{k_c} C$ assumes four states $A, B1, B2, C$ and three rates k_a, k_b, k_c. We analyzed the evolution of FRET induced acceptor intensity over time $I_{Aem|Dexc}(t)$. For both models, the raw data was fitted with the following function body:

$$I_{Aem|Dexc}(t) = x_0 \left(\boldsymbol{x}(t) \cdot \boldsymbol{E} \right) + Bg \tag{1}$$

with x_0 as scaling factor and Bg as background (both fitted parameters), $\boldsymbol{x}(t)$ as the vector of time-dependent fractions of monomer and dimer states (including the monomer state fraction x_A), so $\boldsymbol{x}(t) = (x_A(t), \boldsymbol{x}_{Dimer\,states}(t))$, weighted with corresponding donor brightness proportional to efficiencies $\boldsymbol{E}$ for a given FRET pair. Both $\boldsymbol{x}(t)$ and $\boldsymbol{E}$ were fitted parameters. The fit was done globally for either intra or inter FRET datasets. $\boldsymbol{x}(t)$ was always a joined

parameter within one set, while E, x_0 and Bg were local parameters fitted individually for each FRET pair within one set.

$$\boldsymbol{x}(t) = \begin{pmatrix} x_A(t) \\ x_B(t) \\ \vdots \end{pmatrix}; \; \boldsymbol{E} = \begin{pmatrix} E_A \\ E_B \\ \vdots \end{pmatrix}; \; E_i = \left(1 + \left(\frac{R_{DA,i}}{R_0}\right)^6\right)^{-1}, \; i \in \{A, B, \dots\} \tag{2}$$

with E_i as FRET-efficiency for the given pair, determined by the Förster-Radius R_0 of the dye-pair (here given with 52 Å for the Alexa Fluor 488/647 pair) and the fitted inter-dye distance R_{DA}. The raw data was divided by two, to account for a software error, which doubled the photon counts.

For the simplified three state/two rate 2ABC model $2A \overset{k_a}{\rightarrow} B \overset{k_c}{\rightarrow} C$, the evolution of fractions of states (see **Fig. 1F** for graphical depiction of states) was modeled by the expression:

$$\boldsymbol{x}(t)\,(2ABC) = \begin{pmatrix} x_A(t) \\ x_B(t) \\ x_C(t) \end{pmatrix} = \begin{pmatrix} \frac{1}{1+k_a t} \\ f_r(k_c, k_a, t) \\ \left(1 - \frac{1}{1+k_a t} - f_r(k_c, k_a, t)\right) \end{pmatrix} \tag{3}$$

where k_a, k_c are the (fitted) transition rates from states $2A$ to B to C and the function $f_r(k_i, k_a, t)$ with $i \in \{b, c\}$) the (fitted) convolution of the incoming recombination rate and the exponential decay. It also contains the exponential integral function $Ei(x)$. See **Chapter 2-Written supplement 2, Eq. 28-29** for the full derivation of $f_r(k_i, k_a, t)$.

$$f_r(k_i, k_a, t) = -\frac{1}{k_a t + 1} + e^{-k_i t} + \frac{k_i}{k_a} e^{-\left(\frac{k_i}{k_a} + k_i t\right)} \left(Ei\left(\frac{k_i}{k_a} + k_i t\right) - Ei\left(\frac{k_i}{k_a}\right)\right) \tag{4}$$

$$Ei(x) = \int_{-\infty}^{x} \frac{e^v}{v}\, dv$$

For the extended four state/three rate 2ABBC model $2A \overset{k_a}{\rightarrow} B_1 \overset{k_b}{\rightarrow} B_2 \overset{k_c}{\rightarrow} C$, the evolution of fractions of states (see **Fig. 1F** for graphical depiction of states) was modeled by expression:

$$\boldsymbol{x}(t)\,(2ABBC) = \begin{pmatrix} x_A(t) \\ x_{B1}(t) \\ x_{B2}(t) \\ x_C(t) \end{pmatrix} = \begin{pmatrix} \frac{1}{1+k_a t} \\ f_r(k_b, k_a, t) \\ \frac{k_b}{k_b - k_c}\left(f_r(k_c, k_a, t) - f_r(k_b, k_a, t)\right) \\ \left(1 - \frac{1}{1+k_a t} - \frac{k_b}{k_b - k_c} f_r(k_c, k_a, t) + \frac{k_c}{k_b - k_c} f_r(k_b, k_a, t)\right) \end{pmatrix} \tag{5}$$

Where k_a, k_b, k_c are the (fitted) transition rates from states $2A$ to B to C and the function $f_r(k_i, k_a, t)$ is defined the same as above.

Fitted transition rates and inter-dye distances are given in either **Table 1** (**2ABC model**), **Table 2** (**2ABBC model**) or in **Chapter 2-table supplement 1-3** (scaling factors, background).

Table 1: **2ABC model.** Rates derived, including reduced χ^2. Fitted parameters in **bold**. R_{DA} from [9] averaged by using **Eq. 6**. Large uncertainties (>1000%) for $\langle R_{DA} \rangle_E$ are not shown since $\langle R_{DA} \rangle_E$ was mainly used to stabilize the fitting of rates. * = Inter-FRET data was not fitted in dependency of E. ** = Value taken from Vöpel et al. [7]. *** = Cα-Cα distance of pos. 344 + 481 from PDB-ID 1f5n (whole molecule) aligned with PDB-ID 2b92 (dimer interface), see **Fig. 2A**.

	$k_a(2A \rightarrow B)$ [s⁻¹]	$k_c(B \rightarrow C)$ [s⁻¹]	χ^2
inter FRET	**2.740 ± 0.372**	**0.157 ± 0.003**	1.028
intra FRET	2.740 (fixed)	**0.234 ± 0.004**	1.173

Datasets		inter FRET R_{DA}	Intra FRET R_{DA} from Peulen et al. [9]		Avg.	Fitted R_{DA}		
	Dye pair	$\langle R_{DA} \rangle$	$\langle R_{DA}(M_1) \rangle$ x1=0.61	$\langle R_{DA}(M_2) \rangle$ x2=0.39	$\langle R_{DA} \rangle_E$ (Avg. of M1 and M2)	$\langle R_{DA} \rangle_E$ Monomer State A	$\langle R_{DA} \rangle_E$ Dimer State B	$\langle R_{DA} \rangle_E$ Dimer State C
inter FRET	18D + 18A	63±9**	-	-	62	-	*	*
	344D + 481A	194***	-	-	194	-	*	*
intra FRET	344D/A, 540D/A	-	59±4	45±3	53	**28**	**71**	**92**
	344D/A, 525D/A	-	47±3	21±13	42	-	-	-
	254D/A, 540D/A	-	64±5	37±3	53	**49**	**70**	**121**
	254D/A, 344D/A	-	81±17	72±8	75	**75**	**88**	**95**

Table 2: **2ABBC model.** Rates derived, including reduced χ^2. Fitted parameters in **bold**. R_{DA} from [9] averaged by using **Eq. 6**. Large uncertainties (>1000%) for $\langle R_{DA} \rangle_E$ are not shown since $\langle R_{DA} \rangle_E$ was mainly used to stabilize the fitting of rates.

	$k_a(2A \rightarrow B_1)$ [s⁻¹]	$k_b(B_1 \rightarrow B_2)$ [s⁻¹]	$k_c(B_2 \rightarrow C)$ [s⁻¹]	χ^2
intra FRET	2.740 (fixed)	**1.200 ± 0.045**	**0.115 ± 0.005**	1.017

Datasets		Intra FRET R_{DA} from Peulen et al. [9]		Avg.	Fitted R_{DA}			
	Dye pair	$\langle R_{DA}(M_1) \rangle$ x1=0.61	$\langle R_{DA}(M_2) \rangle$ x2=0.39	$\langle R_{DA} \rangle_E$	$\langle R_{DA} \rangle_E$ Mono. State A	$\langle R_{DA} \rangle_E$ Dimer State B_1	$\langle R_{DA} \rangle_E$ Dimer State B_2	$\langle R_{DA} \rangle_E$ Dimer State C
intra FRET	344D/A, 540D/A	59±4	45±3	53	**35**	**51**	**80**	**120**
	344D/A, 525D/A	47±3	21±13	42	-	-	-	-
	254D/A, 540D/A	64±5	37±3	53	**41**	**52**	**70**	**169**
	254D/A, 344D/A	81±17	72±8	75	**67**	**69**	**82**	**90**

The average of E was calculated with

$$\langle E \rangle = \int \rho(E, \langle R_{DAi} \rangle, \sigma_i)\, dE = \int \sum_{i=1}^{2} x_i \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(R_{DA} - \langle R_{DAi} \rangle)^2}{2\sigma_i^2}\right) E(R_{DA})\, dE$$

$$\langle R_{DA} \rangle_E = \sqrt[6]{\frac{1}{\langle E \rangle} - 1} \cdot R_0 \tag{6}$$

Experimental approach

The hGBP1 variants were based on the Cys-9 [7] variant, a WT construct which lacks all native cysteins to avoid random labeling. The non-native cysteins needed for the labeling reaction with the maleimide-primed dyes were concurrently mutated to the desired locations in the sequence. In the reaction chamber, the non-hydrolysable GTP analogue GTPγS was used to trigger dimerization upon mixing. GTPγS along with GDP*AlF$_x$ has been shown to work as the most potent non-hydrolysable GTP agent under standard conditions [20], using it at 250 µM final concentration. The choice of the GTP-analogue can majorly determine experimental outcomes in different ways [6, 20] and should be taken into account when comparing results. Intramolecular FRET (ex. 480 nm) was measured in double labeled hGBP1 variants, both donor- and acceptor-labeled, with Alexa Fluor 488 and Alexa Fluor 647. Intermolecular FRET was measured between single labeled hGBP1 variants, each variant separately labeled with either donor dye (Alexa Fluor 488) or acceptor dye (Alexa Fluor 647). FRET was measured in pulsed interleaved excitation (PIE) mode to allow integrated acceptor dye analysis [21]. Here, 'green photons' refer to the donor decay excited by the 480 nm laser, 'red photons' refer to the FRET sensitized donor decay and 'yellow photons' to the acceptor decay, excited by the 635 nm laser. The individual decays were binned into 10 ms windows and merged into a macrotime. The macrotime signal was fitted with the two kinetic models shown above to extract the transition rates. In addition, we used a fraction-focused approach for the fitting of intramolecular distances given from [9, 20] in keeping distances constant to ensure fit stability and instead monitor the fraction of different FRET or no-FRET distances. Bi- or tri-exponential fits were applied here (see **6. Methods**).

Inter-FRET measurements: 18D + 18A and 344D + 481A (+ unlabeled Cys-9)

For our two intermolecular dye pairs (**Fig. 2A**, dimer based on PDB-IDs 1f5n (full protein structure) and 2b92 (LG-interface alignment)), we monitored the increase in red photons (FRET-sensitized photons, offset corrected) over the course of the macrotime. Conceptually, the 344D + 481A pair probes the consecutive elongation of the dimer, while the 18A + 18D pair simply shows the starting point of the dimerization by monitoring the LG-LG interaction. The 18D + 18A pair thus serves as our basic dimerization control and determines the first rate, k_a, transitioning from monomer to initial dimer. The 18D + 18A assembly process requires two hGBP1 molecules. Accordingly, this dataset had to be fitted with a second order kinetic term (see above and **Chapter 2-Written supplement 2**). Upon global fitting of both inter-FRET datasets, the 18D + 18A macrotime signal yielded a rate constant of k_a = 2.740 ± 0.372 s^{-1} translating into a bulk transition time of 365 ± 50 ms for initial dimerization (**Fig. 2 B**). A plateau ranges on starting from approx. 5,000+ ms macrotime for 18D + 18A, from where the initial dimerization process enters a state of endpoint equilibrium. In measures of distance, the LG-interaction finds the 18D + 18A pair at $\langle R_{DA} \rangle$ = 63 ± 9 Å [7] at an endpoint equilibrium. The 344D + 481A pair is initially far beyond FRET distance limits, with a theoretical α-carbon to α-carbon distance of 194 Å in the head-to-head LG (presumed) starting dimer conformation (**Fig. 2A**). In order to get an increase in FRET for the 344D + 481A positions, the middle domain containing 344D and the kinking-loop with 481A need to move and converge within a <100 Å distance to each other. There are multiple elongated or semi-elongated conformations possible which would allow this distance. Since these 'late' conformations need dimerization as a necessary pre-condition, the 344D + 481A dataset is assumed to consecutively follow the 18D + 18A. Therefore, a consecutive fitting term had to be applied for the 344D + 481A datasets. Upon global fitting of both inter-FRET datasets, the 344D + 481A pair yielded a rate constant of k_c = 0.157 ± 0.003 s^{-1} translating into a bulk transition time of 6373 ± 127 ms for the consecutive rearrangements leading to the converging of 344D and 481A (**Fig. 2 B**). Using this model results in smooth residuals and in a good red. chi-square of χ^2 = 1.028. Applying a 2ABBC model (see chapter: Intra-FRET measurements) to the inter-FRET data gained no increase in goodness of fit. A sketch of the conformational arrangements likely linked to the observed data is depicted in **Fig. 2C**. There, the 18D + 18A pair indicates in respect to interaction between the LG-domains and subsequent initial dimer formation why 344D + 481A indicates for the 'late' dimers in respect to the middle-hinge-interaction and subsequent dimer elongation.
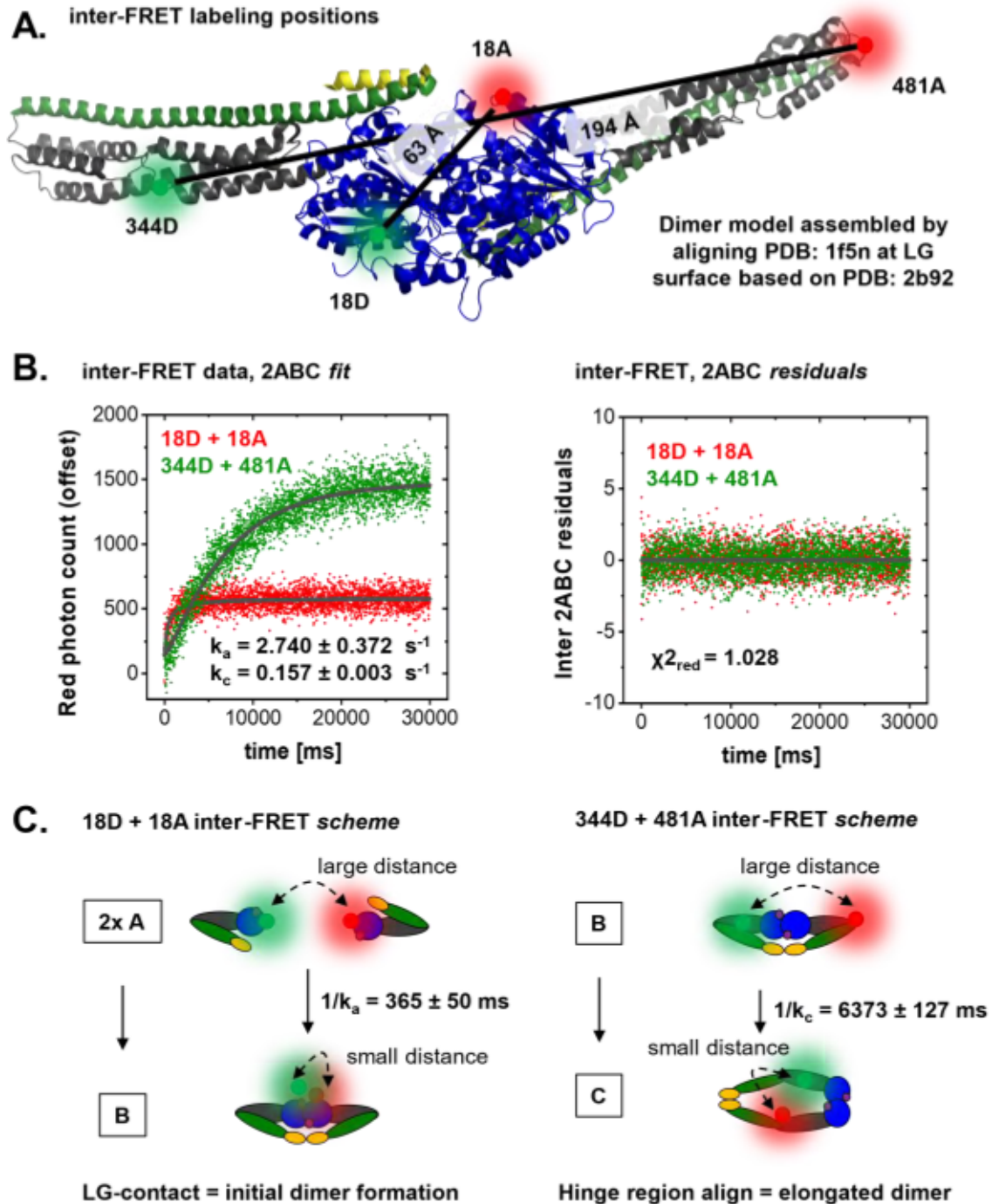
Figure 2: Inter-FRET pairs, data and schematic explanations. Concentrations given with: 250 µM GTPγS, 10 µM acceptor-labeled hGBP1, 200 nM donor-labeled hGBP1, filled up to 25 µM total hGBP1 concentration with unlabeled hGBP1. A. Pairs used for inter-FRET detection. 18D + 18A distance has been determined in [20], 344D + 481A from a Cα-Cα distance in the PDB-ID: 1f5n crystal structure. **B.** Global 2ABC fit on both inter-datasets yields two rates k_a and k_b. Residuals and χ2 show high goodness of fit for the applied model. **C.** Schematic display of the proposed domain movements detectable over the course of the macrotime.

Intra-FRET measurements: (254D/A, 344D/A), (254D/A, 540D/A), (344D/A, 525D/A), (344D/A, 540D/A) + unlabeled Cys-9

For our four intramolecular dye pairs (**Fig. 3A**, PDB 1f5n), we monitored the decrease in red photons (FRET-sensitized photons, offset corrected) over the course of the macrotime. Since the hGBP1 monomer cycles between two main conformations, two distances are assigned to every dye pair (see **Fig. 3A**). Conceptually, the (254D/A, 344D/A) pair probes intramolecular distance between the middle domain and the LG domain, while the three other intra-pairs all detect opening of the α12-helix, away from the middle domain, on different labeling positions. The macrotime trace for all pairs needed to be fitted with an at least bi-exponential decay function. The fit model had to be consecutive since the intramolecular dynamics only change after initial dimerization. The resulting extended kinetic model, 2ABBC (see **2. Results**, Eq. 5) assumes a consecutive conformational change from B1 -> B2, with B1 being the dimer conceived from 2A. The simplified 2ABC kinetic model (see **2. Results**, Eq. 3) assumes an immediate conformational change from 2A -> B. Upon fitting, we noticed a constant misfit of the (344D/A, 525D/A) dataset. Cross checking the photon histograms revealed a roughly tenfold higher relative number of yellow photons (compared to the other three datasets) found in the PIE-window, which translates into increased acceptor concentration. Since we could not exclude an experimental error in the concentration of labeled protein for this pair, the dataset was excluded from fitting. The 2ABC fit yields a slow component B -> C of k_c = 0.234 ± 0.004 s^{-1} (see **Fig. 3B**) translating into a bulk transition time of = 4274 ± 73 ms, using k_a = 2.740 (fixed) from the inter-FRET dataset as fast 2A -> B component. There is a visible misfit in the residuals for 2ABC, resulting in a mediocre red. chi-square of χ^2 = 1.173. The 2ABBC fit yields the slowest component B2 -> C with k_c = 0.115 ± 0.005 s^{-1} (see **Fig. 3C**) translating into a bulk transition time of = 8696 ± 350 ms. The additional B1 -> B2 component yields a rate of k_b = 1.200 ± 0.045 s^{-1} (see **Fig. 3C**) translating into a bulk transition time of = 833 ± 31 ms. We were also using k_a = 2.740 (fixed) from the inter-FRET dataset as fast 2A -> B1 component in the 2ABBC model. The residuals are very smooth, resulting in a good red. chi-square of χ^2 = 1.017. The overall clearly better goodness of fit for the 2ABBC model suggests that at least this one additional component is necessary to fully describe the data. A sketch of the conformational arrangements that might be linked to the observed data is depicted in **Fig. 3D**. There, the (254D/A, 344D/A) pair indicates intramolecular spacing between the middle and LG-domain why the other pairs follow the opening of the α12-helix and concurrent dimer elongation.

A. intra-FRET labeling positions

525D/A   540D/A

42 A

53 A

53 A

PDB: 1f5n

344D/A

75 A

254D/A

B. intra-FRET data, 2ABC fit

k_a = 2.740 s^{-1} (fixed)
k_c = 0.234 ± 0.004 s^{-1}

intra-FRET data, 2ABC residuals

(254D/A, 344 D/A)
(254D/A, 540 D/A)
(344D/A, 540 D/A)

$\chi 2_{red}$ = 1.173

C. intra-FRET data, 2ABBC fit

k_a = 2.740 s^{-1} (fixed)
k_b = 1.200 ± 0.045 s^{-1}
k_c = 0.115 ± 0.005 s^{-1}
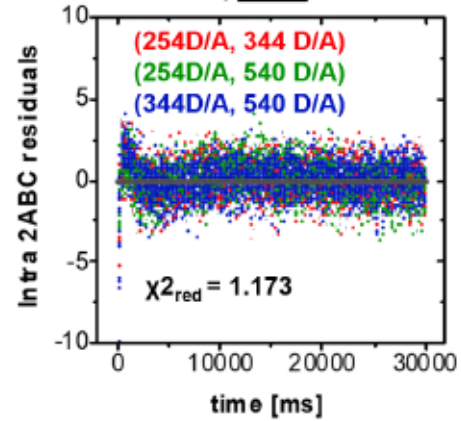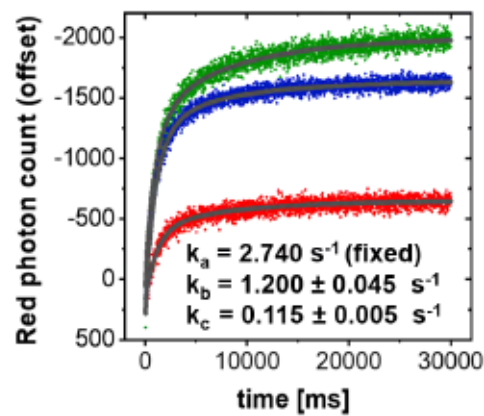
intra-FRET data, 2ABBC residuals

(254D/A, 344 D/A)
(254D/A, 540 D/A)
(344D/A, 540 D/A)

$\chi 2_{red}$ = 1.017
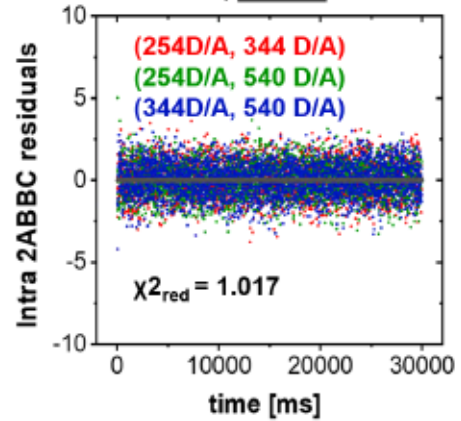
D. (254D/A, 344D/A) intraFRET scheme, 2ABBC

[B1 | B2]

smaller distance

$1/k_a$ = 365 ms (fixed)
$1/k_b$ = 833 ± 31 ms
$1/k_c$ = 8696 ± 350 ms

C

larger dist.

Spacing between middle and LG domain
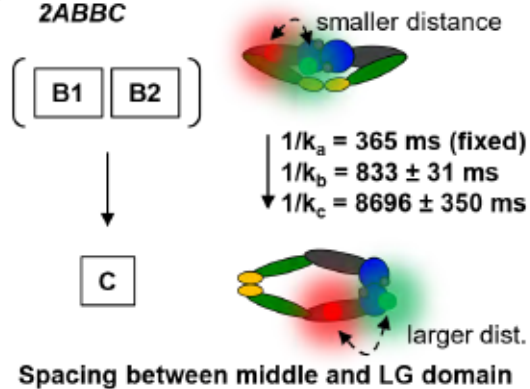
(254D/A, 540D/A) and (344D/A, 540D/A) intraFRET scheme, 2ABBC

[B1 | B2]

small distance

$1/k_a$ = 365 ms (fixed)
$1/k_b$ = 833 ± 31 ms
$1/k_c$ = 8696 ± 350 ms
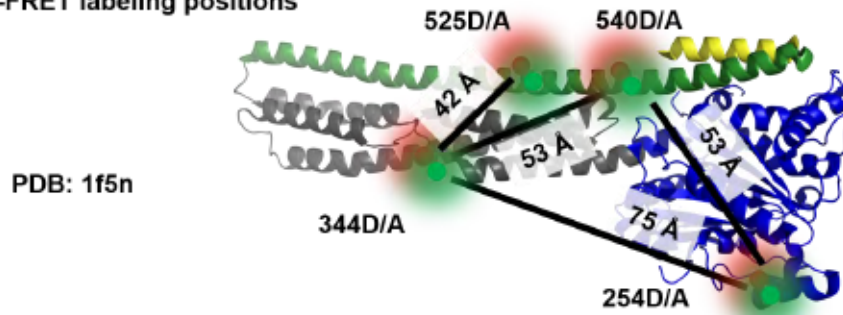
C

large distance

α12 helix hinge out = elongated dimer
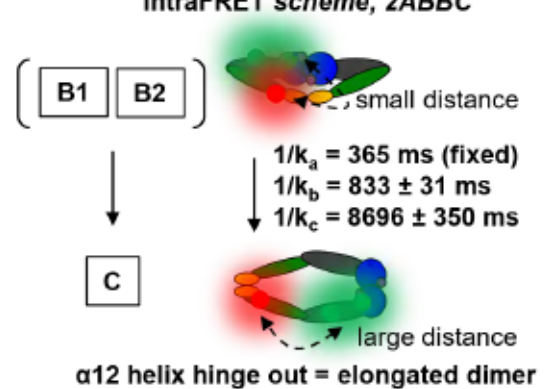
Figure 3: Intra-FRET pairs, data and schematic explanations. Concentrations given with: 250 µM GTPγS, 200 nM donor/acceptor-labeled hGBP1, filled up to 25 µM total hGBP1 concentration with unlabeled hGBP1. A. Pairs used for intra-FRET detection. Monomer distances have been determined in [9]. Structure: PDB 1f5n. **B.** Global 2ABC fit on three intra-datasets yields a rate k_c using a fixed fast rate k_a from inter-FRET pairs. Residuals and χ2 show mediocre goodness of fit for the applied model. **C.** Global 2ABBC fit on three intra-datasets yields two rates k_b and k_c using the fixed fast rate k_a from inter-FRET pairs. Residuals and χ2 show high goodness of fit for the applied model. **D.** Schematic display of the proposed domain movements detectable over the course of the macrotime.
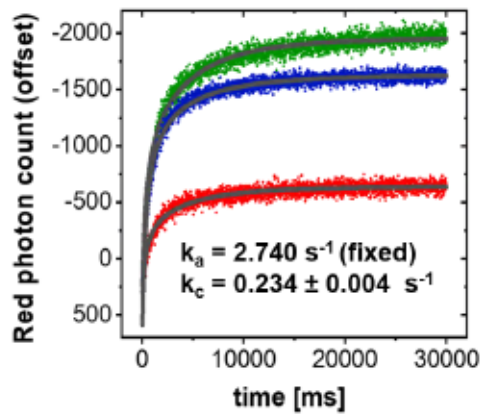
3. Discussion

In this work, we were successfully following the transition of labeled Cys-9 hGBP1 from monomer to dimer with our eTCSPC/SF-setup. We were able to extract transition rates by applying and comparing two multi-exponential, consecutive kinetic models. The conclusion we can make agree with and expand on data and models presented by Hengstenberg et al [20], Peulen et al [9] and Völpel et al [7]. Minor discrepancies can be well explained by varying conditions. Especially the choice of the GTP-analogous nucleotide has an impact as all of the different nucleotides show a unique affinity and even multimerization potential [20] as well as differences in the dimerization interface [6]. The GTPγS used in this work has been shown to be one of the most potent non-hydrolyzable nucleotides for dimerization of hGBP1, used at a final concentration of 250 µM [20]. When comparing the data between different GTP-analogues, one should always consider the possibility of a unique behavior of the nucleotide used and refrain from generalization. Also note that the Cys-9 variant has been shown to have a higher GTP hydrolyzation rate than the wildtype (WT), which might lead to a divergence from the actual wildtypic time axis in which the transitions happen.

Within the complex energy landscape of a multi-conformational protein like hGBP1, all FRET pairs (intra- and inter-FRET) follow spatial changes, which can usually be seen as a sequence of domain movements relative to each other. In our experiments, this happens over the course of 20 s after GTPγS-induced dimerization, under the necessary condition that the distance of the dye pair changes for a least a few Angstrom from the original (starting) value. In general, we see a decrease of red photons for all intra-FRET variants (as depicted in **Fig. 3**) and an increase in red photons for all inter-FRET variants (as depicted in **Fig. 2**). The timescale on which the red signal changes happen depends on the variants used (each variant probes a unique set of distances) and the observed domain movement and arrangement. For the sake of simplicity, we handle conformations as discrete states in our interpretation of the data. Conformational ensemble or a mix of transient domain-rearrangements is possible while transitioning to the (semi-)elongated form.

Looking at the intermolecular FRET measurements (see **Fig. 2B**) there is a very clear idea of the conformational changes behind the data. More importantly, we were able to resolve the time-axis on which these changes happen. Applying the necessary second-rate kinetic model, the timescale of the 18D + 18A interaction appears to be more than one order of magnitude faster than the one found for 344D + 481A. This proves conceptually that the initial LG-dimerization and the subsequent domain rearrangement happen on different timescales within the first seconds after addition of GTPγS. The LG-domain, where 18D + 18A is located, functions as first-line dimerization interface of hGBP1. Consequentially, we expect that of all intermolecular dimer-induced effects, the 18D + 18A interaction will happen first. The two hGBP1 molecules will likely dock as monomers into a head-to-head conformation, but subsequent dynamics might take place instantly (in a very short time-window after initial dimerization) . We assume that the dimer interface remains unchanged while the rest of the molecule transitions, this assumption is backed by the 18D + 18A signal staying constant during measurement. There is a possibility that the interface positioning changes slightly while maintaining constant distance between 18D +18A. In the rigid head-to-head conformation depicted in **Fig. 2A**, the 344D + 481A would not give any FRET signal. The distance between the two labels is far above FRET range (194 Å Cα-Cα). To have FRET between 344D + 481A, the two hGBP1 molecules need to adopt a conformation where the middle domains of both molecules are bended towards each other. Any elongated or semi-elongated dimer with close proximity hinge-regions would be a likely candidate. The 344D + 481A data must be treated consecutively to the 18D + 18A since the dimer needs to assemble first (see **Fig. 4A**). Additionally, the conformational transition is clearly slower and happens in a different time regime. The energy barriers to overcome for transition into an (semi-)elongated state likely have a higher bulk transition time.

In the intra-FRET measurements (see **Fig. 3A/B**) all pairs show a similar and immediately decreasing red signal after mixing. This suggests that α12-helix (pos. 540) moves away from the middle (pos. 344) and LG-domain (pos. 254). These results strongly imply an elongated or elongated-kinked (semi-elongated) dimer conformation. Like the 344D + 481A pair in inter-FRET, a change in red signal for all intra-FRET pairs first needs a dimerized LG-domain. That puts all the intra-FRET data consecutively behind the transition event observed by 18D + 18A inter-FRET. Prior to dimerization, the monomers should be in an equilibrium between their M1 and M2 states. The conformation between the monomeric states and the first dimer might be the same if only the LG-interfaces dock while maintaining monomeric M1/M2 conformation in the rest of the molecule. There might also an immediate conformational change upon dimerization which happens within our resolution of the 2A -> B assembly and can thus by represented by k_a. Until now, the two models (2ABC vs 2ABBC, see **Fig. 4A/B**) both use a fixed k_a as a reference point for the consecutive transitions. We would then fit one

or two additional rates, which should in some sense represent the lower (elongation) rate k_c already extracted from the 344D + 481A inter-FRET pair. Using only one additional rate k_c in the 2ABC fit yielded a faster transition at approx. 4300 ms bulk transition time B -> C but was clearly missing a rate when referring to residuals (Fig. 3B) and red. chi-square. Feeding the fit an additional rate k_b in the 2ABBC model improved the goodness of fit tremendously and clearly shows that the 2ABBC is the superior model to analyse the intra-FRET pairs. We now end up with an approx. 8700 ms bulk transition time for the slowest transition B2 -> C, which is a likely representative for the dimer elongation also observed in the 344D + 481A inter-FRET pair. The difference might be explainable by a transient transition ensemble in which pos. 344 and 488 get into FRET distance slightly before the intra-FRET pairs getting into no-FRET distance, but this is heavy speculation. Looking at the additional transition B1 -> B2 found at approx. 830 ms, we likely see the transitioning towards an (semi-)elongated form. An educated guess would be the interpretation of B2 as the 'bridged dimer' (see **Fig. 4B**), a conformation predicted and shown by Vöpel et al [7]. The publication suggests increased spacing between the α12-helix and LG/middle domains, while still not transitioning in the (semi-)elongated 'late' conformations which would fit our data very well. In equilibrium, the bridged dimer seems to be the favorable conformation (see **Fig. 4E**). The spacing is supposed to be driven by the alignment of the two α13-helices, making it an obvious target for our technique. Unfortunately, we lack the appropriate measurements with dyes positioned at the α13-helices to experimentally verify this theory.

The (254D/A, 344D/A) pair is somewhat different from the other intra-FRET pairs since it probes the intramolecular distance between LG and middle domain. Interestingly, it fits very well within the same 2ABBC model, which enabled us to do global fitting. The transitions seen here likely means that the pronounced movement from the α12-helix away LG/middle is feedbacking into the dynamics between LG and middle domain. As such, pretty much the whole molecule seems to be in movement. The internal LG/middle dynamics could work as an energetic counterweight to compensate the vast movement of the α12-helix into a somewhat open, likely unfavorable state. We find very good agreement with the 2ABBC macrotime fits when looking at FRET-fractions obtained from lifetime fits of (254D/A, 344D/A) at different micro-timepoints (see **Fig. 4C** and chapter supplement). After the fast initial dimerization window (interpreted as 2A -> B1), the longer M1 distance starts to peak towards 1000 ms while the shorter M2 distance goes down concurrently (interpreted as B1 -> B2). This observation is clearly indicating a consecutive conformational hierarchy. The no-FRET distance appears for the first time after 1000 s and starts to raise in fraction while M1 keeps decreasing (interpreted as B1 -> C). In terms of conformations, right after mixing, the contact between the LG and middle domain needs to loosen/start spacing to push the low-FRET fraction M1 up. Until now, the middle domain is probably not fully turned away from the LG.

A. Simplified model: Three states, two rates: 2A-B-C. χ^2_r (inter) = 1.028, χ^2_r (intra) = 1.173

$$2x\ A \xrightarrow{1/k_a \approx 370\ ms} B \xrightarrow[1/k_c \approx 4300\ ms\ (intra\ 2ABC)]{1/k_c \approx 6400\ ms\ (inter\ 2ABC)} C$$

monomeric hGBP1 — Initial hGBP1 dimer — (semi-)elongated hGBP1 dimer
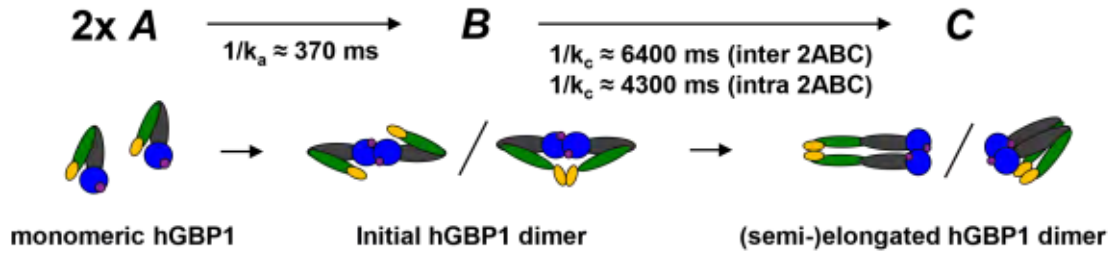
B. Extended model: Four states, three rates: 2A-B1-B2-C. χ^2_r (intra) = 1.017

$$2x\ A \xrightarrow{1/k_a \approx 370\ ms} B1 \xrightarrow{1/k_b \approx 830\ ms} B2 \xrightarrow{1/k_c \approx 8700\ ms} C$$

monomeric hGBP1 — flexible dimer — bridged dimer — (semi-)elongated dimer
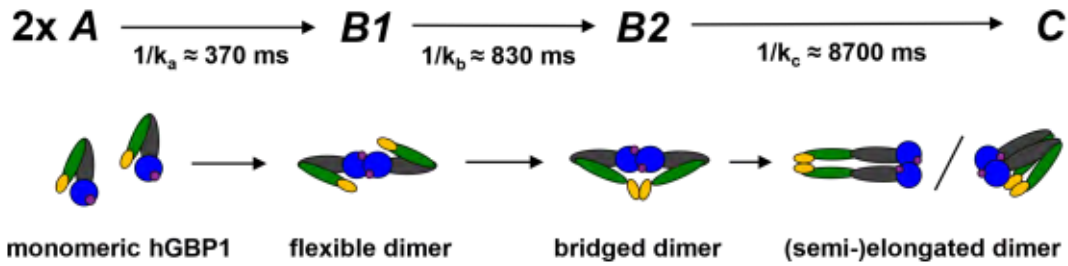
C. (254D/A, 344D/A) fraction change over time
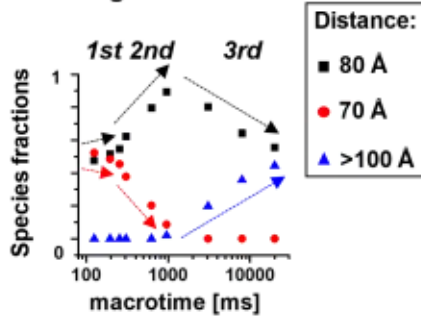
D. *Sketch* of data from [20] shows three rates
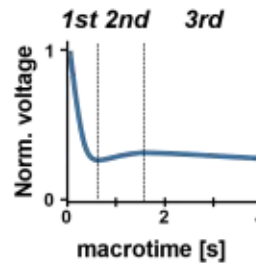
E. Dimer equilibrium dynamics
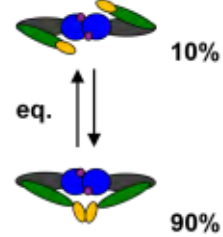
Figure 4: Concluding schemes and hypothesis for hGBP1 transition. **A.** Simplified 2ABC model yielded two transition rates from both intra- and inter-FRET datasets: k_a from monomer to initial dimer and k_c (inter-FRET and intra-FRET) towards the consecutive (semi-)elongated dimer. **B.** Extended 2ABBC model yields three transition rates from intra-FRET datasets: k_a from monomer to flexible dimer, k_b from flexible dimer to bridged dimer and or k_c towards the consecutive (semi-)elongated dimer. *The 2ABBC model yields higher goodness of fit and is therefore preferred for the intra-FRET measurements.* **C.** (254D/A, 344D/A) fraction change over time. Before/while initiating dimerization, the FRET-pair is within FRET-range with zero no-FRET contribution. Approx. between 300-1000 ms, the longer distance fraction goes up while the shorter distance fraction decreases. From >1000 ms, the no-FRET distance comes into account. **D.** (Sketched) data from [20] already suggests three rates as well for the (344D/A, 540D/A) and other pairs. Measurement made with GTP and different protein concentration might cause slightly different behavior. **E.** Dimer equilibrium dynamics suggested by Vöpel et al [7] could explain the B1 -> B2 transition observed in our data. The bridged dimer dominates the equilibrium with a fraction of 90% making it possible to approximate the back reaction contribution as zero in our 2ABBC model.

For the final transition, the middle domain turns more into its (semi-)elongated positioning, increasing the distance between pos. 254 and pos. 344 further into the no-FRET distance. An (semi-)elongated conformation suggested by Hengstenberg et al. [20] or the one simulated for hGBP5 [24], where the molecules are semi-elongated/partially closed/kinked, but aligned, seems likely. We still want to suggest the possibility for a fully elongated dimer. Note, that a kinked dimer in the way suggested by [20] and [24] would likely mean a late (B2 -> C) return of all intra-FRET pairs into FRET-distance, which is not visible in our data. Possibly, the domain movements observed could be specific for GTPγS while Hengstenberg et al. [20] did their experiment with GTP.

4. Conclusion

Based on our work and the work of others we established a working model which can temporally resolve and structurally explain the full transition from monomeric, unfarnesylated, fluorescently labeled Cys-9 hGBP1 to an endpoint dimer (termed as '(semi-)elongated dimer') upon addition of GTPγS. Additionally, the workflow of our stopped flow TCSPC-setup was successfully established. The conformational changes observed over the course of the macrotime was based on five dye-pairs available from our measurements. Fitting the FRET photons with the models introduced in **2. Results**, we got three distinct transition rates. While the inter-FRET data could be fitted using two rates and the simplified 2ABC model, the intra-FRET data showed an increased goodness of fit using an additional third rate, rendering the 2ABBC model superior. The 2ABBC model should therefore be used for further datasets, while the 2ABC can serve as back-up to stabilize the fits if there is obvious overfitting with the 2ABBC. The rates could be interpreted based on structural knowledge and dye positioning. In addition, we were able to extract distances from individual microtime windows by fluorescent lifetime fitting. As endpoint in our measurement window, we propose a (semi-)elongated dimer where the two molecules align most of their domains in a stretched-out fashion. Before the elongation, the dimer could be in a bridged configuration, a structure that has been proposed before. We can detect a clear consecutive order of the transitions. The proposed states are likely solutions based on the data we have but possible additional states or ensembles as well as the exact movement trajectories remain elusive.

5. Outlook

To clarify things further, the measurements could be repeated with additional pairs, different GTP-analogues, in concentration curves or with the farnesylated species. In the light of our results, especially variants with α13-dyes could be interesting to see if the B2 state is indeed the bridged dimer. Nonetheless, the experiments in this work shaped the conceptual basis

for further experiments to elucidate the consecutive transition. Cleaner and more extensive datasets could enable even more precise fits. Bioinformatics can provide comparative modeling to further understand how hGBP1 goes through its complex transition.

6. Methods

Experimental procedure

Purified hGBP1 was available in concentrated (>µM) labeled stock solutions from previous [7] and running projects. hGBP1 variants used were thawed and filtered through 100k MWCO spin-filter (Merck, Darmstadt, Germany). The concentration of each respective variant was determined on a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham (MA), USA). The labels were Alexa Fluor 488 (Donor, D) and Alexa Fluor 647 (Acceptor, A), respectively (both Thermo Fisher Scientific). For SF-TCSPC measurements, two buffers were freshly prepared, a protein buffer with hGBP1 and a GTP-buffer with GTPγS. The following final concentrations refer to the value <u>after</u> 1:1 mixing in the stopped flow chamber. In syringe 1, we loaded protein buffer: DA-labeled (intra-FRET) hGBP1 at 200 nM, D-labeled (inter-FRET) hGBP1 200 nM and A-labeled (inter-FRET) hGBP1 for 10 µM in a spectroscopically clean PBS. The protein buffer also contained the amount of unlabeled hGBP1 calculated to fill up to a final hGBP1 concentration of 25 µM (total labeled + unlabeled protein) after mixing. In syringe 2, the GTP-buffer contained GTPγS (Jena Bioscience, Jena, Germany) for a final concentration of 250 µM after mixing in a spectroscopically clean PBS buffer.

Experimental setup

SF-TCSPC was done in a SX Series Sequential Stopped-Flow Sample Handling Unit (Applied Photophysics, Leatherhead, UK) coupled to two Photomultiplier Tubes H72422P-40 (Hamamatsu Photonics, Hamamatsu, Japan), one upstream with a ET535/30m emission filter (Chroma Technology, Bellows Falls (VT), USA) and the other one upstream with a HQ730/140 emission filter (Chroma Technology). PMTs were controlled via a DCC-110 card (Becker & Hickl, Berlin, Germany) and TCSPC was operated via a SPC-830 card (Becker & Hickl). Excitation was done with LDH-P-C-470 and LDH-P-C-635M laser diode heads (PicoQuant, Berlin, Germany) operated by a PDL 800 and a PDL 800-B driver unit. Experiments were done in PIE-mode and under magic-angle conditions. Repetition rate was set to 20 MHz, Time-amplitude converter (TAC) window was set to 60 ns. The two syringes of the stopped-flow unit were holding a max. volume of approx. 1.2 ml each for 20 shots at 120 µl total volume per shot. The link to a detailed operational guide can be found in **Chapter 2-Methods supplement 1.**

FRET photon data analysis over macrotime (mathematical solution)

The full derivation can be found in **2. Results** and **Chapter 2-Written supplement 1.**

Diffusion limit of enzymatic reactions

To avoid wrong conclusions in the case of very fast rates, we must show that the rate yielded for our initial, bimolecular reaction $2A \xrightarrow{k_{bi}} B$ is below the threshold of diffusion limited rates. The diffusion limit of rates of enzymatic reactions in PBS-buffer at 20 °C is given with

$$k_{bi}(max) = \frac{8}{3}\left(\frac{RT}{\eta}\right) = 6.5 \cdot 10^9 \ s^{-1}M^{-1} \tag{7}$$

and by comparing the limit (using a concentration of $c(hGBP1) = 25 \ \mu M$ and $k_a = 2.74 \ s^{-1}$)

$$k_{bi}(hGBP1) = \frac{k_a}{c(hGBP1)} = 1.1 \cdot 10^5 \ s^{-1}M^{-1} \tag{8}$$

we can conclude that the initial reaction is not diffusion limited, since $k_{bi}(max) \gg k_{bi}(hGBP1)$.

Fluorescence lifetime data analysis

Raw data was prepared using the software "*AnI*-3SF" developed in *Seidel* group, as part of the software package for multiparameter TCSPC/imaging. Macrotime data was merged from the start of each shot by using the stopped flow injection as a line trigger. Additionally, all data within a 10 ms time-well was merged to further enhance photon statistics, starting at 50 ms after mixing. Constant merging window of 10 ms was applied over the whole macrotime. All fits and detailed parameter are available in **Chapter 2-Table supplement 4**. As an example, the decay fitted for the 55 ms time point represents the average/merged data from approx. 20 repeated shots and all TAC windows (166,692) in between 50 ms and 60 ms. TAC resolution was set to a 512-channel binning, yielding a time resolution of dt = 0.11717 ns. Fluorescence decay analysis was done using "ChiSurf", a global analysis platform for fluorescence data developed as open-source software https://github.com/fluorescence-tools.
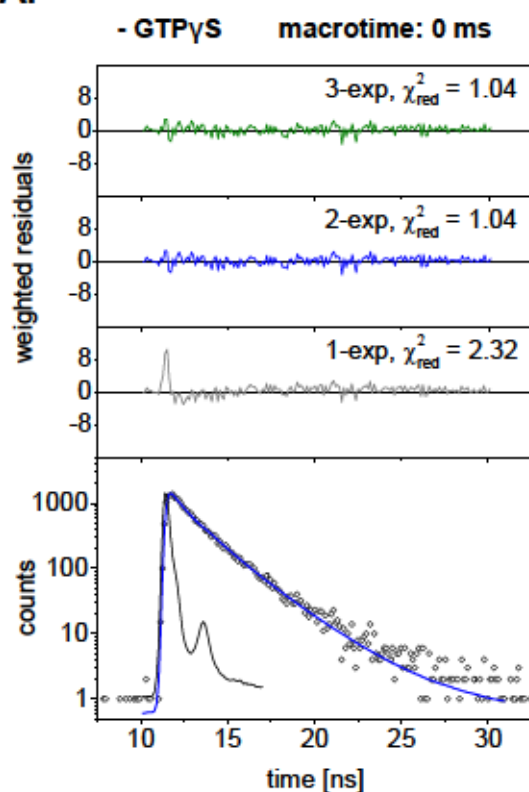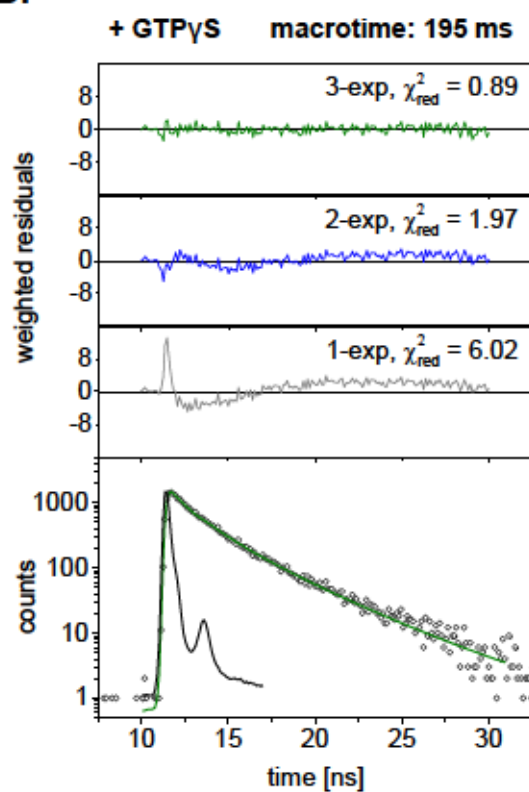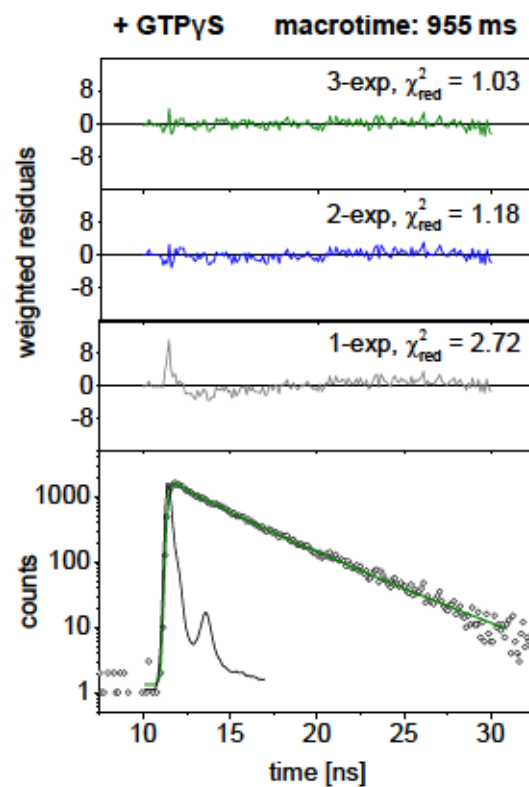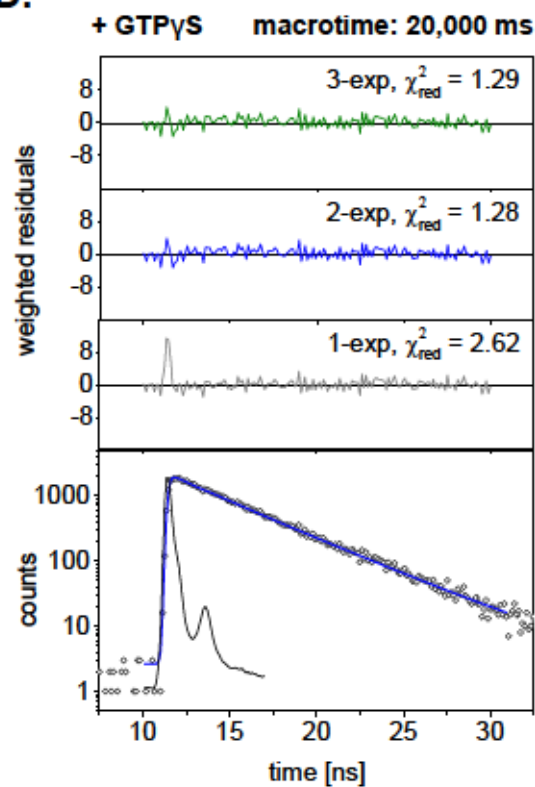
Figure 5: A. Exemplary fitting sequence of the 344-525 FRET-pair, displaying the changes in fluorescence lifetime over time. Mono- (grey), bi- (blue) or tri-exponential (green) fits of the 344-525 intramolecular FRET pair to exemplify lifetime behavior and resulting distance distributions. **A. Without GTPγS**, the bi-exponential fit (blue) is sufficient to describe the two main conformers of the hGBP1 monomer, regardless of the macrotime chosen. **B. With GTPγS and at a macrotime of 195 ms**, the tri-exponential fit (green) shows the best goodness, the distance distribution is almost even between the two main conformers of the monomer and the emerging low/no-FRET distance (>100 Å) from the dimeric state, making three lifetime components mandatory. **C. With GTPγS and at a macrotime of 955 ms**, the distance distribution is already shifted towards the low/no-FRET distance of the dimeric state but the other two fractions, in particular the shorter one, are still present in the fit, thus favoring again the tri-exponential fit (green). **D. With GTPγS and at a macrotime of 20 s**, the low/no-FRET distance dominates. Still, the remains of the monomeric distances' fractions are existing in this near-equilibrium state but can be reduced to one component, closely making the bi-exponential fit (blue) the most eligible.

In general, FRET effects will add additional energy transfer modes, increasing the number of exponents needed to describe the fluorescence decay, as exemplary shown in **Fig. 5**. Fluorescent decays of donor (Eq. 9) dye without the possibility to engage in FRET serves as a reference for dye behavior and ubiquitous quenching effects implied by the protein (e.g. quenching amino acids next to the dye), buffer background and setup-specific features.

$$f_{D|D}^{DO}(t) = \sum_i x_D^i \exp\left(-\frac{t}{\tau_D^i}\right) \tag{9}$$

with *DO* representing the donor in absence of the acceptor, *D|D* representing donor emission upon donor excitation, *i* representing the number of exponents needed to fit the system, x_D representing the donor fraction of each exponential expression and τ_D representing the donor lifetime of each exponential expression.

Knowing that the donor will be quenched by FRET in addition to Eq. 9, we derive Eq. 10, describing the FRET-induced donor decay on top of the donor-only decay.

$$f_{D|D}^{DA}(t) = f_{D|D}^{DO}(t) \cdot \sum_i x_{FRET}^i \exp\left(-t \cdot k_{FRET}^i\right) \tag{10}$$

with *DA* representing the donor in presence of the acceptor, x_{FRET} representing the fraction of molecules undergoing FRET corresponding to each FRET-rate k_{FRET} (with *i* depending of the number of energy transfer modes described by different transfer rates).

The same equation can also be expressed using the DA inter-dye distance R_{DA} as shown in Eq. 11.

$$f_{D|D}^{DA}(t) = f_{D|D}^{DO}(t) \cdot \left(x_{DO} + \int_{R_{DA}} p(R_{DA}) \cdot \exp\left(-t \cdot k_0 \cdot \left(\frac{R_0}{R_{DA}} \right)^6 \right) dR_{DA} \right) \qquad (11)$$

with $p(R_{DA})$ representing the DA inter-dye distance distribution, R_0 representing the Förster-radius (R_0 = 52 Å for Alexa Fluor 488) with a constant dipole orientation factor κ^2 = 2/3, and k_0 representing the inverted donor-only lifetime $1/\tau_{DO}$ (τ_{DO} = 4.0 ns for Alexa Fluor 488). If the protein was labeled insufficiently with acceptor dye, the acceptor dye was bleached or the dyes are above the maximum FRET-distance for the dye pair, x_{DO} accounts for the fraction of the donor only species associated to these effects. Accounting for uncertainty, $p(R_{DA})$ was modeled as a superposition of Gaussian distributions as shown in Eq. 12 (exemplary for two distributions)

$$p(R_{DA}) = \frac{x_1 e^{-\left(\frac{\sqrt{2}\left(R_{DA}-R_{DA}^{(1)}\right)}{w_{DA}} \right)^2} + (1-x_1)e^{-\left(\frac{\sqrt{2}\left(R_{DA}-R_{DA}^{(2)}\right)}{w_{DA}} \right)^2}}{\sqrt{\frac{\pi}{2}} w_{DA}} \qquad (12)$$

With the width of the Gaussian distribution represented by w_{DA} = 2σ.

$f(t)$ functions were fitted with an iterative re-convolution approach of $f(t)$ to the instrument response function (IRF), using a modified Levenberg-Marquardt algorithm [25]. In incorporating experimentally conditioned parameters such as light scatter and background, the model function $g(t)$ based on $f(t)$ is derived in Eq. 13.

$$g(t) = N_F \cdot f(t) \otimes IRF(t) + N_{BG} \cdot IRF(t) + BG \qquad (13)$$

With N_F representing the number of fluorescence photons from all processes, N_{BG} representing the number of background photons (either as background file or as fitting parameter) and BG representing the detector offset. The absolute value of $g(t)$ scales with the number of photons in the data used for the respective fit. Reduced chi-square and Durbin-Watson test are utilized to describe the agreement between the model function $g(t)$ and the data.

7. Chapter 2 References

1. Prakash, B., Praefcke, G. J. K., Renault, L., Wittinghofer, A., Herrmann, C. Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins. *Nature*, 2000, **403**, 567-571. DOI: 10.1038/35000617.

2. Degrandi, D., Konermann, C., Beuter-Gunia, C., Kresse, A., Würthner, J., Kurig, S., et al. Extensive characterization of IFN-induced GTPases mGBP1 to mGBP10 involved in host defense. *Journal of Immunology*, 2007, **179**, 7729-7740. DOI: 10.4049/jimmunol.179.11.7729.

3. Britzen-Laurent, N., Bauer, M., Berton, V., Fischer, N., Syguda, A., Reipschlager, S., et al. Intracellular trafficking of guanylate-binding proteins is regulated by heterodimerization in a hierarchical manner. *PLOS One*, 2010, **5**, 11. DOI: 10.1371/journal.pone.0014246.

4. Kravets, E., Degrandi, D., Ma, Q. J., Peulen, T. O., Klumpers, V., Felekyan, S., et al. Guanylate binding proteins directly attack Toxoplasma gondii via supramolecular complexes. *Elife*, 2016, **5**. DOI: 10.7554/eLife.11479.

5. Kunzelmann, S., Praefcke, G. J. K., Herrmann, C. Nucleotide binding and self-stimulated GTPase activity of human guanylate-binding protein 1 (hGBP1). 2005, **404**, 512-527. DOI: 10.1016/s0076-6879(05)04045-0.

6. Ghosh, A., Praefcke, G. J. K., Renault, L., Wittinghofer, A., Herrmann, C. How guanylate-binding proteins achieve assembly-stimulated processive cleavage of GTP to GMP. *Nature*, 2006, **440**, 101-104. DOI: 10.1038/nature04510.

7. Vöpel, T., Hengstenberg, C. S., Peulen, T. O., Ajaj, Y., Seidel, C. A. M., Herrmann, C., et al. 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*, 2014, **53**, 4590-4600. DOI: 10.1021/bi500524u.

8. Zhu, S., Bradfield, C. J., Mamińska, A., Park, E., Kim, B., Kumar, P., et al. Cryo-ET of a human GBP coatomer governing cell-autonomous innate immunity to infection. *bioRxiv*, 2021, 2021.08.26.457804. DOI: 10.1101/2021.08.26.457804.

9. Peulen, T. O., Hengstenberg, C. S., Biehl, R., Dimura, M., Lorenz, C., Valeri, A., et al. Integrative dynamic structural biology unveils conformers essential for the oligomerization of a large GTPase. *arXiv: Biological Physics*, 2020. DOI: arXiv:2004.04229.

10. Sistemich, L., Kutsch, M., Hämisch, B., Zhang, P., Shydlovskyi, S., Britzen-Laurent, N., et al. The molecular mechanism of polymer formation of farnesylated human

guanylate-binding protein 1. *Journal of Molecular Biology*, 2020, **432**, 2164-2185. DOI: 10.1016/j.jmb.2020.02.009.

11. Dimura, M., Peulen, T. O., Hanke, C. A., Prakash, A., Gohlke, H., Seidel, C. A. M. Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. *Current Opinion in Structural Biology*, 2016, **40**, 163-185. DOI: 10.1016/j.sbi.2016.11.012.

12. Henzler-Wildman, K., Kern, D. Dynamic personalities of proteins. *Nature*, 2007, **450**, 964-972. DOI: 10.1038/nature06522.

13. Gansen, A., Felekyan, S., Kühnemuth, R., Lehmann, K., Toth, K., Seidel, C. A. M., et al. High precision FRET studies reveal reversible transitions in nucleosomes between microseconds and minutes. *Nature Communications*, 2018, **9**. DOI: 10.1038/s41467-018-06758-1.

14. Sanabria, H., Rodnin, D., Hemmen, K., Peulen, T. O., Felekyan, S., Fleissner, M. R., et al. Resolving dynamics and function of transient states in single enzyme molecules. *Nature Communications*, 2020, **11**. DOI: 10.1038/s41467-020-14886-w.

15. Kilic, S., Felekyan, S., Doroshenko, O., Boichenko, I., Dimura, M., Vardanyan, H., et al. Single-molecule FRET reveals multiscale chromatin dynamics modulated by HP1 alpha. *Nature Communications*, 2018, **9**. DOI: 10.1038/s41467-017-02619-5.

16. Kunzelmann, S., Praefcke, G. J. K., Herrmann, C. Transient kinetic investigation of GTP hydrolysis catalyzed by interferon-gamma-induced hGBP1 (human guanylate binding protein 1). *Journal of Biological Chemistry*, 2006, **281**, 28627-28635. DOI: 10.1074/jbc.M604911200.

17. Lakowicz, J. R. Principles of fluorescence spectroscopy. 3rd ed. 2006: Springer.

18. Holden, S. J., Uphoff, S., Hohlbein, J., Yadin, D., Le Reste, L., Britton, O. J., et al. Defining the limits of single-molecule FRET resolution in TIRF microscopy. *Biophysical Journal*, 2010, **99**, 3102-3111. DOI: 10.1016/j.bpj.2010.09.005.

19. Becker, W. The bh TCSPC handbook. 8th ed. 2019: Becker & Hickl GmbH.

20. Hengstenberg, C. S., Structural dynamics and implications for dimer formation of human guanylate-binding protein 1, *Physical and biophysical chemistry*. 2013, Ruhr University: Bochum.

21. Müller, B. K., Zaychikov, E., Bräuchle, C., Lamb, D. C. Pulsed interleaved excitation. *Biophysical Journal*, 2005, **89**, 3508-3522. DOI: 10.1529/biophysj.105.064766.

22. Lakowicz, J. R., Gryczynski, I., Wiczk, W., Laczko, G., Prendergast, F. C., Johnson, M. L. Conformational distributions of melittin in water-methanol mixtures from frequency-domain measurements of nonradiative energy-transfer. *Biophysical Chemistry*, 1990, **36**, 99-115. DOI: 10.1016/0301-4622(90)85014-w.

23. Hellenkamp, B., Schmid, S., Doroshenko, O., Opanasyuk, O., Kühnemuth, R., Adariani, S. R., et al. Precision and accuracy of single-molecule FRET measurements-a multi-laboratory benchmark study. *Nature Methods*, 2018, **15**, 669-+. DOI: 10.1038/s41592-018-0085-0.

24. Cui, W., Braun, E., Wang, W., Tang, J. H., Zheng, Y. Y., Slater, B., et al. Structural basis for GTP-induced dimerization and antiviral function of guanylate-binding proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 2021, **118**. DOI: 10.1073/pnas.2022269118.

25. Phillips, D., Drake, R. C., O'Connor, D. V., Christensen, R. L. Time correlated single-photon counting (TCSPC) using laser excitation. *Analytical Instrumentation*, 1985, **14**, 267-292. DOI: 10.1080/10739148508543581.

Chapter 2 Supplement

Chapter 2-Written supplement 1: DATA FINDER

A) An operation manual for the TCSPC-stopped flow is available at:

:\Measurements@New_Server\Stopped Flow\Operation manual stopped flow setup.pdf

B) Raw data from TCSPC-SF is available at:

:\Measurements@New_Server\Stopped Flow\2017\hGBP1 measurements…

… 7_20_17 for (344D/A, 525D/A)

… 8_08_17 for 18D + 18A

… 8_14_17 for (344D/A, 540D/A)

… 8_15_17 for (254D/A, 540D/A)

… 8_21_17 for (254D/A, 344D/A)

… 8_24_17 for 344D + 481A

C) Kinetic model fit files (ASCII/txt) are available at:

:\KochJ\Thesis\hGBP1 story\Kinetic models

D) ChiSurf fit files from TCSPC-SF is available at:

:\Measurements@New_Server\Stopped Flow\2017\hGBP1 measurements\Fits and processed data

E) Images and text used in this work is available at:

:\KochJ\Thesis\hGBP1 story

Chapter 2-Written supplement 2: Full solution for multistate exchange in the presence of a recombination

Let us consider the kinetic scheme consisting of initial irreversible recombination step ($2A \overset{k_r}{\rightarrow} D$) followed by multistate dimer (D) exchange ($B_1 \rightarrow B_2 \rightarrow C$)

$$2A \overset{k_r}{\rightarrow} \overbrace{\boxed{D_1 \leftrightarrow D_2 \leftrightarrow \dots}}^{D} \qquad (14)$$

The corresponding system of ODE describing the kinetic of reagent concentration can be written as:

$$\begin{cases} c_A'(t) = -k_r\, c_A^2(t) \\ 2c_D'(t) = k_r\, c_A^2(t)x_{D,r} + 2\,\mathbf{K}_D\, c_D(t)\,' \end{cases} \qquad (15)$$

where c_A is concentration of monomers, $c_D = (c_{D1}, c_{D2}, \dots)^T$ is column-vector consisting of concentrations of dimer in different states, $x_{D,r}$ is vectors of probabilities that recombination

lead to the particular dimer state, k_r is the specific recombination rate constant and $\mathbf{K}_D$ is the transition rate matrix describing exchange between dimer states.

We assume that at the beginning all molecules are not dimerized and the concentration of monomers $c_A(t = 0) = c_{A0}$. We can rewrite Eq. 16 in the form for fractions of monomer units in all reagents:

$$\begin{cases} x_A'(t) = -k_{r,0}\, x_A^2(t) \\ \boldsymbol{x}_D'(t) = k_{r,0}\, x_A^2(t)\, \boldsymbol{x}_{D,r} + \mathbf{K}_D\, \boldsymbol{x}_D(t) \end{cases}, \text{where} \begin{cases} x_A(t) = c_A(t)/c_{A0} \\ \boldsymbol{x}_D(t) = 2c_D'(t)/c_{A0} \\ k_{r,0} = c_{A0}\, k_r \end{cases} \tag{16}$$

The solution should fulfil initial and conservation conditions:

$$\begin{cases} x_A(0) = 1 \\ \boldsymbol{x}_D(0) = \mathbf{0} \end{cases}; \; x_A(t) + \mathbf{1}^T \cdot \boldsymbol{x}_D(t) = 1 \tag{17}$$

The equation for monomers $x_A(t)$ can be solved separately:

$$x_A(t) = \frac{1}{k_{r,0}\, t + 1} \tag{18}$$

Solution for $\boldsymbol{x}_D(t)$

For shortness, let us rewrite Eq. 16 in the form:

$$\boldsymbol{x}_D'(t) = r(t)\boldsymbol{x}_{D,r} + \mathbf{K}_D\, \boldsymbol{x}_D(t); \quad r(t) = -x_A'(t) = \frac{k_{r,0}}{\left(k_{r,0}\, t + 1\right)^2} \tag{19}$$

This is the first order non-homogeneous ODE which can be solved by the integrating factor method. Let us multiply both side of equation by factor $e^{-\mathbf{K}_D t}$ then, move the term with $e^{-\mathbf{K}_D t}\mathbf{K}_D\, \boldsymbol{x}_D(t)$ to the left-hand side:

$$e^{-\mathbf{K}_D t}\boldsymbol{x}_D'(t) - \mathbf{K}_D e^{-\mathbf{K}_D t}\, \boldsymbol{x}_D(t) = r(t)e^{-\mathbf{K}_D t}\boldsymbol{x}_{D,r} \tag{20}$$

Now the left-hand side can be written as derivative of the product $e^{-\mathbf{K}_D t}\, \boldsymbol{x}_D(t)$, and we get:

$$\left(e^{-\mathbf{K}_D t}\, \boldsymbol{x}_D(t)\right)' = r(t)e^{-\mathbf{K}_D t}\boldsymbol{x}_{D,r} \tag{21}$$

Now the right-hand side is the antiderivative of $e^{-\mathbf{K}_D t}\, \boldsymbol{x}_D(t)$. So, applying the Fundamental Theorem of Calculus we can write:

$$e^{-\mathbf{K}_D t}\, \boldsymbol{x}_D(t) = \int_0^t r(\tau)e^{-\mathbf{K}_D \tau}\boldsymbol{x}_{D,r}d\tau + \boldsymbol{const} \tag{22}$$

Now we can apply the initial condition $x_D(0) = 0$ and get $\mathbf{const} = 0$. At last step we multiply both sides by $e^{\mathbf{K}_D\, t}$ and get the solution:

$$x_D(t) = \int_0^t r(\tau)e^{-\mathbf{K}_D\,(t-\tau)}x_{D,r}d\tau \qquad (23)$$

Note that the factor $e^{-\mathbf{K}_D\,t}x_{D,r}$ is the solution of the complementary (c) homogeneous ODE:

$$x_D^{(c)'}(t) = \mathbf{K}_D\, x_D^{(c)}(t); \quad x_D^{(c)}(0) = x_{D,r} \qquad (24)$$

Which describe the kinetic of dimers ($B_1 \to B_2 \to C$) in the absence of recombination and with initial fractions equal to $x_{D,r}$.

So, the formal solution for the monomer fractions in the dimer states is the convolution ($*$):

$$\boxed{x_D(t) = -x_A'(t) * x_D^{(c)}(t)} = -\int_0^t x_A'(\tau)e^{\mathbf{K}_D(t-\tau)}x_{D,r}\ d\tau$$

$$x_D^{(c)}(t) = e^{\mathbf{K}_D\,t}x_{D,r} \qquad , \qquad (25)$$

$$x_A'(t) = -\frac{k_{r,0}}{(k_{r,0}\,t + 1)^2}$$

The matrix exponential $e^{\mathbf{K}_D\,t}$ can be calculated using eigen value decomposition (EVD) of $\mathbf{K}_D$ matrix. This gives the next expression for $x_D^{(c)}(t)$:

$$x_D^{(c)}(t) = \sum_{i=0}^{n-1} a_i e^{-k_i t}, \qquad a_i = u_i(v_i \cdot x_{D,r}) \qquad (26)$$

where n is the number of dimer states, k_i are negated eigen-values of $\mathbf{K}_D$ and u_i, v_i are corresponding left and right eigen-vectors.

Then the whole solution in the presence of a recombination is:

$$\left\{ \begin{aligned} &x_A(k_{r,0}, t) = \frac{1}{k_{r,0}\,t + 1} \\ &x_D(t) = \sum_{i=0}^{n-1} a_i f_r(k_i, k_{r,0}, t) \\ &a_i = u_i(v_i \cdot x_{D,r}) \\ &f_r(k, k_0, t) \overset{\text{def}}{=} -x_A'(k_0, t) * e^{-kt} \end{aligned} \right. \qquad (27)$$

Derivation of function $f_r(k, k_0, t)$

First, we can use integration by parts and rewrite the convolution in the form involving $x_A(t)$ fraction instead of its derivative:

$$f_r(k, k_0, t) = -x_A'(k_0, t) * e^{-kt} = -x_A(k_0, t) + e^{-kt} + k(x_A(k_0, t) * e^{-kt})$$

$$f_r(k, k_0, t) = -\frac{1}{k_0 t + 1} + e^{-kt} + k\left(\frac{1}{k_0 t + 1} * e^{-kt}\right)$$

$$k\left(\frac{1}{k_0 t + 1} * e^{-kt}\right) = k \underbrace{\int_0^t \frac{1}{1 + k_0 \tau} e^{-k(t-\tau)} d\tau}_{\substack{v = k\tau + \frac{k}{k_0} \\ \tau = \frac{v}{k} - \frac{1}{k_0} \\ d\tau = \frac{1}{k} dv}} = \frac{k}{k_0} e^{-\left(\frac{k}{k_0} + kt\right)} \underbrace{\int_{\frac{k}{k_0}}^{\frac{k}{k_0} + kt} \frac{e^v}{v} dv}_{Ei\left(\frac{k}{k_0} + kt\right) - Ei\left(\frac{k}{k_0}\right)} \tag{28}$$

This results in (indices harmonized to Eq. 1-5 as $k_0 = k_a$; $k = k_i$; $i \in \{b, c\}$)

$$\boxed{f_r(k_i, k_a, t) \overset{\text{def}}{=} -x_A'(k_a, t) * e^{-k_i t} = -\frac{1}{k_a t + 1} + e^{-k_i t} + \frac{k_i}{k_a} e^{-\left(\frac{k_i}{k_a} + k_i t\right)}\left(Ei\left(\frac{k_i}{k_a} + k_i t\right) - Ei\left(\frac{k_i}{k_a}\right)\right)} \tag{29}$$

where $Ei(x)$ is the exponential integral function:

$$Ei(x) = \int_{-\infty}^x \frac{e^v}{v} dv \tag{30}$$

The fitting of the stop-flow data presented in this work was done using the numerical EVD and Eq. 29 above. However, for simple exchange schemes the relaxation rates k_i and amplitudes vectors a_i can be calculated analytically. Thus, for the 3-state dimer with $x_{D,r} = (1,0,0)^T$ (in this work we assumed that monomers always recombinate into the state B_1) and exchange scheme $B_1 \overset{k_b}{\to} B_2 \overset{k_c}{\to} C$ they are:

$$k_0 = 0; \qquad k_1 = k_b; \qquad k_2 = k_c;$$

$$a_0 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}; \quad a_1 = \begin{pmatrix} \frac{1}{k_b} \\ -\frac{k_b}{k_b - k_c} \\ \frac{k_c}{k_b - k_c} \end{pmatrix}; \quad a_2 = \begin{pmatrix} 0 \\ \frac{k_b}{k_b - k_c} \\ -\frac{k_b}{k_b - k_c} \end{pmatrix}; \tag{31}$$

So, the dimer part of solution is:

$$\text{3 state dimer: } x_D(t) = \begin{pmatrix} x_{B1}(t) \\ x_{B2}(t) \\ x_C(t) \end{pmatrix} = \begin{cases} f_r(k_b, k_{r,0}, t) \\ \dfrac{k_b}{k_b - k_c}\left(f_r(k_c, k_{r,0}, t) - f_r(k_b, k_{r,0}, t)\right) \\ 1 - \dfrac{1}{1 + k_a t} - \dfrac{k_b}{k_b - k_c} f_r(k_c, k_{r,0}, t) + \dfrac{k_c}{k_b - k_c} f_r(k_b, k_{r,0}, t) \end{cases} \tag{32}$$

The expression for the 2-state scheme $B \xrightarrow{k_c} C$ can be obtained from 3-state solution above by setting $k_b \to \infty, k_c = k$ and $x_B = x_{B1}(t) + x_{B2}(t)$:

$$\text{2 state dimer: } x_D(t) = \begin{pmatrix} x_B(t) \\ x_C(t) \end{pmatrix} = \begin{cases} f_r(k, k_{r,0}, t) \\ 1 - \dfrac{1}{1 + k_a t} - f_r(k, k_{r,0}, t) \end{cases} \tag{33}$$

Chapter 2-Written supplement 3: Time-dependency of k$_a$

In general, a specific reaction rate is time-dependent: $k_a = k_a(t)$. We assumed a time-independent k_a to make our solution possible. This assumption will be validated in the following supplemental part. We can assume that the dimerization kinetics follows the Smoluchowski theory. This means, that the dimerization followed the diffusion-controlled collision happens with probability $p_{dim} = 1$. Then the Smoluchovski theory predicts that the specific reaction rate is:

$$k_a(t) = 4\pi D R N_A \left(1 + \frac{R}{\sqrt{\pi D t}}\right) \tag{34}$$

Then, the condition for k_a = constant is:

$$\frac{R}{\sqrt{\pi D t}} \ll 1 \quad \Rightarrow \quad t \gg \frac{R^2}{\pi D} \tag{35}$$

Using the Einstein diffusion relation

$$ED = \frac{k_B T}{6\pi \eta R_M} \tag{36}$$

where R_M is molecular radius. So we can write:

$$t \gg \frac{6 \eta R^2 R_M}{k_B T} \tag{37}$$

Then, taking $R_M \approx 10^{-8} m$, $R \approx 2R_m$, $\eta = 10^{-3} \frac{kg}{m\,s}$, $T = 300K$, we get:

$$t \gg \frac{6\,\eta\,R^2 R_M}{k_B T} \approx \frac{6 \times 4 \times 10^{-3} \times 10^{-24}}{300 \times 1.38 \times 10^{-23}}\,s \approx 10^{-5}s \tag{38}$$

In our measurements the smallest time is $\Delta t = 10^{-2}s \gg 10^{-5}s$, this validates our assumption that k_a can be treated time-independently.

Chapter 2-Table supplement 1: Inter-FRET - 2ABC fit results (red photons)

The result of joint fitting of inter-FRET pairs 18D + 18A, 344D + 481A with 2ABC model

Inter-FRET data was not weighted with corresponding donor brightness proportional to efficiencies since, therefore no distances were fitted. "init" = starting value.

Parameter	Value
Reduced chi-square	1.0284
Akaike information criterion	174.1894
Bayesian information criterion	214.3845
kb	1.5691e-04 +/- 3.1972e-06 (2.04%) (init = 0.0001392)
r (kb/ka)	0.0572 +/- 0.006604 (11.54%) (init = 0.01)
Offset 18D + 18A	4132.5404 +/- 22.493831 (0.54%) (init = 3900)
Scaling 18D + 18A	449.5134 +/- 22.084485 (4.91%) (init = 600)
Offset 344D + 481A	10354.6911 +/- 8.280897 (0.08%) (init = 10300)
Scaling 344D + 481A	1337.5482 +/- 7.776847 (0.58%) (init = 1350)
1/kb	6372.9395 +/- 130.009200 ms
1/r (ka/kb)	17.4813 +/- 2.017342 ms
1/ka	364.5568 +/- 49.506813 ms

Chapter 2-Table supplement 2: Intra-FRET - 2ABC fit results (red photons)

The result of joint fitting of intra-FRET pairs (254D/A, 344D/A); (254D/A, 540D/A); (344D/A, 540D/A) with 2ABC model)

Large uncertainties for distances and offset not shown since E has not enough restrictions. Distances were only used to stabilize the fitting of rates and are not explicit.

Parameter	Value
Reduced chi-square	1.1726
Akaike information criterion	1446.7878
Bayesian information criterion	1560.4408
ka (fixed to ka from inter-FRET)	0.0027 (fixed)
kb	0 (fixed)
kc	2.3436e-04 +/- 4.0029e-06 (1.71%) (init = 0.000126)
Offset (254D/A, 344D/A)	966.7315 (init = 1120)
Scaling (254D/A, 344D/A)	2420.9098 (init = 2528)
R1 (state A) (254D/A, 344D/A)	75.3306 (init = 73)
R2 (state B) (254D/A, 344D/A)	87.8194 (init = 82)

R3 (state C) (254D/A, 344D/A)	95.0998 (init = 89)
Offset (254D/A, 540D/A)	2191.6570 (init = 2183)
Scaling (254D/A, 540D/A)	3112.7710 (init = 3828)
R1 (state A) (254D/A, 540D/A)	49.3747 (init = 52)
R2 (state B) (254D/A, 540D/A)	69.7142 (init = 69)
R3 (state C) (254D/A, 540D/A)	120.6849 (init = 150)
Offset (344D/A, 540D/A)	1060.7583 (init = 1130)
Scaling (344D/A, 540D/A)	3012.7347 (init = 3357)
R1 (state A) (344D/A, 540D/A)	28.3349 (init = 44)
R2 (state B) (344D/A, 540D/A)	70.6606 (init = 77)
R3 (state C) (344D/A, 540D/A)	91.9179 (init = 123)

Chapter 2-Table supplement 3: Intra-FRET - 2ABBC fit results (red photons)

(The result of joint fitting of intra-FRET pairs (254D/A, 344D/A); (254D/A, 540D/A); (344D/A, 540D/A) with 2ABBC model)

Large uncertainties for distances and offset not shown since E has not enough restrictions. Distances were only used to stabilize the fitting of rates and are not explicit. Fractions x2, x3, x4 were added and fitted globally to stabilize the fit further.

Parameter	Value
Reduced chi-square	**1.0171**
Akaike information criterion	173.7358
Bayesian information criterion	322.9054
ka (fixed to ka from inter-FRET)	**0.0027 (fixed)**
kb	**0.0012 +/- 4.4789e-05 (3.75%) (init = 0.0015)**
kc	**1.1516e-04 +/- 4.6280e-06 (4.02%) (init = 0.000126)**
Offset (254D/A, 344D/A)	1128.28105 (init = 1120)
Scaling (254D/A, 344D/A)	2537.38576 (init = 2528)
R1 (state A) (254D/A, 344D/A)	66.6258388 (init = 73)
R2 (state B1) (254D/A, 344D/A)	68.5952036 (init = 69)
R3 (state B2) (254D/A, 344D/A)	82.4853213 (init = 82)
R4 (state C) (254D/A, 344D/A)	89.6373952 (init = 89)
Offset (254D/A, 540D/A)	2180.07871 (init = 2183)
Scaling (254D/A, 540D/A)	3920.44226 (init = 3828)
R1 (state A) (254D/A, 540D/A)	41.2189782 (init = 52)
R2 (state B1) (254D/A, 540D/A)	52.4469301 (init = 51)
R3 (state B2) (254D/A, 540D/A)	70.4076247 (init = 69)
R4 (state C) (254D/A, 344D/A)	168.997720 (init = 150)
Offset (344D/A, 540D/A)	2180.07871 (init = 2183)
Scaling (344D/A, 540D/A)	3450.72683 (init = 3357)
R1 (state A) (344D/A, 540D/A)	35.0241359 (init = 44)
R2 (state B1) (344D/A, 540D/A)	51.1380156 (init = 50)
R3 (state B2) (344D/A, 540D/A)	79.9306324 (init = 77)
R4 (state C) (254D/A, 344D/A)	119.667322 (init = 123)
x2 (global)	0.7864 (init = 0.5)
x3 (global)	0.21350191 (init == '1-x1')
x4 (global)	0 (fixed)

Chapter 2-Table supplement 4: 18D-18A inter-FRET

Parameters which are fitted are given in **bold**, linked parameters in *italic*. fixed parameters are plain text.

Time-point [ms]	55	125	195	255	305	625	955	3065	7997	20001
DOnly	**0.992**	**0.953**	**0.874**	**0.887**	**0.810**	**0.642**	**0.586**	**0.417**	**0.281**	**0.295**
R(G,1)	*62.000*	*62.000*	*62.000*	*62.000*	*62.000*	*62.000*	*62.000*	*62.000*	*62.000*	*62.000*
R(G,2)	*45.000*	*45.000*	*45.000*	*45.000*	*45.000*	*45.000*	*45.000*	*45.000*	*45.000*	*45.000*
R0	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000
bg	**0.189**	**0.709**	**0.940**	**0.346**	**0.000**	**0.252**	**0.397**	**0.792**	**0.542**	**0.083**
dt	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117
g	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
k(G,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
k(G,2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
k2	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667
kappa2	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667
l1	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031
l2	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037
lb	**78.262**	**99.999**	**1.521**	**99.998**	**100.000**	**15.219**	**99.999**	**99.573**	**99.270**	**99.590**
n0	6943.128	6819.967	6942.008	6876.990	6912.758	6992.679	6987.664	6966.846	7173.251	7060.908
r0	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380
rep	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000
s(G,1)	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
s(G,2)	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
sc	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
start	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
stop	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991
t0	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000
tBg	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
tDead	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000
tL1	*3.986*	*3.986*	*3.986*	*3.986*	*3.986*	*3.986*	*3.986*	*3.986*	*3.986*	*3.986*
tL2	*0.234*	*0.234*	*0.234*	*0.234*	*0.234*	*0.234*	*0.234*	*0.234*	*0.234*	*0.234*
tL3	*1.926*	*1.926*	*1.926*	*1.926*	*1.926*	*1.926*	*1.926*	*1.926*	*1.926*	*1.926*
tMeas	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ts	-0.006	-0.006	-0.006	-0.006	-0.006	-0.006	-0.006	-0.006	-0.006	-0.006
win-size	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000
x(G,1)	0.860	*0.860*	*0.860*	*0.860*	*0.860*	0.860	*0.860*	*0.860*	*0.860*	*0.860*
x(G,2)	0.140	*0.140*	*0.140*	*0.140*	*0.140*	0.140	*0.140*	*0.140*	*0.140*	*0.140*
xL1	0.747	*0.747*	*0.747*	*0.747*	*0.747*	0.747	*0.747*	*0.747*	*0.747*	*0.747*
xL2	0.114	*0.114*	*0.114*	*0.114*	*0.114*	0.114	*0.114*	*0.114*	*0.114*	*0.114*
xL3	0.139	*0.139*	*0.139*	*0.139*	*0.139*	0.139	*0.139*	*0.139*	*0.139*	*0.139*
Chi2r	1.0803	1.0571	1.0328	1.0180	1.2256	1.0054	0.8934	1.1330	0.8359	0.9589

Chapter 2-Table supplement 5: 344D-481A inter-FRET

Parameters which are fitted are given in **bold**, linked parameters in *italic*. fixed parameters are plain text.

Time-point [ms]	55	125	195	255	305	625	955	3065	7997	20001
DOnly	**0.952**	**0.983**	**0.951**	**0.922**	**0.953**	**0.941**	**0.735**	**0.609**	**0.497**	**0.379**
R(G,1)	*58.500*	*58.500*	*58.500*	*58.500*	*58.500*	*58.500*	*58.500*	*58.500*	*58.500*	*58.500*
R(G,2)	*41.000*	*41.000*	*41.000*	*41.000*	*41.000*	*41.000*	*41.000*	*41.000*	*41.000*	*41.000*
R0	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000
bg	**0.642**	**0.792**	**0.543**	**0.583**	**0.850**	**0.379**	**1.023**	**0.978**	**0.692**	**0.040**
dt	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117
g	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
k(G,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
k(G,2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
k2	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667
kappa2	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667
l1	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031
l2	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037
lb	**99.986**	**99.913**	**3.046**	**99.999**	**99.899**	**100.000**	**100.000**	**99.999**	**100.000**	**100.000**
n0	15082.398	14742.786	14921.011	15254.903	14982.091	14864.790	14525.629	14712.883	14642.859	14878.717
r0	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380
rep	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000
s(G,1)	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
s(G,2)	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
sc	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
start	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
stop	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991
t0	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000
tBg	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
tDead	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000
tL1	*4.090*	*4.090*	*4.090*	*4.090*	*4.090*	*4.090*	*4.090*	*4.090*	*4.090*	*4.090*
tL2	*2.070*	*2.070*	*2.070*	*2.070*	*2.070*	**2.070**	**4.000**	**4.000**	**4.000**	**4.000**
tL3	*0.290*	*0.290*	*0.290*	*0.290*	*0.290*	**0.290**	**4.000**	**4.000**	**4.000**	**4.000**
tMeas	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ts	*0.071*	*0.071*	*0.071*	*0.071*	*0.071*	*0.071*	*0.071*	*0.071*	*0.071*	*0.071*
win-size	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000
x(G,1)	*0.626*	*0.626*	*0.626*	*0.626*	*0.626*	*0.626*	*0.626*	*0.626*	*0.626*	*0.626*
x(G,2)	*0.374*	*0.374*	*0.374*	*0.374*	*0.374*	*0.374*	*0.374*	*0.374*	*0.374*	*0.374*
xL1	*0.815*	*0.815*	*0.815*	*0.815*	*0.815*	*0.815*	*0.815*	*0.815*	*0.815*	*0.815*
xL2	*0.103*	*0.103*	*0.103*	*0.103*	*0.103*	*0.103*	*0.103*	*0.103*	*0.103*	*0.103*
xL3	*0.082*	*0.082*	*0.082*	*0.082*	*0.082*	*0.082*	*0.082*	*0.082*	*0.082*	*0.082*
Chi2r	0.9242	0.9883	1.0799	0.9422	1.0059	1.0683	1.0704	0.9246	0.9633	1.2172

Chapter 2-Table supplement 6: 254-344 intra-FRET

Parameters which are fitted are given in **bold**, linked parameters in *italic*. fixed parameters are plain text.

Time-point [ms]	55	125	195	255	305	625	955	3065	7997	20001
DOnly	*0.170*	*0.170*	*0.170*	*0.170*	*0.170*	*0.170*	*0.170*	*0.170*	*0.170*	*0.170*
R(G,1)	*71.900*	*71.900*	*71.900*	*71.900*	*71.900*	*71.900*	*71.900*	*71.900*	*71.900*	*71.900*
R(G,2)	*62.500*	*62.500*	*62.500*	*62.500*	*62.500*	*62.500*	*62.500*	*62.500*	*62.500*	*62.500*
R(G,3)	*100.00*	*100.000*	*100.000*	*100.000*	*100.000*	*100.000*	*100.000*	*100.000*	*100.000*	*100.000*
R0	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000
bg	**0.001**	**0.000**	**0.006**	**0.000**	**0.218**	**0.566**	**0.000**	**0.000**	**0.003**	**0.000**
dt	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117
g	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
k(G,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
k(G,2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
k(G,3)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
k2	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667
kappa2	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667
l1	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031
l2	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037
lb	**99.299**	**96.896**	**99.783**	**100.000**	**100.000**	**99.777**	**100.000**	**89.887**	**100.000**	**99.774**
n0	40822.704	41158.799	41320.668	41388.856	40910.026	40909.216	41282.146	41307.180	41398.608	41414.720
r0	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380
rep	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000
s(G,1)	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
s(G,2)	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
s(G,3)	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
sc	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
start	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
stop	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991
t0	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000
tBg	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
tDead	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000
tL1	*4.300*	*4.300*	*4.300*	*4.300*	*4.300*	*4.300*	*4.300*	*4.300*	*4.300*	*4.300*
tMeas	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ts	*-0.459*	*-0.459*	*-0.459*	*-0.459*	*-0.459*	*-0.459*	*-0.459*	*-0.459*	*-0.459*	*-0.459*
win-size	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000
x(G,1)	**0.646**	**0.477**	**0.515**	**0.546**	**0.622**	**0.797**	**0.893**	**0.802**	**0.642**	**0.555**
x(G,2)	**0.354**	**0.523**	**0.485**	**0.454**	**0.378**	**0.203**	**0.087**	**0.000**	**0.000**	**0.000**
x(G,3)	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.020**	**0.198**	**0.358**	**0.445**
xL1	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*
Chi2r	1.0444	1.0531	1.2812	1.3305	1.3645	1.2042	1.1729	1.1244	1.2904	1.2097

Chapter 2-Table supplement 7: 254-540 intra-FRET

Parameters which are fitted are given in **bold**, linked parameters in *italic*. fixed parameters are plain text.

Time-point [ms]	55	125	195	255	305	625	955	3065	7997	20001
DOnly	0.150	*0.150*	*0.150*	*0.150*	*0.150*	0.150	*0.150*	*0.150*	*0.150*	*0.150*
R(G,1)	51.800	*51.800*	*51.800*	*51.800*	*51.800*	51.800	*51.800*	*51.800*	*51.800*	*51.800*
R(G,2)	31.500	*31.500*	*31.500*	*31.500*	*31.500*	31.500	*31.500*	*31.500*	*31.500*	*31.500*
R(G,3)	100.00	*100.000*	*100.000*	*100.000*	*100.000*	100.000	*100.000*	*100.000*	*100.000*	*100.000*
R0	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000
bg	**0.560**	**0.682**	**0.711**	**1.056**	**0.775**	**0.287**	**0.152**	**0.873**	**0.818**	**1.071**
dt	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117
g	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
k(G,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
k(G,2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
k(G,3)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
k2	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667
kappa2	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667
l1	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031
l2	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037
lb	**1.462**	**99.996**	**99.050**	**1.921**	**99.992**	**99.980**	**35.796**	**79.950**	**98.279**	**100.000**
n0	37889.430	42119.081	41451.067	42344.931	41765.892	41545.073	41752.160	42023.241	41830.009	40940.836
r0	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380
rep	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000
s(G,1)	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
s(G,2)	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
s(G,3)	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
sc	0.000	*0.000*	*0.000*	*0.000*	*0.000*	0.000	*0.000*	*0.000*	*0.000*	*0.000*
start	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
stop	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991
t0	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000
tBg	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
tDead	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000
tL1	4.000	*4.000*	*4.000*	*4.000*	*4.000*	4.200	*4.200*	*4.200*	*4.200*	*4.200*
tMeas	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ts	0.025	*0.025*	*0.025*	*0.025*	*0.025*	0.025	*0.025*	*0.025*	*0.025*	*0.025*
win-size	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000
x(G,1)	**0.568**	**0.447**	**0.445**	**0.422**	**0.396**	**0.391**	**0.333**	**0.240**	**0.148**	**0.102**
x(G,2)	**0.363**	**0.457**	**0.434**	**0.442**	**0.436**	**0.401**	**0.394**	**0.348**	**0.339**	**0.311**
x(G,3)	**0.069**	**0.096**	**0.121**	**0.135**	**0.168**	**0.207**	**0.273**	**0.412**	**0.513**	**0.586**
xL1	1.000	*1.000*	*1.000*	*1.000*	*1.000*	1.000	*1.000*	*1.000*	*1.000*	*1.000*
Chi2r	1.0638	1.1709	1.2223	1.1625	1.3257	1.2676	0.9531	1.0100	1.0623	1.3552

Chapter 2-Table supplement 8: 344-525 intra-FRET

Parameters which are fitted are given in **bold**, linked parameters in *italic*. fixed parameters are plain text.

Time-point [ms]	55	125	195	255	305	625	955	3065	7997	20001
DOnly	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
R(G,1)	*26.500*	*26.500*	*26.500*	*26.500*	*26.500*	*26.500*	*26.500*	*26.500*	*26.500*	*26.500*
R(G,2)	*45.900*	*45.900*	*45.900*	*45.900*	*45.900*	*45.900*	*45.900*	*45.900*	*45.900*	*45.900*
R(G,3)	*100.00*	*100.000*	*100.000*	*100.000*	*100.000*	*100.000*	*100.000*	*100.000*	*100.000*	*100.000*
R0	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000
bg	**0.000**	**0.000**	**0.239**	**0.143**	**0.000**	**0.000**	**0.482**	**0.000**	**0.435**	**0.040**
dt	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117
g	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
k(G,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
k(G,2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
k(G,3)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
k2	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667
kappa2	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667
l1	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031
l2	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037
lb	**100.00**	**100.000**	**1.000**	**1.000**	**35.148**	**100.000**	**100.000**	**100.000**	**100.000**	**100.000**
n0	31544.284	38029.797	36312.795	34560.211	36126.284	32791.667	38074.174	35783.291	33552.940	33577.342
r0	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380
rep	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000
s(G,1)	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
s(G,2)	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
s(G,3)	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
sc	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
start	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
stop	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991
t0	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000
tBg	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
tDead	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000
tL1	*3.750*	*3.750*	*3.750*	*3.750*	*3.750*	*4.080*	*4.080*	*4.080*	*4.080*	*4.080*
tMeas	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ts	-0.533	-0.533	-0.533	-0.533	-0.533	-0.533	-0.533	-0.533	-0.533	-0.533
win-size	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000
x(G,1)	**0.478**	**0.576**	**0.531**	**0.501**	**0.509**	**0.411**	**0.477**	**0.400**	**0.356**	**0.337**
x(G,2)	**0.495**	**0.357**	**0.342**	**0.314**	**0.282**	**0.239**	**0.156**	**0.077**	**0.032**	**0.040**
x(G,3)	**0.027**	**0.067**	**0.127**	**0.185**	**0.208**	**0.350**	**0.367**	**0.522**	**0.612**	**0.624**
xL1	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*
Chi2r	1.0542	0.9248	0.8590	1.2003	1.3341	1.1239	1.3417	1.1760	1.2539	0.9917

Chapter 2-Table supplement 9: 344-540 intra-FRET

Parameters which are fitted are given in **bold**, linked parameters in *italic*. fixed parameters are plain text.

Time-point [ms]	55	125	195	255	305	625	955	3065	7997	20001
DOnly	**0.017**	**0.041**	**0.053**	**0.000**	**0.000**	**0.100**	**0.000**	**0.000**	**0.000**	**0.000**
R(G,1)	*42.200*	*42.200*	*42.200*	*42.200*	*42.200*	*42.200*	*42.200*	*42.200*	*42.200*	*42.200*
R(G,2)	*57.900*	*57.900*	*57.900*	*57.900*	*57.900*	*57.900*	*57.900*	*57.900*	*57.900*	*57.900*
R(G,3)	*100.00*	*100.000*	*100.000*	*100.000*	*100.000*	*100.000*	*100.000*	*100.000*	*100.000*	*100.000*
R0	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000	52.000
bg	**0.677**	**0.717**	**0.484**	**1.000**	**0.487**	**0.000**	**0.787**	**0.517**	**1.104**	**0.626**
dt	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117
g	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
k(G,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
k(G,2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
k(G,3)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
k2	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667
kappa2	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667
l1	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031
l2	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037
lb	**99.995**	**100.000**	**1.000**	**99.999**	**1.000**	**16.620**	**100.000**	**100.000**	**100.000**	**100.000**
n0	22510.944	23079.933	22635.402	23020.122	22530.216	23112.289	22795.211	22768.240	22350.947	22557.932
r0	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380
rep	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000
s(G,1)	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
s(G,2)	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
s(G,3)	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
sc	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
start	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
stop	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991	59.991
t0	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000
tBg	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
tDead	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000
tL1	3.900	3.900	3.900	3.900	3.900	4.200	4.200	4.200	4.200	4.200
tMeas	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ts	-0.009	-0.009	-0.009	-0.009	-0.009	-0.009	-0.009	-0.009	-0.009	-0.009
win-size	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000	17.000
x(G,1)	**0.256**	**0.284**	**0.267**	**0.261**	**0.232**	**0.270**	**0.198**	**0.137**	**0.070**	**0.076**
x(G,2)	**0.738**	**0.713**	**0.677**	**0.584**	**0.571**	**0.515**	**0.395**	**0.187**	**0.117**	**0.066**
x(G,3)	**0.006**	**0.003**	**0.056**	**0.156**	**0.198**	**0.215**	**0.407**	**0.676**	**0.813**	**0.859**
xL1	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*
Chi2r	1.0409	0.9907	0.9040	1.2031	1.2590	1.1727	1.2256	0.8428	0.9345	0.9347

Chapter 2-Table supplement 10: Exemplary fits 1-exp, 2-exp, 344-525 intra-FRET

Parameters which are fitted are given in **bold**, linked parameters in *italic*. fixed parameters are plain text.

Time-point, exp	0 ms, 1-exp	195 ms, 1-exp	955 ms, 1-exp	20 s, 1-exp	Time-point, exp	0 ms, 2-exp	195 ms, 2-exp	955 ms, 2-exp	20 s, 2-exp
DOnly	0.000	0.000	0.000	0.000	DOnly	0.000	0.000	0.000	0.000
R(G,1)	**47.828**	**54.741**	**73.112**	**102.805**	R(G,1)	26.500	26.500	**77.737**	120.000
-	-	-	-	-	R(G,2)	**49.690**	**59.224**	26.500	22.000
R0	52.000	52.000	52.000	52.000	R0	52.000	52.000	52.000	52.000
bg	0.200	0.200	0.200	1.000	bg	0.200	0.200	0.200	1.000
dt	0.117	0.117	0.117	0.117	dt	0.117	0.117	0.117	0.117
g	1.000	1.000	1.000	1.000	g	1.000	1.000	1.000	1.000
k(G,1)	0.000	0.000	0.000	0.000	k(G,1)	0.000	0.000	0.000	0.000
-	-	-	-	-	k(G,2)	0.000	0.000	0.000	0.000
k2	0.667	0.667	0.667	0.667	k2	0.667	0.667	0.667	0.667
kappa2	0.667	0.667	0.667	0.667	kappa2	0.667	0.667	0.667	0.667
l1	0.031	0.031	0.031	0.031	l1	0.031	0.031	0.031	0.031
l2	0.037	0.037	0.037	0.037	l2	0.037	0.037	0.037	0.037
lb	2.000	2.000	2.000	2.000	lb	2.000	2.000	2.000	2.000
n0	21470.7	19064.99	20228.87	24466.95	n0	37294.24	46368.4	36207.1	43764.72
r0	0.380	0.380	0.380	0.380	r0	0.380	0.380	0.380	0.380
rep	20.000	10.000	20.000	20.000	rep	20.000	10.000	10.000	20.000
s(G,1)	6.000	6.000	6.000	6.000	s(G,1)	6.000	6.000	6.000	6.000
-	-	-	-	-	s(G,2)	6.000	6.000	6.000	6.000
sc	0.000	0.000	0.000	0.000	sc	0.000	0.000	0.000	0.000
start	0.000	0.000	0.000	0.000	start	0.000	0.000	0.000	0.000
stop	59.054	59.054	59.054	59.054	stop	59.054	59.054	59.054	59.054
t0	3.800	3.800	3.800	3.800	t0	3.800	3.800	3.800	3.800
tBg	1.000	1.000	1.000	1.000	tBg	1.000	1.000	1.000	1.000
tDead	85.000	85.000	85.000	85.000	tDead	85.000	85.000	85.000	85.000
tL1	3.990	*3.990*	3.990	*3.990*	tL1	3.990	*3.990*	*3.990*	*3.990*
tL2	0.230	*0.230*	0.230	*0.230*	tL2	0.230	*0.230*	*0.230*	*0.230*
tL3	1.930	*1.930*	1.930	*1.930*	tL3	1.930	*1.930*	*1.930*	*1.930*
tMeas	1.000	1.000	1.000	1.000	tMeas	1.000	1.000	1.000	1.000
ts	-0.690	-0.580	-0.493	-0.700	ts	-0.690	-0.580	-0.493	-0.700
win-size	17.000	17.000	17.000	17.000	win-size	17.000	17.000	17.000	17.000
x(G,1)	**1.000**	**1.000**	**1.000**	**1.000**	x(G,1)	**0.529**	**0.685**	**0.519**	**0.542**
x(G,2)	0.750	*0.750*	0.750	*0.750*	x(G,2)	**0.471**	**0.315**	**0.481**	**0.458**
xL1	0.110	*0.110*	0.110	*0.110*	xL1	0.750	*0.750*	*0.750*	*0.750*
xL2	0.140	*0.140*	0.140	*0.140*	xL2	0.110	*0.110*	*0.110*	*0.110*
xL3	2.3204	6.0231	2.7234	2.6234	xL3	0.140	*0.140*	*0.140*	*0.140*

Chapter 3: Specificity of phase transition for murine guanylate binding proteins in live cells

Version: 14.09.2021

Chapter digest

Aims	Methods	Outcome
Phase separation properties, localization and interaction study of mGBP7 with mGBP2 and mGBP3.	Stable cell line models, MFIS, FRAP, SIM, imaging analysis.	Specific and dynamic interaction and assembly of mGBP7 with mGBP2 (no colocalization/ separated phase) and mGBP3 (colocalization/ mixed phase) depends on the protein's consensus motifs.

Contributions

This chapter is a **manuscript in preparation**. See **Manuscript in preparation** section (**page 188**) for a **full clarification on author list, individual contributions and the state of the manuscript.**

Abstract

Murine Guanylate Binding Proteins (mGBPs) localize in vesicle-like structures (VLS) and at the parasitophorous vacuole membrane (PVM) after infection with parasites, where especially mGBP2 and mGBP7 are assumed to promote parasite membrane disruption. mGBP7 colocalizes with several mGBPs, but it is unclear with which mGBPs it physically interacts. Furthermore, it is elusive which rules apply for dimerization and for intrinsic phase transition into VLS. We characterized MEF mGBP7$^{-/-}$ cells stably transduced with eGFP-mGBP7 and mCherry-mGBP1, -2, -3, -5, -6, -7, identifying mGBP3 and mGBP7 as interacting and mGBP1, -2, -5, -6 as non-interacting partners. We chose mGBP2 and mGBP3 as prototypic partners for further characterization by MFIS-FRET, FRAP, SIM and confocal microscopy. We determined specific high affinity interactions and oligomerization, finding homo-interactions as favored over hetero-interactions. While mGBP7 and mGBP3 colocalized in a mixed and mobile protein phase, mGBP7 and mGBP2 never colocalized, but were present at phase boundaries. We conclude that cytoplasmic dimerization via the LG domain is a necessary condition for further oligomerization. The low C-terminal helix of mGBP7 might act as an anchor to recruit and orient the protein into the VLS protein phase.

1. Introduction

Guanylate binding proteins are a family of interferon (INF) inducible 65-73 kDa GTPases, which are strongly upregulated after INF stimulation [1]. They exhibit protective functions against several pathogens and play important roles in the cell autonomous defense [2]. Usually, mGBPs are studied in the context of immunity to intracellular pathogens, such as *C. trachomatis* or *T. gondii* [3-5]. In an earlier study [6], we used a Multiparameter Fluorescence Image Spectroscopy-FRET (MFIS-FRET) approach to show homo-interactions of mGBP2 and hetero-interactions with mGBP1 and mGBP3 by forming supramolecular cytoplasmic complexes which directly attack *T. gondii* after infection. In the infected host cell, *T. gondii* has two membranes: The outer layer PVM (parasitophorous vacuole membrane) from the host cell and the inner layer pPM (parasite plasma membrane). The PVM serves as interface and modulation platform [7] and is the first target of mGBPs and other proteins of the host immune defense [8]. The pPM is exposed after PVM disruption and has been shown to be attacked by mGBP2 [9]. The exact molecular mechanism underlying the disintegration of the PVM remains elusive [8]. To increase membrane affinity, the C-termini of mGBP1, mGBP2 and mGBP5 can undergo posttranslational prenylation [10]. Intriguingly, loss of the prenylation motif prevents assembly of mGBP2 to the PVM [3]. mGBP3, mGBP4, mGBP6, mGBP7, mGBP9 and mGBP10 do not have a CAAX prenylation motif [11]. However, comparative models and simulations suggest that the C-terminal tail of mGBP7 might act as a lipid anchor by adopting an α-helical/coil-coil conformation as well [12]. Furthermore, deletion of the C-terminal sequence prevented formation of VLS and efficient assembly at the PVM.

In addition to the differences in posttranslational modifications, mGBP homologs also show differences at sequence level [11, 13]. mGBP7 shares a high sequence homology with mGBP3 (identity 73%, similarity 87%), but much less with mGBP2 and hGBP1 as shown in [12]. Phylogenetic analysis indicated, that mGBP7 and mGBP3 are much closer to each other than to mGBP2 [14]. Until now, only comparative models for mGBP7 are available [12] based on the hGBP1 crystal structure [15] as well as structure predictions from the deep mind network AlphaFold2, that is now available on the AlphaFold protein structure database [16].

All GBPs are composed of three sub-domains, the N-terminal LG-domain, middle and a C-terminal α12/13 domain. By mapping the protein dynamics with sm-FRET, we identified two distinct sub-domains playing a key role of free energy minimization and in the initiation of oligomerization (=GBP polymerization) [17]: While the first contact is made via the LG-domain and thereby initiating the dimer formation, the C-terminal α12/13 interaction is stabilizing the protein conformation in its oligomer phase. The hGBP1 protein undergoes a

GTP-triggered conformational change [17] from a compact (monomer) to an extended "pocketknife" (oligomer) state [18, 19]. Moreover, oligomerization, which triggers formation of highly concentrated vesicle-like structures (VLS), is already extensively characterized for mGBP2 in MEF cells [6].

The formation of such highly concentrated protein phases is controlled by external parameters (temperature, pH, concentration, pressure, post-translational modifications) and is known as phase transition. Each phase is characterized by distinct macroscopic properties, such as intrinsic concentration, pressure and composition. Minimization of the Gibbs-free energy, promotes molecular polymerization to regular or amorphous structures (aggregates) [20]. While phase transitions are mainly linked to IDPs such as Nucleoli [21], ribosomes [22], Cajal- and P-bodies [23], as well as centrosomes [24] and proteasomes [25], highly ordered proteins are also capable of fast state reversibility [17, 26] and can assemble into concentrated protein phases such as microtubuli [27], actin [28] and hGBP1 [29].

As the mGBP7 interaction network with other mGBPs is largely unknown, we want to understand the underlying interaction organization principles leading to the fundamental questions: Which rules apply during dimerization and intrinsic phase transition? Which mGBP members interact directly with mGBP7 in cytoplasm, VLS and at the PVM? What are the specific compartment-wise differences (protein phases, localization and sub-structural arrangement, affinity and mobility) between interacting and non-interacting mGBP members?

To answer these questions, we combined SIM and confocal fluorescence imaging spectroscopy in live cells to determine compartment specific prerequisites for mGBP7 interaction with other mGBP species. We studied:

(i) oligomerization properties and interaction affinities by MFIS-FRET measurements in cytoplasm, VLS and at the PVM after *T. gondii* invasion,

(ii) diffusion dependent recovery of mGBP species of VLS after FRAP,

(iii) (co-)localization and distribution within the different cellular compartments by SIM (structured illumination microscopy) and

(iv) protein species dependent complex assembly by using scatterplot analysis and a cellular back mapping approach.

2. Results

mGBP7 interacts with itself and mGBP3, but not with other mGBP members in VLS and at the PVM.

In our initial FRET-based study, we studied the interactions of mGBP7 with other mGBPs. We took advantage from strong and heterogeneous inter- and intra-cellular expression levels of the target proteins in uninfected and infected cells, since our interaction studies and analysis of protein phases required a wide range of concentrations (**Figure 1**). In addition to a colocalization study (that reports on relationship between the analyzed proteins), our FRET-based approach is able to resolve direct interactions on a molecular level. Still, colocalization is a necessary condition for any FRET experiment in cells. Consequently, we match our findings to already published colocalization data. The data obtained from the colocalization study [30] agree very well to our interaction analysis (**Figure 1-table supplement 1**). The interaction of mGBP7 with mGBPs in our FRET-based experiments is defined by changes in the fluorescence-weighted lifetime of the donor eGFP-mGBP7, $\langle \tau_D \rangle_f$, in cytoplasm (CP), in vesicle-like structures (VLS) and at the parasitophorous vacuole membrane (PVM), caused by the close spatial distance to the mCherry-tagged mGBPs [6]. In this study, we used our established Multiparameter Fluorescence Image Spectroscopy FRET (MFIS-FRET) for a comprehensive analysis of live cell FRET data free of anisotropy artifacts [31, 32]. MFIS-FRET is superior to traditional FLIM (Fluorescence Lifetime Imaging Microscopy) by multiparameter detection, high spatial resolution and photon boost by noise-reduction through selective pixel averaging. An analogous workflow was established earlier for mGBP2 [6], where we likewise characterized the FRET properties compartment-wise. The fluorophores, genetically encoded fluorescent proteins eGFP and mCherry, were fused to the N-terminus of mGBPs in live cells. Both eGFP and mCherry are commonly used as a FRET pair with a Förster radius of $R_0 = 52$ Å [33]. eGFP serves as the donor molecule and mCherry as the corresponding acceptor. As depicted in **Figure 1**, we observed a highly significant ($p < .0001$) decrease in $\langle \tau_D \rangle_f$ of mGBP7 in the VLS and at the PVM when co-expressed with itself or mGBP3. We also observed a highly significant ($p < .0001$) decrease in $\langle \tau_D \rangle_f$ in the CP of mGBP7 when co-expressed with itself (mCh-mGBP7). All other combinations yielded no detectable decrease in $\langle \tau_D \rangle_f$ in any compartment (for details see **Figure 1-table supplement 2**).
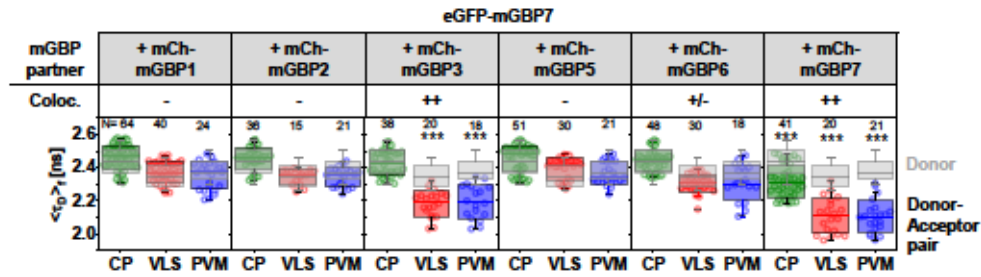
Figure 1: Lifetime based FRET study of eGFP-mGBP7 with mCh-mGBP members. *Donor-only (eGFP-mGBP7 only, grey boxes) showed fluorescence weighted averaged lifetimes between 2.3 to 2.5 ns in respect to the designated subcellular compartments (CP green, VLS red, PVM blue). mGBP7 combined with mGBP1, mGBP2, mGBP3 (CP), mGBP5 and mGBP6 overlaps with the donor-only boxes. mGBP7 combined with itself and mGBP3 (VLS, PVM) differs from donor-only boxes with high significance. The colocalization data [30] is presented qualitatively, with ++ representing very strong, +/- moderate to low and - negligible colocalization. See **Figure 1-table supplement 1** for full information. Significance was determined with two-tailed t-test. Null-hypothesis tested: Data of donor-acceptor combination is different from corresponding donor-only data. Significance level *** = p < .0001 Due to the environmental sensitivity of the eGFP fluorescence lifetime and heterogeneous expression levels, accuracy is limited. Deviations are possible and naturally occurring. Therefore, we only show experiments with highest significance level. Refer to **Figure 1-table supplement 2** for all p values of the dataset.*

In the following, we chose two prototypic models for non-interacting and interacting mGBPs to study the molecular basis for interaction and assembly in VLS and at the PVM intensively: mGBP2-mGBP7 was chosen as a non-interacting mGBP pair, while mGBP3-mGBP7 was chosen as the interacting one. We subsequently used these combinations for

(i)   SR microscopy and scatterplot analysis to map compartment-wise differences between the mGBPs in uninfected and infected cells **(Figure 2-3)**,

(ii)   a FRET-based study to determine the interaction and oligomerization affinity **(Figure 4)**

mGBP3 and mGBP7 colocalize in vesicle-like structures, but mGBP7 and mGBP2 do not colocalize and remain separated at phase boundaries.

We analyzed images from confocal microscopy of the two prototypic models and created scatterplots. We identified populations within the scatterplot based on signal intensities and mapped them back to their cellular origin by using the freely available Fiji plugin ScatterJ [34]. This is advantageous, as the usual standard analysis based on Pearson correlation coefficients (PCC) averages over subpopulations and important details get lost. We first performed a conventional colocalization analysis by comparing PCC values (0.19 ± 0.16 for mGBP7 and mGBP2; 0.61 ± 0.24 for mGBP7 and mGBP3) with already published [30] PCC values, as shown in **Figure 2-figure supplement 2 and Figure 2-table supplement 1.** The PCC values were very similar. On top to the conventional analysis, we used the scatterplot approach to highlight the importance of cellular resolution in two prototypic mGBP-systems by identifying (i) VLS heterogeneities, (ii) pattern changes in infected cells (PVM vs VLS due to changes in the green-to-red intensity ratio between the mGBPs) and (iii) a unique population in non-interacting mGBPs. In the following, we explain these findings: In **Figure 2A and 2B**, we performed scatterplot analysis and cellular back mapping in uninfected and infected cells (with *T. gondii*). The corresponding scatterplots (**Figure 2C**) show the fluorescence intensity of the eGFP and mCherry channel. We identified three distinct VLS groups in the mGBP2-mGBP7 scatterplot (**Figure 2C, uninfected**). First, mGBP2 and mGBP7 never colocalize. mGBP2 alone (red subpopulation) mainly localized at the border of a large VLS with mGBP7 mostly unmixed in the center (orange subpopulation). Additionally, mGBP7 formed small VLS (green population), which fuse with the larger red-orange particles. The situation at the PVM is more complex, as displayed by the scatterplot populations. Especially the orange population distributed over a wide intensity range. The heterogeneous distribution of this particular population (orange) is -at least partly- caused by the axial and lateral resolution limit of confocal microscopy to clearly resolve non-colocalizing fractions of proteins in VLS (particles smaller than approx. 200 nm in diameter, see **Figure 3A**). Scatterplot analysis of mGBP7 and mGBP3 indicates a high degree of overlap (orange population) and a small mGBP7-only (green) population (**Figure 2C, uninfected**). A population representing mGBP3 alone was not detectable. Evidently, the green populations assemble and merge with the orange (double positive) populations (**Figure 2B**). In cells infected with *T. gondii*, the scatterplot analysis indicates a more heterogeneous population of mGBP7-mGBP3 VLS (orange) with a clear change of the scatterplot slope, thus altering the intensity ratio between VLS and PVM (blue). Consequently, both populations are distinguishable (**Figure 2B/C**, infected). Overall, different degrees of colocalization of both proteins were clearly distinguishable.
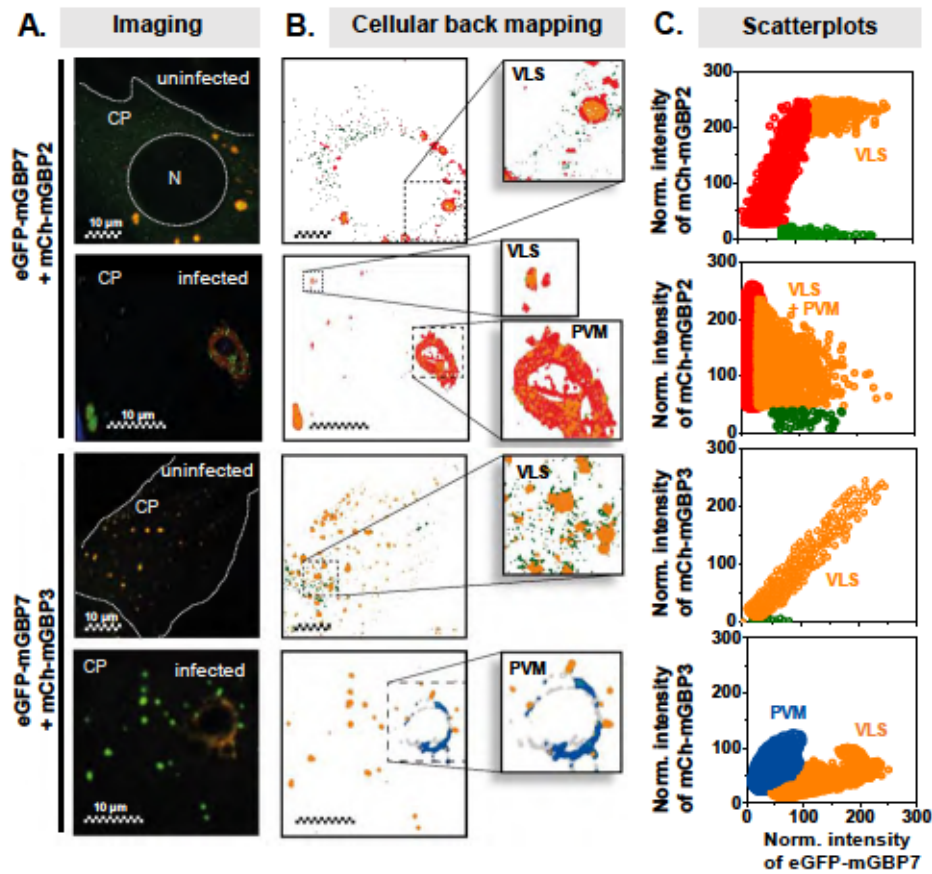
Figure 2: Detailed colocalization analysis of mGBP7 with mGBP3 and mGBP2. *A. Confocal images used in scatterplot analysis. Compartments and cell shape are highlighted in the merged image. B. Cellular back mapping approach of selected scatterplot populations. Top section: Back mapping of the mGBP2 and mGBP7 populations. The zoom in shows the heterogeneous VLS: Yellow, mixed species in the middle of a condensate, free mGBP2 (red) at the condensate rim and small VLS, and free mGBP7 (green) in small separate VLS of mGBP7. In infected cells, VLS behave similar while the PVM appears heterogeneous, partially mixed or single-species (mGBP2) with no clear pattern.   Bottom section: Localization of the colocalizing population containing mGBP7 and mGBP3 (orange), as well as a mGBP7 only fraction (green) within the measured MEF cell. The zoom-in shows how the green populations assemble the orange population. Although mGBP7 and mGBP3 assemble homogeneously in infected cells, a difference in stoichiometry is already visible at the PVM (blue) compared to the VLS (orange). C. mGBP7 and mGBP2 show separate green and red populations. A third population with constant high mGBP2 intensity and increasing mGBP7 intensity (yellow) was identified. In infected cells, VLS and PVM follow this trend but are far more heterogeneous in the mixed population (orange). mGBP7 and mGBP3 show a high degree of colocalization (orange population) and only little amount of free mGBP7 (green). Situation in T. gondii reproducibly changes from one into two intensity ratio populations, forming distinct slopes, which we clearly assigned to either VLS (orange) or PVM (blue)*

mGBP3 and mGBP7 form sub-structured mixed VLS, whereas mGBP2 and mGBP7 assemble in spatially separated VLS

To improve resolution of the mGBP7 containing condensates and verify the hypothesis that heterogeneity occurs within the protein phases, we imaged fixed MEF cells by Structure Illumination Microscopy (SIM) along with a concurrent confocal particle analysis. Our findings suggest the presence of heterogeneous sub-structures in very small VLS and larger structures. As suggested from the images (**Figure 2-figure supplement 1**), a particle size analysis shown in **Figure 3-table supplement 1** indicates differences in particle size and number: mGBP7 in a homo phase and non-interacting environment (as with mCh-mGBP2) formed very small particles of $0.13 \pm 0.34\ \mu m^2$. Interestingly, mGBP7-mGBP3 particles have average particle sizes up to $0.37 \pm 0.61\ \mu m^2$, three times larger than in mGBP7 cells without mGBP3. In contrast, mGBP2 was extremely variable in size and form, reaching an average particle size of more than tenfold ($1.78 \pm 6.61\ \mu m^2$). The differences in distribution of eGFP-mGBP7 VLS and mCh-mGBP2 VLS in MEF mGBP7$^{-/-}$ cells (n = 9) is described in a histogram and highlighted by the cumulative probability mass function Px_i as shown in **Figure 3A**. Using the 2σ criterion of size distribution, we identified a size threshold at approx. $4.5\ \mu m^2$ (5%), from where we start to term the VLS as large particles. VLS of eGFP-mGBP7 and mCh-mGBP2 are distinctively different in size with eGFP-mGBP7 forming the smaller VLS, as visible in the cumulative probability mass function Px_i and shown in **Figure 3-table supplement 1**. Concurrently, the number of small eGFP-mGBP7 VLS near the diffraction limit is higher than for mCh-mGBP2 VLS, which follow a broader distribution towards larger VLS.

Of note, visual inspection and cell counting of MEF mGBP7$^{-/-}$ cells (n = 500) with eGFP-mGBP7 and mCh-mGBP2 ± INFγ indicated that at least in 65% cells (n = 325) particles with a length of >5 μm and clearly separated mCh and eGFP species can be found. Representative eGFP-mGBP7 and mCh-mGBP2 particles are shown in **Figure 3B**. We cannot rule out very small mGBP7 VLS (< 1 μm) within the designated single-positive mGBP2 VLS (n = 155), as the scatterplot analysis (**Figure 2B**, infected) already suggested the presence of VLS smaller than 1 μm consisting of both non-interacting mGBPs. Thus, the real number of mGBP2-mGBP7 particles is higher. As these particles are extremely large and not necessarily globular, we differentiate between the small particles = VLS and the larger ones with clearly separated mGBPs, that we demark in the following "large particles". The heterogeneity occurs due to cell-to-cell variations in protein expression and differences in the respective stoichiometry. In **Figure 3B**, we show a representative confocal imaging gallery of eGFP-mGBP7 and mCh-mGBP2 double positive VLS. Different degrees of co-localization of both proteins were clearly distinguishable. The VLS could be resolved by

confocal microscopy at high resolution if large optical zoom (according to the Nyquist criterion) and very low power (nW) were used. The largest VLS (**Figure 3C**, white box) were selected for magnification in both combinations. As observed before with confocal microscopy and scatterplot analysis, a clear separation between mGBP7 and mGBP2 was confirmed in **Figure 3C**. Small VLS were visible which fused into the existing perinuclear, larger VLS with mGBP2 encompassing mGBP7 for unknown reasons. Compared to confocal imaging (**Figure 3B**), we were now able to resolve sub-structures in the mGBP7 phase (**Figure 3C**). Furthermore, the VLS can be distinguished into mGBP2-poor phases and into mGBP2-dense phases. The interior mGBP2-poor phase was characterized by the presence of a heterogeneously sub-structured mGBP7 phase. However, the mGBP2-dense phase was always free of mGBP7 protein. Also shown in **Figure 3D**, mGBP7 and mGBP3 particles were nearly always colocalized, but with fluctuating green to red ratios indicating a variable mGBP7/mGBP3 stoichiometry. This trend was already observed in cellular back mapping (**Figure 2**) and fits reasonably well to the heterogeneity observed in FRAP recovery and dynamics, as detailed below.
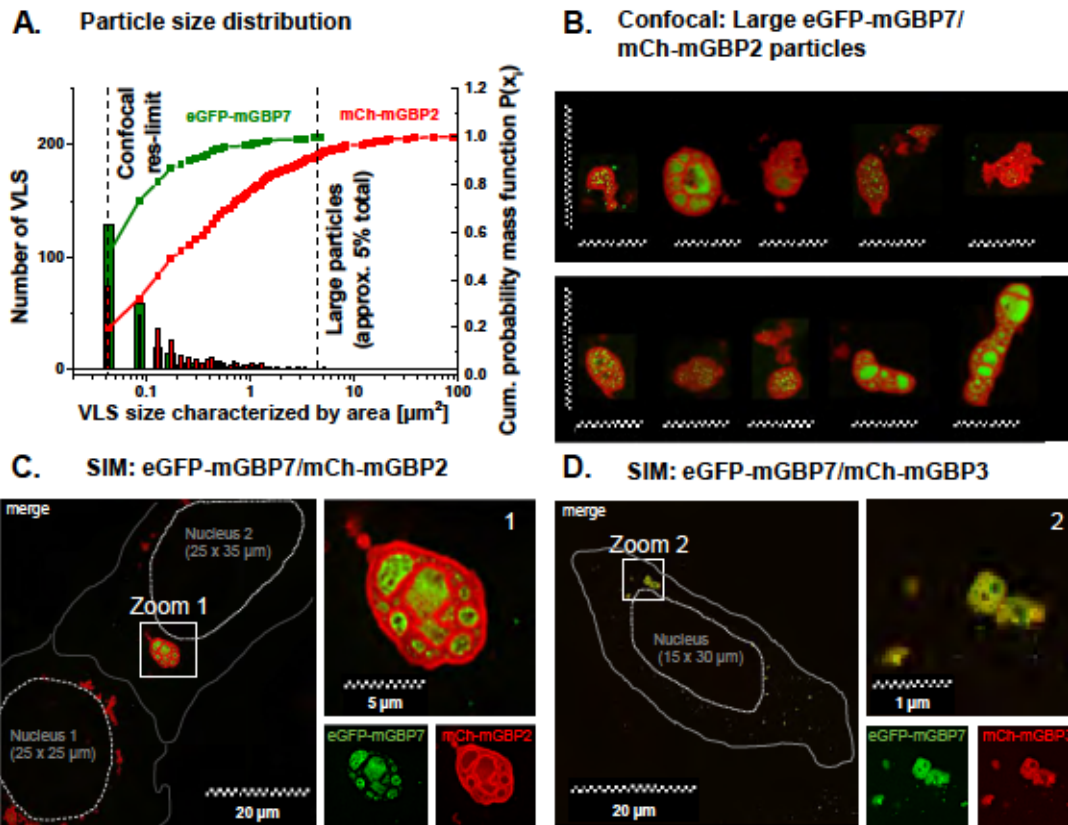
Figure 3: SIM reveals different degrees of heterogeneity in mGBP7 containing condensates.

A. Distribution of VLS size characterized by area of particles in the confocal plane (n = 9 cells). Highest frequency of VLS is found at the confocal resolution limit (here: 207 nm). eGFP-mGBP7 VLS are smaller than mCh-mGBP2. Cumulative probability mass function $P(x_i)$ reveals the main fraction (95%) of VLS as smaller than 4.5 μm^2. VLS larger >4.5 μm^2 are termed 'large particles'. B. Representative gallery of eGFP-mGBP7/mCh-mGBP2 particles. The vertical scale bar is 20 μm, the horizontal scale bar is 10 μm C. eGFP-mGBP7 (green) in presence of mCh-mGBP2 (red.) Left image shows two representative cells with predominantly perinuclear eGFP-mGBP7 (green) and mCh-mGBP2 (red) positive large particles of different areas >10 μm^2 as well as mGBP2 or mGBP7 homo-VLS. For orientation, the nuclei and membranes were indicated in grey. Right image (1) is a zoom of this giant VLS. A grid size of 34 μm for GFP and 42 μm for mCherry was used, rotating 5 times. D. eGFP-mGBP7 (green) in presence of mCh-mGBP3 (red). Left image: Both protein species are co-localized, but a perinuclear, large VLS (white box, zoom 2) clearly shows sub structure (right image). For orientation, the cell nucleus was outlined.

Formation of high affinity mGBP7 homo-oligomers and medium affinity hetero-oligomers.

Our previous analysis clearly indicated a significant colocalization of mGBP7 and mGBP3. Therefore, we use the established MFIS-FRET technique to measure and analyze the oligomerization properties dependent on cellular heterogeneity (concentration effects) under live-cell conditions. In detail, we determined:

(i)   the fraction of FRET-active molecules, x_{FRET}, *judged by* comparison of time-resolved, FRET-induced donor quenching, $\varepsilon_D(t)$,

(ii) FRET-based affinities by the concentration-dependence of x_{FRET}. Here, we use the observed intra- and inter-cell expression variability of the respective mGBPs to determine concentrations from measured fluorescence intensities under calibrated conditions.

(iii) k_{FRET} to estimate and compare differences in oligomerization between the compartments.

As observed before, eGFP-mGBP7 protein in absence or presence of its acceptors mCh-mGBPs exists in the cytoplasm in differently concentrated assemblies (**Figure 1**). The FRET properties of the system were also visualized directly by time-resolved, FRET-induced donor quenching, $\varepsilon_D(t)$ [35], [17]. It directly resolves the fraction of FRET-inactive species ($x_{NoFRET}=1-x_{DA}$) as constant offset. **Figure 4A** displays different $\varepsilon_D(t)$ curves of eGFP-mGBP7 in absence or presence of mCh-mGBP2, mCh-mGBP3 and mCh-mGBP7 in CP (green), VLS (red) and PVM (blue). Accordingly, the fraction of FRET-active mGBP7 molecules, x_{FRET}, increases concentration-dependent, from CP to VLS and PVM. Consistent with the results in **Figure 1**, almost no cytoplasmic hetero-FRET was observed for mGBP7 in presence of the acceptors, mGBP3 and mGBP2. No changes in x_{FRET} were detected in VLS and only minor changes (concentration effect) at the PVM in mGBP7/mGBP2 cells. For the acceptor mGBP3, increasing x_{FRET} was detected in a concentration-dependent manner in VLS and PVM.

The distinct protein concentration and fraction of FRET-active molecules can be used to estimate FRET-based affinities (**Figure 4B**). For determination of K_D, we fit the relation between x_{FRET} and the acceptor concentration. The observed variation in data is caused by the heterogeneous properties of mCherry in cells. The apparent K_D indicates that affinity between the same molecules, namely mGBP7, is much higher and therefore preferred over interaction with other molecules, in this case mGBP3 (weaker affinity). We never observed affinity of mGBP7 towards mGBP2. The FRET rate constant, k_{FRET}, is utilized as an additional measure to compare the increasing amount of FRET upon oligomerization due to the rising acceptor concentration (Figure 4C).

Cytoplasmic mGBP7 in presence of mGBP3 shows the lowest k_{FRET} and twofold increase in VLS and at the PVM. The FRET rate is higher for homo-interaction of mGBP7 and is comparable to mGBP2 from our previous study [6]. Therefore, we suggest that mGBP7 forms dimers and higher-order oligomers similar to mGBP2.
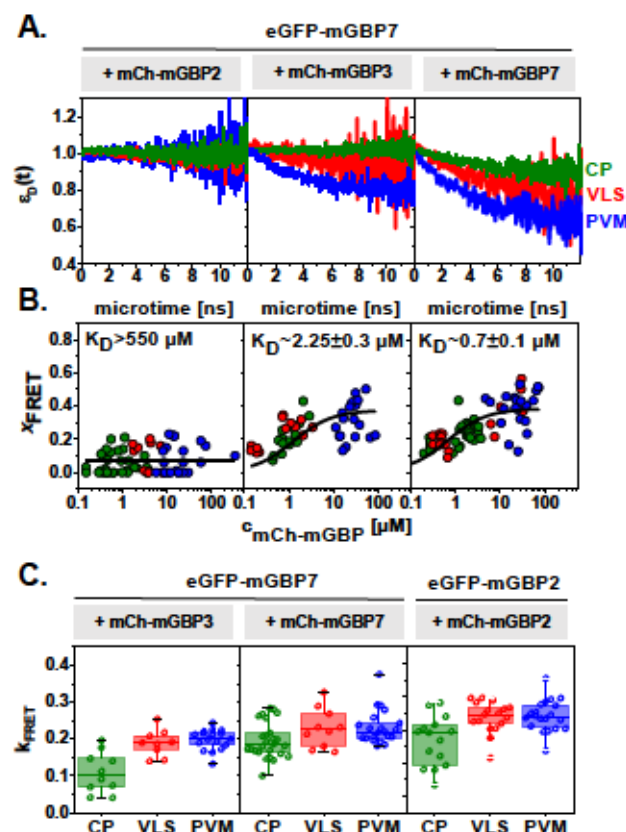

Figure 4: FRET-based analysis on mGBP7 oligomerization and affinity. A. *The slope and saturation of the curve of the FRET-induced donor quenching $\varepsilon_D(t)$ displays the ratio of the two fluorescence decays $f_{DA}(t)/f_D(t)$ of eGFP-mGBP7 (D) in presence of the acceptors (A) in CP (green, 4 µM protein), VLS (red, 20 µM protein) and at the PVM (blue, 100 µM protein) in a concentration-dependent manner.* **B.** *FRET-based affinity of eGFP-mGBP7 was determined by plotting the acceptor concentration versus x_{FRET}. K_D were calculated to be in the lower µM range. The affinity was highest for mGBP7 self-interaction (K_D~0.7, n=84, X^2 =0.00644), weaker for mGBP3 hetero-interaction (K_D~2.25, n=65, X^2 =0.00436) and absent with mGBP2 (K_D~550 µM, n=80).* **C.** *k_{FRET} derived from hetero-interaction of eGFP-mGBP7 with mCh-mGBP3 are the lowest in average due to the lowest affinity between the species. k_{FRET} derived for mGBP7 self-interaction is comparable and slightly lower than mGBP2 self-interaction [6].*

mGBP7 VLS are dynamic but at low recovery rates.

While mGBP7 and mGBP2 have the ability to localize at spatially separated phase boundaries, mGBP7 and mGBP3 form hetero-oligomers within VLS. Scatterplot and cellular back mapping already indicated minor heterogeneities between and inside VLS, suggesting that VLS are composed of irregular structural elements. Therefore, we used FRAP (fluorescence recovery after photobleaching) to analyze the dynamic behavior and composition of mGBP7 containing VLS in different MEF mGBP7$^{-/-}$ cell lines with stably transduced with eGFP-mGBP7 in presence of mCh-mGBP2 or -mGBP3).

We performed FRAP measurements on VLS with an average size of approximately 3-7 µm and recorded the fluorescence recovery within an observation window of 360 s. FRAP on smaller VLS gave no reasonable results, due to fast movements of the observed particles. Furthermore, we only showed the first 250 seconds of the recovery curves, as fluctuations and noise increased. eGFP and mCherry were completely bleached (**Figure 5A**), leading to the background corrected recovery curves shown in **Figure 5B**. We were unable to observe full recovery for either eGFP nor mCherry attached to the respective mGBP. Describing the observed recovery curves by a two-component bi-exponential diffusion model consisting of two recovery times and corresponding fractions (see **eq. (1)**, material and methods and **Figure 5-figure Supplement 1** for model selection), leads to observed values shown in **Figure 5C**. While the recovery level can be used to make a statement about the fraction of freely diffusing (mobile) molecules within the respective VLS, the recovery time quantifies the velocity of mobile molecules.

Interestingly, FRAP properties determined for eGFP-mGBP7 were strongly dependent on the compartment. While mGBP7 and mGBP3 show similar recovery properties within the VLS, we observed differences for mGBP7 and mGBP2 with respect to saturation level and recovery time. eGFP-mGBP7 in the presence of mGBP3 shows a faster recovery ($\langle \tau_{1/2} \rangle_{eGFP}^{mGBP7/3}$ = 63 ± 9 s) and higher mobile fraction (38 ± 8%) compared to eGFP-mGBP7 in presence of mGBP2 (18 ± 2%, $\langle \tau_{1/2} \rangle_{eGFP}^{mGBP7/2}$= 14 ± 3 s). These findings may indicate that the observed recovery rates of mGBP7-containing VLS are due to heterogeneities within the different structures in accordance with the interaction ability (little heterogeneity in double positive mGBP7-mGBP3 VLS, stronger heterogeneity in mGBP7-mGBP2 VLS). Previously observed sub-structures of mGBP7-mGBP2 VLS may work as a diffusion barrier for mGBP7. The mGBP7 phase is always encompassed by mGBP2, so that the recovery times are significantly slower, leading to a lower mobile fraction.

Furthermore, local differences in cytoplasmic protein concentrations may also influence the recovery in all cases. We also observed faster recovery in proximal regions and slower recovery in the structure's core (spatial diffusion heterogeneity), explaining the need to fit data bi-exponentially and resulting in different recovery rates (**Figure 5-figure supplement 2A/B**). Moreover, we observed fusion, fluctuations and high mobility of VLS, which result in broad heterogeneity of the observed recovery curves (**see Figure 5-figure supplement 2A**).
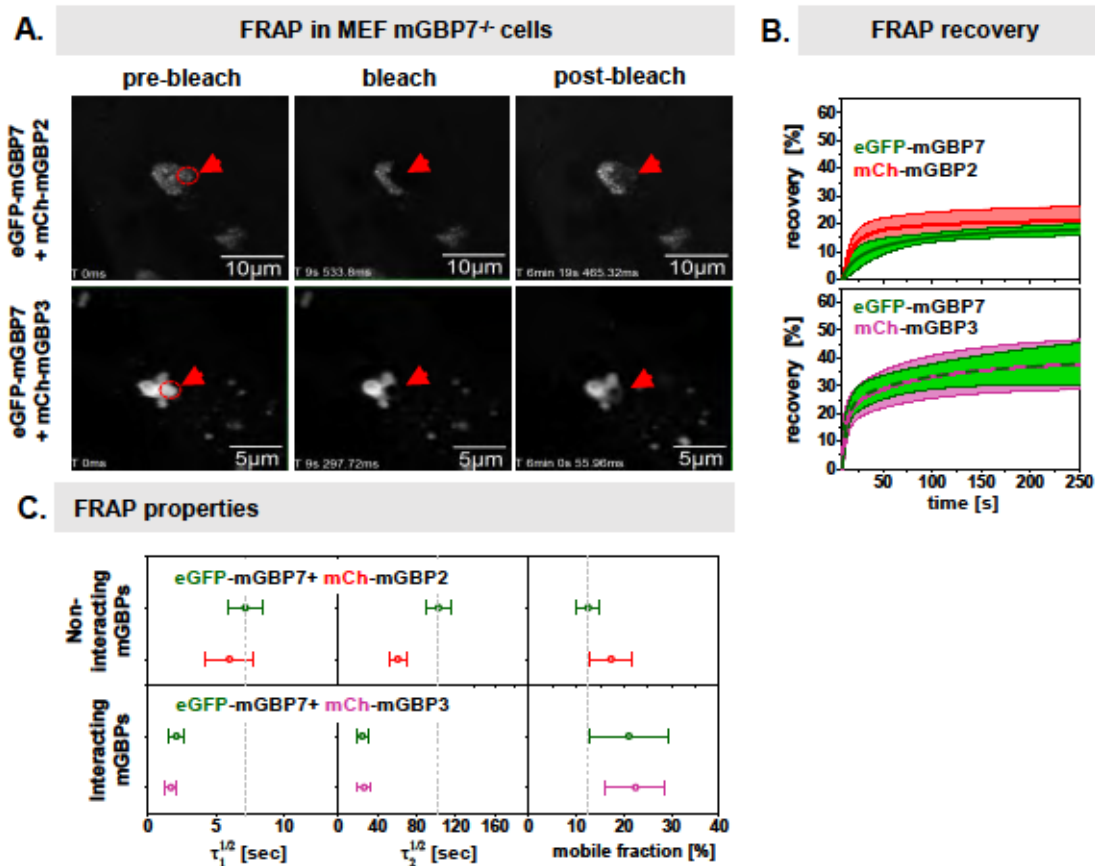

Figure 5: Fluorescence recovery of fluorescently labeled mGBPs. *A. Exemplary images of mGBP7-mGBP3 (top row) and mGBP7-mGB2 (bottom row) at different times within the FRAP experiment (t=0s, 9s and 6 min; scale bars 5 µm and 10 µm, respectively). The red arrows indicate the bleached area (red circle at t = 0s). Corresponding time dependent recoveries [%] are shown in B. for mGBP7-mGBP2 (top), mGBP7-mGBP3 (middle). Recovery curves are fitted with a bi-exponential model (mean eGFP recovery: dark green; mean mCherry recovery: red/purple dashed line with error shown in green and red/purple, respectively) (eq. 1). The obtained mGBP7 recovery times and mobile fractions are shown in C, allowing to identify two types of FRAP recovery in mGBP7 VLS. Individual FRAP curves are shown in Figure 5-figure supplement 2.*

3. Discussion

mGBPs specifically assemble in dynamic and concentrated protein phases within the VLS and at the PVM.

In this study, we analyzed mobility, (co-)localization and interaction affinities between mGBPs with fluorescence spectroscopy and imaging based techniques in three MEF mGBP7$^{-/-}$ cell lines stably transduced with FP-fused mGBP7, mGBP3 and mGBP2. MFIS-FRET measurements indicated a high affinity of mGBP7 with itself, a lower affinity with mGBP3 and no affinity towards mGBP2 (**Figure 1**). As a necessary condition for FRET, confocal and SIM imaging (**Figure 3**) with subsequent scatterplot analysis and cellular back mapping (**Figure 2**) indicated a high degree of colocalization between mGBP7 and mGBP3. The colocalization was restricted to VLS and to the PVM, while mGBP7 and mGBP2 never directly colocalize and interact via FRET (**Figure 4**). Although the majority of these VLS consist of either mGBP7 or mGBP2, both proteins were also found together at the boundary of the two phases. In these condensates, mGBP2 always encompassed mGBP7 and never *vice versa*. This process is possibly driven by fusion of VLS and spontaneous unmixing of the non-interacting proteins, creating the large spatially separated VLS shown in **Figure 3A**. After infection, the degree of colocalization and resulting FRET based affinities between mGBP7 and the analyzed mGBPs at the PVM appear more heterogeneous and slightly weaker compared to the VLS. To assess whether the VLS are still dynamic or irreversibly aggregated, we performed FRAP experiments **(Figure 5)**. We observed, that all mGBP species are mobile but to a different extend, even in very large VLS (3-7 µm).

Structural similarity between LG-domains, low complexity features and the C-terminal region of mGBPs determine specific phase transitions into dynamic VLS.

Almost all human and murine Guanylate Binding Proteins (mGBPs) are reported to form vesicle-like structures (VLS) [36]. While some GBPs colocalize in the same VLS, others stay clearly spatially separated (phase boundary). We attribute this to the structural similarity between the GBPs. We used the mGBP7 homology model [12] for further analysis. As shown exemplarily in **Figure 6A**, most members of the GBP family share the highest sequence identity in their N-terminal, structured LG domain (**Figure 6-figure supplement 1**). The degree of sequence similarity and key residues, especially in the LG, the presence of a membrane anchor, C-terminal low complexity regions and their post-translational modifications may influence the homo- and hetero-oligomerization and therefore the ability to form VLS. Throughout our characterization, we observed frequent changes in size, shape and number of VLS among the MEF mGBP7$^{-/-}$ cell lines. Based on our experimental findings, we propose a phase separation behavior for mGBPs in their cellular environment, which

causes the mGBPs to transit from the solvated cytosolic dimeric form into a dense oligomeric protein phase, ultimately assembling as VLS. This transition is dependent on

(i)   the ability of mGBPs to dimerize via their LG domain and

(ii)   biophysical properties such as hydrophobicity (largely determined by the C-terminus) and electrostatic interactions, making it thermodynamically favored to align the proteins in the same or spatially separated phase.

(i) Considering sequence similarity, we noticed a strong tendency of all mGBPs towards homo-oligomerization. This observation is in accordance with studies with hGBPs [29, 37], and mGBP2 [6]. Consequentially, a higher similarity in sequence would always correlate with higher interaction potential. This applies especially to the LG, were the first contact of the dimer occurs. Evidently, the higher similarity (86.6%) of mGBP7 with mGBP3 favors colocalization but not between mGBP7 and mGBP2 (71.9%). Our FRET experiments (**Figures 1,4**) further support this finding, revealing mGBP7 and mGBP2 as non-interacting. For mGBP7 with mGBP3, FRET exclusively revealed the homo-dimerization via the LG domain of mGBP7 as having the highest affinity, followed by hetero-dimerization with mGBP3. As already reported in literature [6, 38], mutational deletion of key residues in the LG domain prevent dimerization. Intriguingly, inability to dimerize would also abolish the ability of mGBP2 to assemble in VLS. Therefore, we conclude that LG-dimerization of GBPs is a key step to form oligomers and VLS formation [29]. In an effort to explain the inability of mGBP2 to interact with mGBP7, we compared the full interaction interfaces found in three hGBP1 dimer crystal structures [39] (PDB: 2b92, 2bc9, 2b8w, see **Figure 6-table supplement 1**) to the corresponding residues of all three mGBPs. While most of the contacts appear well conserved over all homologues, we find a strong accumulation of dimeric polar contacts impaired in the guanine cap due to substitutions of key residues (**Figure 6B**). The guanine cap (GC) is a highly flexible region, which can undergo many smaller and large conformational changes [40]. It has a relatively low homology compared to the rest of the LG domain. Upon GTP binding, it changes from the open conformation into a closed conformation and subsequently allows dimerization [41]. In hGBP1, the most important residues for this process, R240 and R244, are responsible for roughly half of the $\Delta\Delta G$ in the dimerization process [42], maintaining crucial contact within the dimer interface during nucleotide hydrolysis [29, 41]. R240 is conserved over hGBP1 and all three mGBPs in this study. However, R244 is mutated to lysine (K244 in mGBP7) in mGBP7 and mGBP3, but not in mGBP2. Crucially, at the -1 position of R244, mGBP7 and mGBP3 both introduce an additional aspartate (D242 in mGBP7) into the motif, effectively elongating the guanine cap by one residue. (**Figure 6D**). Comparing the guanine cap crystal structure of nucleotide-bound hGBP1 (PDB: 1f5n) to the mGBP7 homology structure model [12], we found a distinct

gap between the structures in the respective helix turn (**Figure 6D**). Therefore, the angle of the K243 in mGBP7 is distorted against the R244 in hGBP1, deviating from the original topology. Intriguingly, the shorter 'gap-R/X' type found in hGBP1 and mGBP2 seems to be a typical motif for prenylated GBPs (h/mGBP1, h/mGBP2 and, to a lesser extent, h/mGBP5) [11, 39]. The +1 longer 'DK' cap found in mGBP7 and mGBP3 typically exists in GBPs without CAAX-motif (mGBP3, hGBP4, mGBP6, mGBP7, mGBP10) [11, 39]. Since a weak interaction between mGBP2 and mGBP3 was reported earlier [6], a lower affinity interaction must still be possible and the cap motif and length does not appear as fully deterministic for the interaction. Still, it might play a fundamental role as one key structural element to mediate not only homo- but especially hetero-dimerization within the same type/motif, explaining a part of the difference we see in LG hetero-interaction between mGBP7 and mGBP2 or mGBP3.
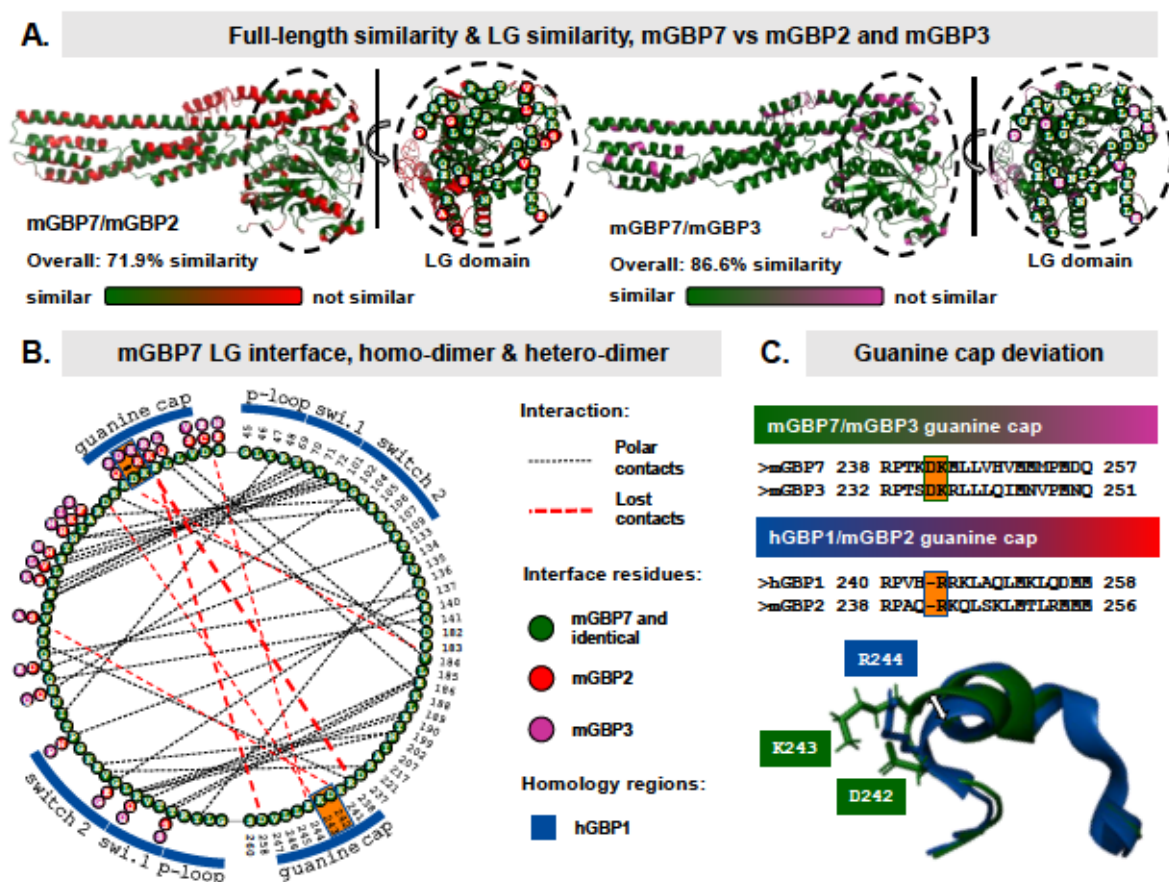

Figure 6: Structural analysis of mGBP7 and homologues. *A. Side-view: Full-length similarity between mGBP7 and mGBP2 (72.9%), mGBP7 and mGBP3 (86.6%). Front-view: LG domain with interface residues highlighted. Green/solid circle: Identical, green/dotted circle: Similar, red/magenta: not identical. B. LG-interface of mGBP7 with itself, with mGBP2 and with mGBP3. Polar contacts taken from corresponding residues hGBP1 dimer structures, DSSP-based standard via pymol.*

*Identical residues and mGBP7-residues in green, residues found specifically in mGBP2 in red, residues found specifically in mGBP3 in magenta. Catalytically important regions are highlighted (blue). Maintained polar contacts displayed as black dashed line. Lost contacts displayed as red dashed line, thickness represents the number of contacts lost at the respective position between the homologues. **C.** Guanine cap sequences aligned. The gap found in the guanine cap of hGBP1 and mGBP2 is marked in orange. Crystal structure alignment reveals a strong deviation around R244 (hGBP1). RMSD of guanine cap alignment goes up to 3.70 vs 1.22 of the whole LG-domain alignment.*

(ii) After the high-affinity LG dimerization step, a re-orientation occurs in the dimer. This was recently shown for hGBP1 [17, 43] and results in a low-affinity interaction between the α-13 helices, which is sufficient due to the close proximity in the dimer. The re-orientation, while retaining structural plasticity, can affect hydrophobicity and polar contacts and promote assembly within protein phase rather than a solvated species. Again, high similarity will increase the chance of the features being beneficial for homo-phase formation. Accordingly, a local concentration gradient (**Figure 3C**) as observed in the VLS could be a consequence of a favorable homo assembly rather than mixed species. An essential difference concerning hydrophobicity lies within the prenylation of mGBP2 and other mGBPs at its C-terminus, inducing a janus-headed, highly amphiphilic momentum into the protein. CAAX-mutated variants of mGBP2, unable to be modified via prenylation, fail to form any visible VLS [6], rendering the C-terminal modification and interaction as the second fundamental step towards oligomerization, alongside the earlier LG-interaction. The C-terminal membrane binding helix recently detected in mGBP7 [12] could mimic the concept of mGBP2, although the hydrophobic potential of a prenylated tail harbors a considerably stronger amphiphilic character. Apart from the mismatching LG domains, the described difference in C-terminal hydrophobicity might also determine whether mGBPs assemble in mixed protein phases or unmix spontaneously.

VLS as dynamic protein phase reservoir for immediate attack of mGBPs on intracellular parasites

VLS as mobile protein phases serve as reservoir for multiple mGBP species, keeping cytosolic mGBP-levels effectively under control and induce cooperativity [44]. Upon INF-γ stimulation, cellular mGBP levels rise drastically [45]. The resulting increase in cytosolic concentration needs to be controlled to avoid uncontrolled aggregation and side effects [46] in the cytoplasm, eventually causing severe cellular stress by disturbing vital cytosolic processes. The high throughput of GTP might disturb cellular energy metabolism [47], if not locally controlled. The assembly of mGBPs in highly concentrated, locally restricted and dynamic phases appear as a spatial control mechanism [26] to maintain full cellular functionality and likewise swiftly attack intracellular parasites upon infection [29].

Plant guanylate-binding proteins like GTPases (GBPL) show similar mechanisms, even undergoing liquid-liquid phase separation triggered by long IDRs, that mammalian GBPs lack [48]. We further speculate, that the locally high concentration of mGBPs is regulated by yet unidentified proteins, such as intrinsically disordered proteins (IDPs), which crosslink proteins and RNAs [49, 50].

Mode of interaction of mGBPs is connected to specific domain and phase interactions.

Individual GBPs undergo phase transition from a monomeric and dimeric phase (primarily cytoplasm) to an oligomeric phase (VLS). Based on the structural differences, mGBP7 can form different types of mobile protein phases with other mGBPs. The VLS containing mGBP7 and mGBP2 form immiscible protein phases, while mGBP7 and mGBP3 form a miscible, sub-structured protein phase. We conclude that cytoplasmic (homo-) dimerization via the LG-domain is a pre-requisite for higher order oligomerization presumably mediated by the mGBP C-terminus and subsequent formation of protein phases (=VLS). Whether mGBPs assemble in mixed or separated phases might be determined by LG-affinity and similarity of structural properties (electrostatic potential, hydrophobicity, C-terminal tails) as well as further interaction partners mediating the phase transition. However, (from a thermodynamic perspective) the high homo-affinity of mGBP7 and mGBP2 (homo dimerization) might initiate a nucleation process at a molecular level with lower free energy and therefore allow the system further minimize the free energy by the formation of two distinguishable phases. Once the nucleation is initiated, the growth of VLS is driven coalescence, leading to larger demixed assemblies as shown in **Figure 3**. Once the thermodynamic equilibrium is reached, the concentration of GBPs in solution and in the dense phase are highly different. Due to constant chemical potential in both phases and across the phase boundary, there is no temporal change in chemical potential causing the diffusion flux. The GBPs stochastically diffuse in and out of the different phases undergoing conformational changes and protein interactions. Based on these implications, we propose an interaction model as shown in **Figure 7**.
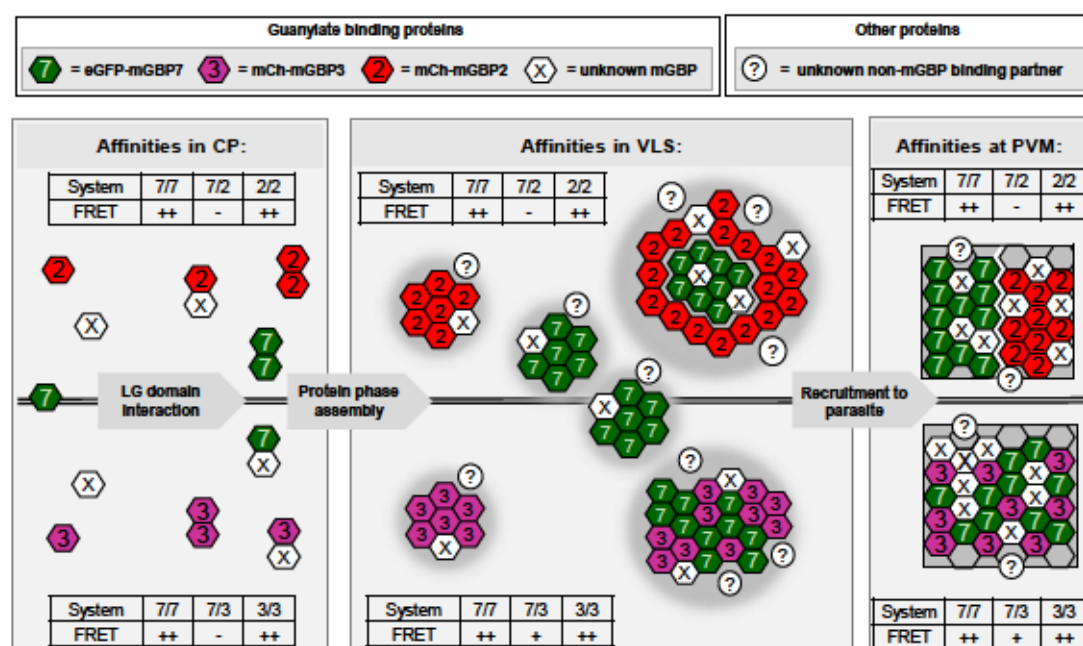
Figure 7: Proposed mode of interaction and phase behavior of mGBP7 with mGBP2 and mGBP7 with mGBP3. *FRET properties as suggested from MFIS-FRET measurements indicate high affinity (++), medium affinity (+) or negligible affinity (-). All species, mGBP7 (green), mGBP2 (red) and mGBP3 (magenta), form homotypic dimer units in the cytoplasm. mGBP7 and mGBP2 are separated in small VLS and larger VLS where mGBP2 always encloses mGBP7 without mixing. The situation at the PVM is more complex, mGBP2 can quickly interact with the membrane due to the prenylation and appears to distribute homogeneously around the membrane. In addition to homotypic assemblies, mGBP7 and mGBP3 form mixed assemblies in VLS and at the PVM. Based on their concentration, oligomerization degree, unknown mGBPs (white hexagons) and possible unknown interactors (white circles) the FRET properties differ between species and compartments. Especially at the PVM, both proteins tend to be more clustered in contrast to VLS, where only micro-heterogeneities are visible.*

Future research is needed to investigate potential interactors involved in VLS formation and recruitment to the PVM, as well as super-resolution imaging and spectroscopy techniques to resolve the underlying heterogeneities and affinities in protein phases in VLS, at phase boundaries and at the PVM with higher precision.

102

4. Material and Methods

Cloning and constructs

NIH 3T3 mouse embryonic fibroblasts (MEFs) from ECACC (Public Health England, Salisbury) were used as model organisms. The WT open reading frame of mGBP7 (NCBI accession number NM_001083312.2), mGBP2 (NCBI accession number NM_010260.1) and mGBP3 (NCBI accession number NM_001289492.1) were cloned into the pWPXL plasmid (Trono lab) as N-terminal GFP-fusion or mCherry-fusion constructs. The lentiviral envelope vector pLP/VSVG (Invitrogen) and the packaging vector psPAX2 were used for the lentiviral genetic transfer.

Cell culture

MEF cells were cultivated in Glutamax™ Dulbecco's modified Eagle's medium (DMEM, Gibco) supplemented with 10% heat inactivated fetal bovine serum (FBS, Gibco) and 100 U/ml penicillin/100 µg/ml streptomycin at 37 °C with 5% CO_2.

Infection of MEFs with *T. gondii*

Cells were stimulated with 200 U/mL IFNγ (R&D Systems) 16 h before infection. For imaging and measurements, MEFs were cultured in 6-well plates on cover glasses (ø 15 mm, VWR International) and inoculated with freshly harvested *T. gondii* at a ratio of 100:1. To remove extracellular parasites 2 hours post infection, cells were washed with PBS and fixed for downstream processing.

Live cell measurements

Live cell measurements were carried out in Nunc™ LabTek™ II 8-well chambers (ThermoFisher). Cells were seeded in supplemented DMEM medium (as described above) and grown until 70-80% confluence. For live cell measurements, medium was changed to pre-warmed FluoroBrite™ DMEM (Gibco).

Microscopy

Confocal imaging: Confocal imaging with live or fixed cells were performed on a confocal laser scanning microscope FV1000 IX81 inverted microscope, Olympus using a 60x water immersion UPLSAPO NA 1.2 objective. DAPI and eGFP were excited at 405 nm and 488 nm, respectively with the internal FV10-MARAD-2 main laser unit, mCherry was excited at 559 nm with an external Opti λ 559 diode laser (NTT Electronics). Internal PMT detectors (Olympus) were used for detection.

Confocal imaging of mGBP7-mGBP2 VLS: MEF mGBP7$^{-/-}$ cells stably transduced with eGFP-mGBP7 and mCh-mGBP2 were imaged with the settings explained above. Due to a

very bright and high expression levels of mCh-mGBP2, we initially observed "yellow" perinuclear VLS, which were unexpected. To resolve these double positive perinuclear VLS, an optical zoom between 4-7 was used, a resolution of 1024x1024 pixel, a scanning time of 20 μs and the lowest possible laser power for the red diode laser of 0.1%. The resulting image collection is presented in **Figure 3** and nicely shows spatially separated, unmixed mGBP assemblies within the same VLS.

Structured illumination microscopy (SIM): structured illumination microscopy (SIM) was performed on an ELYRA PS.1 (Zeiss) using a Plan-Apochromat 63x/1.4 Oil DIC M27 objective. For imaging of the fluorescent proteins eGFP and mCherry fused to mGBPs in fixed MEF cells, 488 nm was chosen with a BP495-575+LP750 filter for eGFP and 561 nm excitation with a BP570-650+LP750 filter for mCherry, respectively. The resulting images were 75.56 μm x 75.40 μm. During SIM a grid size of 34 μm for eGFP and 42 μm for mCherry was used. Grids were 5 times rotated. Reconstruction was performed using the ZEN structured illumination feature with noise filter of -6.5 for eGFP and -6.2 for mCherry.

Software and Analysis

Image acquisition was performed using the commercial software for the respective microscopes (Zen Black version for Elyra PS.1, Fluoview 1000 version 3 for Olympus IX81).

Our homebuilt software collection Multiparameter Fluorescence Detection and Imaging (MFDI, http://www.mpc.hhu.de/software/software-package.html) was used for fluorescence spectroscopy analysis. The colocalization based analysis (scatterplot generation) was done with Fiji plugin ScatterJ [34]. Particles analysis was done with the ImageJ particle analysis tool [51]. In particular, 'AnI' was used for image based analysis and 'Margarita' was used for subsequent data visualization. Accurate lifetimes were obtained using 'ChiSurf' (https://github.com/Fluorescence-Tools/ChiSurf). Huygens Professional 19.04 was used for chromatic aberration correction in confocal data and for crosstalk correction, stabilization and deconvolution. SIM images were reconstructed using the Zen Black software.

Fluorescence recovery after photo bleaching (FRAP)

MEF cells were seeded in NUNC 8-well chambered slides one day before FRAP measurements. Cells were kept in pre-warmed FluoroBrite™ DMEM (Gibco) during measurement. For FRAP, a bleach region of interest (ROI) of 40x40 pixel was chosen, resulting in bleached areas of 0.5 μm (for mGBP3 containing structures, 11xzoom) and 2 μm (for mGBP2 containing structures, 6xzoom) with a resolution of 256x256 pixel. The bleach ROI was located at the border between cytoplasm and condensate. Both laser lines, 488 nm and 559 nm were simultaneously activated for bleaching. 15 pre-bleach frames were

collected, 10 bleach frames and 985 post-bleach frames with a dwell time of 4 µs/pixel. For analysis, a background ROI (BG) of the same size as the bleach ROI was used to subtract noise from the fluorescence signals S_{GFP} and S_{mCh}. The resulting corrected intensities I_{GFP} and I_{mCh} were plotted versus time. Recovery rate was calculated by the ratio of post-bleach intensities to the average of pre-bleach intensities: Recovery (%) = ($I_{GFP,post-bleach}/ \langle I_{GFP,pre-bleach} \rangle$)*100. No significant photo-bleaching was detected and bleaching efficiency was between 98-100% for both laser lines.

Background corrected recovery curves were fitted by a biexponential model according to equation 1:

$$R(t) = A\left[1 - \exp\left(-\frac{t}{\tau_1}\right)\right] + (1 - A)\left[1 - \exp\left(-\frac{t}{\tau_2}\right)\right] + offset \qquad (1)$$

where A is the fraction of diffusion component 1 and τ_i are the recovery lifetime for the diffusion component 1 and 2, irrespectively. Several models with higher complexity for single component were also evaluated (see **Figure 5-figure supplement 1**).

Confocal MFIS-FRET measurements

<u>**Setup description:**</u> The FV1000 system is additionally equipped with an external time-correlated single photon counting unit (Hydra Harp 400, PicoQuant) and external detectors for MFIS-FRET measurements. eGFP was excited at 485 nm with a LDH-D-C-485 diode laser (PicoQuant) with a linearly polarized beam and a pulse frequency of 32 MHz. mCherry was excited with an Opti λ 559 (NTT Electronics) laser at a continuous wave (CW) setting. MFIS-FRET measurements were done with laser powers of ~ 200 nW at 485 nm and ~500 nW at 559 nm, a pinhole size of 200 µm, a dwell time of 20 µs at a total of 60 frames. Separation into parallel and perpendicular signals was realized by a PBS 101 polarizing beamsplitter cube (Thorlabs). eGFP signal was detected via PD5CTC single-photon avalanche photodiodes (Micro Photon Devices) with upstream HC 520/35 bandpass filters (AHF). Detection of mCherry signal was realized via cooled HMPC-100-40 hybrid-photodetectors (Becker&Hickl) with upstream HC 607/70 bandpass filters (AHF).

<u>**Calibration measurements**</u>: Rhodamine 110 delivered the G-factor $G = Sg_\perp/Sg_\parallel$ for the GFP emission wavelength range (green channels). The G-factor accounts for the detection efficiency difference between detectors of both polarizations $(g_\perp$ and $g_\parallel)$. The instrument response function (IRF) was measured with the back-reflection of the laser beam using a mirror and was used for iterative re-convolution in the fitting process. Furthermore, untransfected cells were measured at 488 nm and 559 nm for background determination.

MFIS-FRET analysis

Multiparameter Fluorescence Lifetime spectroscopy-FRET provides particular advantages to traditional FLIM as all relevant parameters are simultaneously monitored with picosecond accuracy allowing for a multi-step analysis on biological systems [31]. In the first step, pixels were selectively averaged by grouping with the same characteristics. In a second step, the grouped pixels were integrated for subsequent sub-ensemble analysis. This procedure increased the signal-to-noise ratio and ensured high quality data for characterization of structural properties, stoichiometry and interaction affinities.

Pixel-wise analysis: To determine fluorescence-weighted lifetimes in a pixel-wise analysis, the histograms presenting the decay of fluorescence intensity after the excitation pulse were built for each pixel with 128 ps per bin. We used maximum likelihood estimator to determine the experimental anisotropy, and the fluorescence-weighted averaged lifetime of donor molecules $\langle \tau_{D(A)} \rangle_f$ in a single pixel using complex model function according to [6].

2D MFD histograms: To show coordinated changes of FRET indicators (shown in Fig. 5C), we plotted the 2D histograms of donor lifetime $\langle \tau_{D(A)} \rangle_f$ vs the green to yellow fluorescence intensity ratio (F_D/F_A) (see equations (2) and (3)) corrected for crosstalk (characterized by the crosstalk factor α), background $\langle B \rangle$, detection efficiencies of D (g_G) and A (g_Y). The acceptor fluorescence used for 2D-FRET was corrected for additional direct acceptor excitation DE.

$$F_D = \frac{S_G - \langle B_G \rangle}{g_G} \tag{2}$$

$$F_A = \frac{S_Y - (\langle B_Y \rangle + DE) - \alpha(S_G - \langle B_G \rangle)}{g_Y} \tag{3}$$

crosstalk α is determined as the ratio between donor photons detected in the yellow channels and those detected in the green channels for the Donor only sample.

The simultaneous reduction in both FRET indicators $\langle \tau_{D(A)} \rangle_f$ and (F_D/F_A) indicate FRET due to protein interaction. For a given sub-population selection of the donor fluorescence decay histograms with 32 ps time resolution was constructed for further pixel-integrated, sub-ensemble analysis, and the species-averaged fluorescence lifetime of the donor $\langle \tau_{D(A)} \rangle_x$ was calculated based on fit results (species fractions x_i and lifetimes $\tau_{D(A)}$). n is the number of exponents used in donor fluorescence lifetime fitting.

$$\langle \tau_{D(A)} \rangle_x = \sum_{i=1}^{n} x_i * \tau_{D(A),i} \tag{4}$$

Pixel-integrated, time-resolved $\varepsilon_D(t)$ illustration: To identify appropriate pixel in the cells for further pixel-integrated analysis, we computed all fluorescence parameters for each pixel

and selected the pixels in 2D-histograms of several FRET indicators (described above). A pixel population with homogeneous properties was integrated for subsequent pixel-integrated, sub-ensemble analysis. The time-dependent FRET parameter $\varepsilon_D(t)$ contains information on the underlying FRET-rate distribution and is proportional to the probability that FRET occurs at a certain time. After pixel selection, $\varepsilon_D(t)$ was plotted for direct visualization of molecular species with different FRET efficiencies in sub-ensemble data. $\varepsilon_D(t)$ is calculated as the ratio of normalized fluorescence decays of the FRET sample $f_{D(A)}(t)$ and donor-only sample, $f_{D(0)}(t)$ (see eq. 4).

$$\varepsilon_D(t) = \frac{f_{D(A)}(t)}{f_{D(0)}(t)} \tag{5}$$

$\varepsilon_D(t)$ is the probability density function of the occurring FRET governed by FRET rate constant(s), k_{FRET}. The decaying part of $\varepsilon_D(t)$ represents the features of FRET: high- or low-FRET can be directly read out from the decay slope. The amplitude of the decaying part indicates the FRET-active species fraction, x_{FRET}. Accordingly, the offset of $\varepsilon_D(t)$ is the FRET-inactive fraction, $(1 - x_{FRET})$.

Pixel-integrated MFIS-FRET analysis: To determine FRET parameters from pixel-integrated, sub-ensemble data the reference samples were fitted by a multi-exponential relaxation model accounting for a multi-exponential fluorescence decay of the donor in the absence of FRET:

$$f_{D(0)}(t) = \sum_m x_D^{(m)} e^{-t*k_D^{(m)}} \tag{6}$$

in which m =2 considers that FPs in living cells usually show at least a bi-exponential characteristic. Fit parameters in donor decay include three normalized pre-exponential factors $x_D^{(m)}$ $(\sum x_D^{(m)} = 1)$ and two decay rate constants $k_D^{(m)}$, which are the reciprocals of fluorescence lifetimes. The quenched donor decay $f_{D(A)}(t)$ is given by:

$$f_{D(A)}(t) = \sum_m x_D^{(m)} e^{-t*(k_D^{(m)} + k_{FRET})} \tag{7}$$

and k_{FRET} is the FRET rate constant. The fitted parameters in the 1-k_{FRET} model are x_{FRET} and k_{FRET}.

Determination of acceptor and donor concentration from MFIS experiments: To calculate the protein concentrations from fluorescence intensity, the detection volume of our microscope and GFP and mCherry brightness are required. The detection volume was determined as $1.23*10^{-15}$ l from FCS measurements of Cyanine 3B (Cy3B). The fitting model applied to the obtained FCS curve assumes a 3-dimensional Gaussian-shaped volume, and a single diffusing species including transitions to a triplet state as described in [6]. The brightness of enhanced GFP and mCherry in cells were individually characterized from FCS

measurements of freely diffusing FPs in cytoplasm. We found that with 0.6 µW of 559 nm laser excitation at the objective, mCherry brightness is 0.68 kcpm in cytoplasm. With 0.4 µW of 485 nm laser excitation at the objective, GFP brightness is 0.56 kcpm in cytoplasm.

The average mCherry fluorescence intensity of an image with mCherry excitation ($S_{Y,Y}$) was first corrected for detector dead time, and then used to calculate the total concentration of mCherry, $[A]_0$, with the determined detection volume and the mCherry brightness:

$$[A]_0 = \frac{S_{Y,Y}^m}{brightness\ [kcpm] * confocal\ volume\ [fl]}$$

$$= \frac{S_{Y,Y}^m}{0.68\ kcpm * 0.8\ fl}$$

(8)

Assuming the concentration of the FPs reflects the concentration of their host proteins, the mGBP concentration (without non-fluorescent molecules) in µM was determined.

Apparent FRET-based affinity determination, K_D: Affinity of mGBPs was determined based on FRET extracted parameters: x_{FRET} and intensity-based acceptor concentrations by using the following equation 9 implemented in Origin Version 8.6:

$$y = \frac{B_{max} * x}{K_D + x}$$

(9)

5. Chapter 3 References

1. Kim, B. H., Shenoy, A. R., Kumar, P., Das, R., Tiwari, S., MacMicking, J. D. A Family of IFN-gamma-Inducible 65-kD GTPases Protects Against Bacterial Infection. *Science*, 2011, **332**, 717-721. DOI: 10.1126/science.1201711.

2. Tretina, K., Park, E. S., Maminska, A., MacMicking, J. D. Interferon-induced guanylate-binding proteins: guardians of host defense in health and disease. *Journal of Experimental Medicine*, 2019, **216**, 482-500. DOI: 10.1084/jem.20182031.

3. Degrandi, D., Kravets, E., Konermann, C., Beuter-Gunia, C., Klümpers, V., Lahme, S., et al. Murine guanylate binding protein 2 (mGBP2) controls Toxoplasma gondii replication. *Proceedings of the National Academy of Sciences of the United States of America*, 2013, **110**, 294-299. DOI: 10.1073/pnas.1205635110.

4. Finethy, R., Coers, J. Sensing the enemy, containing the threat: cell-autonomous immunity to Chlamydia trachomatis. *Fems Microbiology Reviews*, 2016, **40**, 875-893. DOI: 10.1093/femsre/fuw027.

5. Finethy, R., Jorgensen, I., Haldar, A. K., de Zoete, M. R., Strowig, T., Flavell, R. A., et al. Guanylate binding proteins enable rapid activation of canonical and noncanonical inflammasomes in chlamydia-infected macrophages. *Infection and Immunity*, 2015, **83**, 4740-4749. DOI: 10.1128/iai.00856-15.

6. Kravets, E., Degrandi, D., Ma, Q. J., Peulen, T. O., Klümpers, V., Felekyan, S., et al. Guanylate binding proteins directly attack Toxoplasma gondii via supramolecular complexes. *Elife*, 2016, **5**. DOI: 10.7554/eLife.11479.

7. Laliberte, J., Carruthers, V. B. Host cell manipulation by the human pathogen Toxoplasma gondii. *Cellular and Molecular Life Sciences*, 2008, **65**, 1900-1915. DOI: 10.1007/s00018-008-7556-x.

8. Praefcke, G. J. K. Regulation of innate immune functions by guanylate-binding proteins. *International Journal of Medical Microbiology*, 2018, **308**, 237-245. DOI: 10.1016/j.ijmm.2017.10.013.

9. Saeij, J. P., Frickel, E. M. Exposing Toxoplasma gondii hiding inside the vacuole: a role for GBPs, autophagy and host cell death. *Current Opinion in Microbiology*, 2017, **40**, 72-80. DOI: 10.1016/j.mib.2017.10.021.

10. Vestal, D. J., Jeyaratnam, J. A. The guanylate-binding proteins: emerging insights into the biochemical properties and functions of this family of large interferon-induced guanosine triphosphatase. *Journal of Interferon and Cytokine Research*, 2011, **31**, 89-97. DOI: 10.1089/jir.2010.0102.

11. Degrandi, D., Konermann, C., Beuter-Gunia, C., Kresse, A., Würthner, J., Kurig, S., et al. Extensive characterization of IFN-induced GTPases mGBP1 to mGBP10 involved

in host defense. *Journal of Immunology*, 2007, **179**, 7729-7740. DOI: 10.4049/jimmunol.179.11.7729.

12.   Legewie, L., Loschwitz, J., Steffens, N., Prescher, M., Wang, X., Smits, S. H. J., et al. Biochemical and structural characterization of murine GBP7, a guanylate binding protein with an elongated C-terminal tail. *Biochemical Journal*, 2019, **476**, 3161-3182. DOI: 10.1042/BCJ20190364.

13.   Olszewski, M. A., Gray, J., Vestal, D. J. In silico genomic analysis of the human and murine guanylate-binding protein (GBP) gene clusters. *Journal of Interferon and Cytokine Research*, 2006, **26**, 328-352. DOI: 10.1089/jir.2006.26.328.

14.   Kresse, A., Konermann, C., Degrandi, D., Beuter-Gunia, C., Wuerthner, J., Pfeffer, K., et al. Analyses of murine GBP homology clusters based on in silico, in vitro and in vivo studies. *BMC Genomics*, 2008, **9**. DOI: 10.1186/1471-2164-9-158.

15.   Daumke, O., Praefcke, G. J. K. Mechanisms of GTP hydrolysis and conformational transitions in the dynamin superfamily. *Biopolymers*, 2018, **109**, 1. DOI: 10.1002/bip.23079.

16.   Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 2021, **596**, 583-+. DOI: 10.1038/s41586-021-03819-2.

17.   Peulen, T. O., Hengstenberg, C. S., Biehl, R., Dimura, M., Lorenz, C., Valeri, A., et al. Integrative dynamic structural biology unveils conformers essential for the oligomerization of a large GTPase. *arXiv: Biological Physics*, 2020. DOI: arXiv:2004.04229.

18.   Budde, J. H., Voort, N. T. M. v. d., Felekyan, S., Folz, J., Kühnemuth, R., Lauterjung, P., et al. FRET nanoscopy enables seamless imaging of molecular assemblies with sub-nanometer resolution. *arXiv: Physics Optics*, 2021. DOI: arXiv:2108.00024.

19.   Zhu, S., Bradfield, C. J., Mamińska, A., Park, E., Kim, B., Kumar, P., et al. Cryo-ET of a human GBP coatomer governing cell-autonomous innate immunity to infection. *bioRxiv*, 2021, 2021.08.26.457804. DOI: 10.1101/2021.08.26.457804.

20.   Valadez-Perez, N. E., Benavides, A. L., Scholl-Paschinger, E., Castaneda-Priego, R. Phase behavior of colloids and proteins in aqueous suspensions: theory and computer simulations. *Journal of Chemical Physics*, 2012, **137**. DOI: 10.1063/1.4747193.

21.   Falahati, H., Pelham-Webb, B., Blythe, S., Wieschaus, E. Nucleation by rRNA dictates the precision of nucleolus assembly. *Current Biology*, 2016, **26**, 277-285. DOI: 10.1016/j.cub.2015.11.065.

22. Reinkemeier, C. D., Girona, G. E., Lemke, E. A. Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. *Science*, 2019, **363**, 1415-+. DOI: 10.1126/science.aaw2644.

23. Zhu, L., Brangwynne, C. P. Nuclear bodies: the emerging biophysics of nucleoplasmic phases. *Current Opinion in Cell Biology*, 2015, **34**, 23-30. DOI: 10.1016/j.ceb.2015.04.003.

24. Zwicker, D., Decker, M., Jaensch, S., Hyman, A. A., Jülicher, F. Centrosomes are autocatalytic droplets of pericentriolar material organized by centrioles. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, **111**, E2636-E2645. DOI: 10.1073/pnas.1404855111.

25. Gu, Z. C., Wu, E., Sailer, C., Jando, J., Styles, E., Eisenkolb, I., et al. Ubiquitin orchestrates proteasome dynamics between proliferation and quiescence in yeast. *Molecular Biology of the Cell*, 2017, **28**, 2479-2491. DOI: 10.1091/mbc.E17-03-0162.

26. Xia, S. Y., Chen, Z. H., Shen, C., Fu, T. M. Higher-order assemblies in immune signaling: supramolecular complexes and phase separation. *Protein & Cell*, 2021, **12**, 680-694. DOI: 10.1007/s13238-021-00839-6.

27. Mahserejian, S. M., Scripture, J. P., Mauro, A. J., Lawrence, E. J., Jonasson, E. M., Murray, K. S., et al. Microtubule stutter: a transient dynamic instability phase that is strongly associated with catastrophe. *bioRxiv*, 2020, 2019.12.16.878603. DOI: 10.1101/2019.12.16.878603.

28. Korn, E. D., Carlier, M.-F., Pantaloni, D. Actin polymerization and ATP hydrolysis. *Science*, 1987, **238**, 638-644.

29. Sistemich, L., Kutsch, M., Hamisch, B., Zhang, P., Shydlovskyi, S., Britzen-Laurent, N., et al. The molecular mechanism of polymer formation of farnesylated human guanylate-binding protein 1. *Journal of Molecular Biology*, 2020, **432**, 2164-2185. DOI: 10.1016/j.jmb.2020.02.009.

30. Steffens, N., Beuter-Gunia, C., Kravets, E., Reich, A., Legewie, L., Pfeffer, K., et al. Essential role of mGBP7 for survival of Toxoplasma gondii infection. *mBio*, 2020, **11**. DOI: 10.1128/mBio.02993-19.

31. Weidtkamp-Peters, S., Felekyan, S., Bleckmann, A., Simon, R., Becker, W., Kühnemuth, R., et al. Multiparameter fluorescence image spectroscopy to study molecular interactions. *Photochemical & Photobiological Sciences*, 2009, **8**, 470-480. DOI: 10.1039/b903245m.

32. Widengren, J., Kudryavtsev, V., Antonik, M., Berger, S., Gerken, M., Seidel, C. A. M. Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. *Analytical Chemistry*, 2006, **78**, 2039-2050. DOI: 10.1021/ac0522759.

33. Akrap, N., Seidel, T., Barisas, B. G. Forster distances for fluorescence resonant energy transfer between mCherry and other visible fluorescent proteins. *Analytical Biochemistry*, 2010, **402**, 105-6. DOI: 10.1016/j.ab.2010.03.026.

34. Zeitvogel, F., Schmid, G., Hao, L., Ingino, P., Obst, M. ScatterJ: an ImageJ plugin for the evaluation of analytical microscopy datasets. *Journal of Microscopy*, 2016, **261**, 148-156. DOI: 10.1111/jmi.12187.

35. Greife, A., Felekyan, S., Ma, Q. J., Gertzen, C. G. W., Spomer, L., Dimura, M., et al. Structural assemblies of the di- and oligomeric G-protein coupled receptor TGR5 in live cells: an MFIS-FRET and integrative modelling study. *Scientific Reports*, 2016, **6**. DOI: 10.1038/srep36792.

36. Britzen-Laurent, N., Bauer, M., Berton, V., Fischer, N., Syguda, A., Reipschlager, S., et al. Intracellular trafficking of guanylate-binding proteins is regulated by heterodimerization in a hierarchical manner. *Plos One*, 2010, **5**, 11. DOI: 10.1371/journal.pone.0014246.

37. Ince, S., Kutsch, M., Shydlovskyi, S., Herrmann, C. The human guanylate-binding proteins hGBP-1 and hGBP-5 cycle between monomers and dimers only. *Febs Journal*, 2017, **284**, 2284-2301. DOI: 10.1111/febs.14126.

38. Kravets, E., Degrandi, D., Weidtkamp-Peters, S., Ries, B., Konermann, C., Felekyan, S., et al. 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*, 2012, **287**, 27452-66. DOI: 10.1074/jbc.M112.379636.

39. Ghosh, A., Praefcke, G. J., Renault, L., Wittinghofer, A., Herrmann, C. How guanylate-binding proteins achieve assembly-stimulated processive cleavage of GTP to GMP. *Nature*, 2006, **440**, 101-4. DOI: 10.1038/nature04510.

40. Oates, M. E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M. J., Xue, B., et al. (DP2)-P-2: database of disordered protein predictions. *Nucleic Acids Research*, 2013, **41**, D508-D516. DOI: 10.1093/nar/gks1226.

41. Kunzelmann, S., Molekularer Mechanismus der GTP-Hydrolyse durch das humane Guanylat-bindende Protein 1, *Faculty of Chemistry and Biochemistry*. 2007, Ruhr-University Bochum: Bochum.

42. Wehner, M., Kunzelmann, S., Herrmann, C. The guanine cap of human guanylate-binding protein 1 is responsible for dimerization and self-activation of GTP hydrolysis. *Febs Journal*, 2012, **279**, 203-210. DOI: 10.1111/j.1742-4658.2011.08415.x.

43. Völpel, T., Hengstenberg, C. S., Peulen, T. O., Ajaj, Y., Seidel, C. A. M., Herrmann, C., et al. 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*, 2014, **53**, 4590-4600. DOI: 10.1021/bi500524u.

44.   Xavier, A., Al-Zeer, M. A., Meyer, T. F., Daumke, O. hGBP1 coordinates chlamydia restriction and inflammasome activation through sequential GTP hydrolysis. *Cell Reports*, 2020, **31**. DOI: 10.1016/j.celrep.2020.107667.

45.   Qin, A. P., Lai, D. H., Liu, Q. F., Huang, W. J., Wu, Y. P., Chen, X. Y., et al. Guanylate-binding protein 1 (GBP1) contributes to the immunity of human mesenchymal stromal cells against Toxoplasma gondii. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, **114**, 1365-1370. DOI: 10.1073/pnas.1619665114.

46.   Bolognesi, B., Lehner, B. Reaching the limit. *Elife*, 2018, **7**. DOI: 10.7554/eLife.39804.

47.   Li, G. D., Segu, V. B. G., Rabaglia, M. E., Luo, R. H., Kowluru, A., Metz, S. A. Prolonged depletion of guanosine triphosphate induces death of insulin-secreting cells by apoptosis. *Endocrinology*, 1998, **139**, 3752-3762. DOI: 10.1210/en.139.9.3752.

48.   Huang, S., Zhu, S. W., Kumar, P., MacMicking, J. D. A phase-separated nuclear GBPL circuit controls immunity in plants. *Nature*, 2021, **594**, 424-+. DOI: 10.1038/s41586-021-03572-6.

49.   Harmon, T. S., Holehouse, A. S., Rosen, M. K., Pappu, R. V. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *Elife*, 2017, **6**. DOI: 10.7554/eLife.30294.

50.   Weber, S. C., Brangwynne, C. P. Getting RNA and protein in phase. *Cell*, 2012, **149**, 1188-1191. DOI: 10.1016/j.cell.2012.05.022.

51.   Schneider, C. A., Rasband, W. S., Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 2012, **9**, 671-675. DOI: 10.1038/nmeth.2089.

Chapter 3 Supplement

DATA FINDER:

A) For raw data refer to
P:\Projects\mGBP7-Project_2017\data

B) For analysed data refer to
P:\Projects\mGBP7-Project_2017\Analysis

C) For image files and text refer to
P:\Projects\mGBP7-Project_2017\mGBP7 manuscript

Figure 1-table supplement 1

Comparison between published mGBP7 colocalization data (Steffens et al., 2020) presented by PCC values and quantitative changes in the donor fluorescence lifetime extracted from our FRET study presented in **Figure 1**. For simplification, PCC values below 0.1 were considered to present no relationship (-) between the observables, below 0.3 negligible/weak (-), below 0.4 moderate (+) and between 0.7-1 very strong relationship (++).

mGBP7 partner	PCC (-/+ INF treatment)*	Relationship between observables	Simplification	Qualitative FRET judged by lifetime changes
mGBP1	~0.1-0.3	Negligible, weak	-	No
mGBP2	~0.1-0.3	Negligible, weak	-	No
mGBP3	~0.7	Very strong	++	Yes
mGBP5	~0.05	No	-	No
mGBP6	~0.3-0.4	Moderate	+	No
mGBP7	~0.85	Very strong	++	Yes

Figure 1-table supplement 2

MFIS-FRET t-test results. Hypothesis tested: *Donor-acceptor data are different from donor-only-data*. Two-tailed hypothesis. Significance levels: .05<*p* = N.S., .01<*p*<.05 = *, .001<*p*<.01 = **, *p*<.001 = ***.

Species (+mGBP7)	*p* value by compartment		
	CP	VLS	PVM
mGBP1	.11	.24	.17
mGBP2	.27	.87	.027
mGBP3	.55	<.0001	<.0001
mGBP5	.21	.030	.40
mGBP6	.77	.013	.01
mGBP7	<.0001	<.0001	<.0001
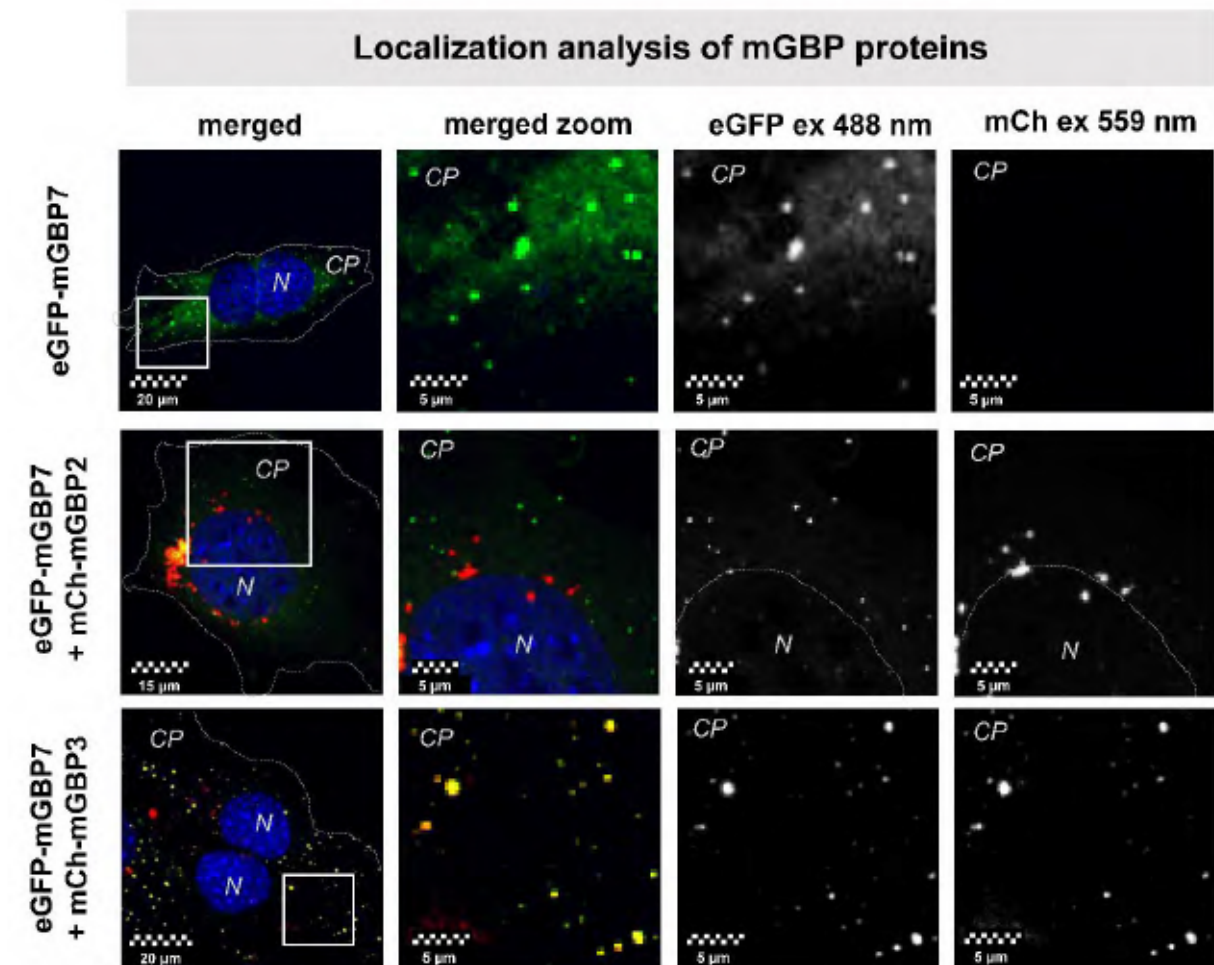
Figure 2-figure supplement 1: Characterization of mGBP7 vesicle like structures in absence or presence with further mGBP members. The nucleus was counterstained with DAPI (blue). **Top row**: eGFP-mGBP7 (green) distribution and localization. eGFP-mGBP7 appeared as small, globular structures of high intensity, as well as green background. **Middle row**: In presence of mCherry-mGBP3 (red), both proteins almost completely co-localized in globular structures (yellow in merged channel). **Bottom row**: In presence of mCherry-mGBP2 (red), both proteins are present perinuclear in very large structures (overview image). Furthermore, both proteins were independently distributed all over the cell and appeared as small globular structures (zoom in). All images were taken with the inverse confocal fluorescence microscope Olympus Fluoview 100 by using the 60x water NA1.2 objective. The Laser intensity for all channels was below 1%. Abbreviations: CP = cytoplasm, N= nucleus.
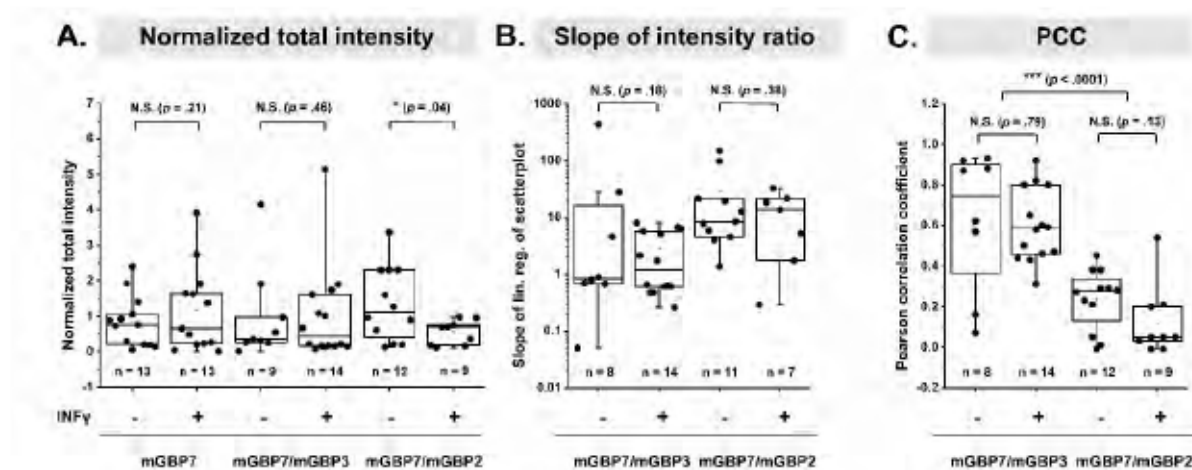
Figure 2-figure supplement 2: PCC and test for INFγ dependency.

MEF mGBP7$^{-/-}$ transduced with eGFP-mGBP7 and mCh-mGBP2 or mGBP3 were tested for (**A.**) differences in intensity, (**B.**) slope of intensity ratio and (**C.**) Pearson Correlation Coefficient. Cells were randomly selected with variable expression levels. No significant difference was found between mGBPs with and without INFγ. Highly significant difference between PCCs of mGBP7-mGBP2 (median approx. 0.05-0.25) and mGBP7-mGBP3 (median approx. 0.6-0.75). T-test results: Hypothesis 1 tested (**A., B.**): *Data of one combination with and without INFγ is different.* Hypothesis 2 tested (**C.** only): *PCC of mGBP7/2 different from mGBP7/3.* Two-tailed hypothesis. Significance level: .05<p = N.S., .01<p<.05 = *, .001<p<.01 = **, p<.001 = ***.

Figure 2-table supplement 1: PCC from Figure 2-figure supplement 2.

MEF mGBP7$^{-/-}$ transduced with eGFP-mGBP7 and mCh-mGBP2 or mGBP3. Pearson's correlation coefficient determined with ImageJ Coloc2 tool (Schneider, Rasband, & Eliceiri, 2012).

Combination:	mGBP7 with mGBP3		mGBP7 with mGBP2	
INFg	- INFg	+ INFg	- INFg	+ INFg
	0.93	0.6	0.01	-0.01
	0.92	0.31	0.38	0.21
	0.88	0.58	0.29	0.05
	0.87	0.82	0.21	0.05
	0.62	0.92	0.29	0.05
	0.57	0.8	0.45	-0.01
PCC (cell-wise)	0.16	0.46	0.38	0.54
	0.07	0.8	0.05	0.03
		0.65	0.23	0.2
		0.43	-0.01	
		0.47	0.28	
		0.5	0.27	
		0.44		
		0.59		
Mean PCC	0.6275	0.59785714	0.23583333	0.12333333
STDDEV	0.32255813	0.17296484	0.14232933	0.16579773

Figure 3-table supplement 1: Size distribution of mGBP7 in absence or presence of mGBP2 and mGBP3. mGBP7 alone and in presence of mGBP2 appears as very small µm sized particles. In presence of mGBP3, both particles are of similar size with approximately three times larger particles than mGBP7 in absence of mGBP3. mGBP2 forms on average the largest particles.

Cell line	eGFP particle		mCherry particle	
	mean size and STDDEV (µm²)	Mean number and STDDEV	mean size and STDDEV (µm²)	Mean number and STDDEV
eGFP-mGBP7	0.13 ± 0.34	344	-	-
eGFP-mGBP7 and mCherry-mGBP3	0.37 ± 0.61	736	0.26 ± 0.44	717
eGFP-mGBP7 and mCherry-mGBP2	0.16 ± 0.80	152 ± 72	0.83 ± 3.73	170 ± 113
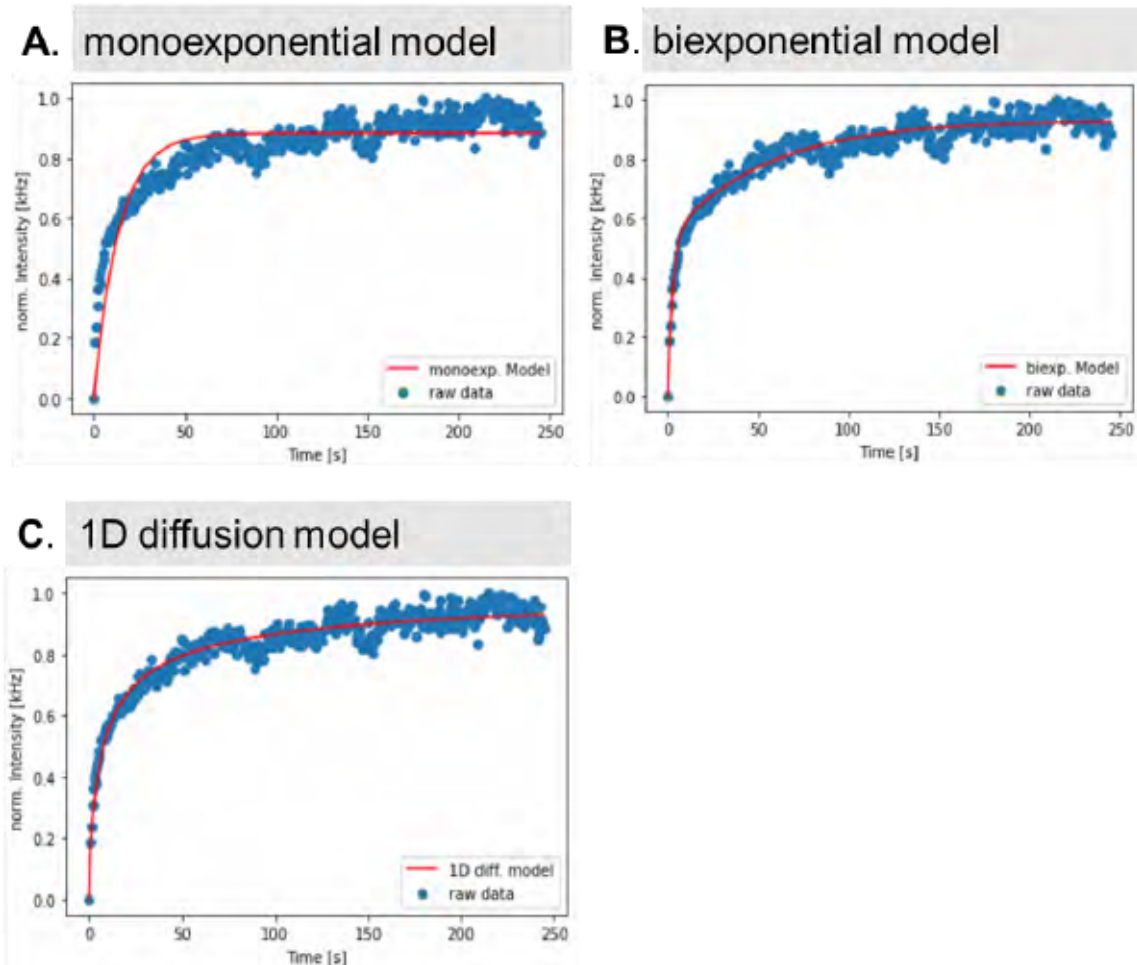
Figure 5-figure Supplement 1: FRAP model evaluation. In order to make a statement regarding the protein dynamics in the respective dense phase (VLS), fluorescence after photobleaching (FRAP) was applied. Different diffusion models were tested, allowing extracting characteristic recovery, which can be used to get specific information about the diffusion inside the sample or interaction with the direct environment. An exemplary background corrected FRAP curve of eGFP from a eGFP-mGBP7/mCh-mGBP3 sample was used for FRAP model evaluation. Different diffusion models with increased complexity are shown in (**A.**) mono-exponential model (see methods **eq. (1)**), (**B.**) bi-exponential (methods **eq. (1)**) and (**C.**) pure 1D-diffusion model (Kang, Day, Kenworthy, & DiBenedetto, 2012; Soumpasis, 1983). Data was insufficiently described by the mono-exponential model and 1D diffusion model, but a bi-exponential model yielded adequate results and was used for data analysis shown in **Figure 5B/C**.
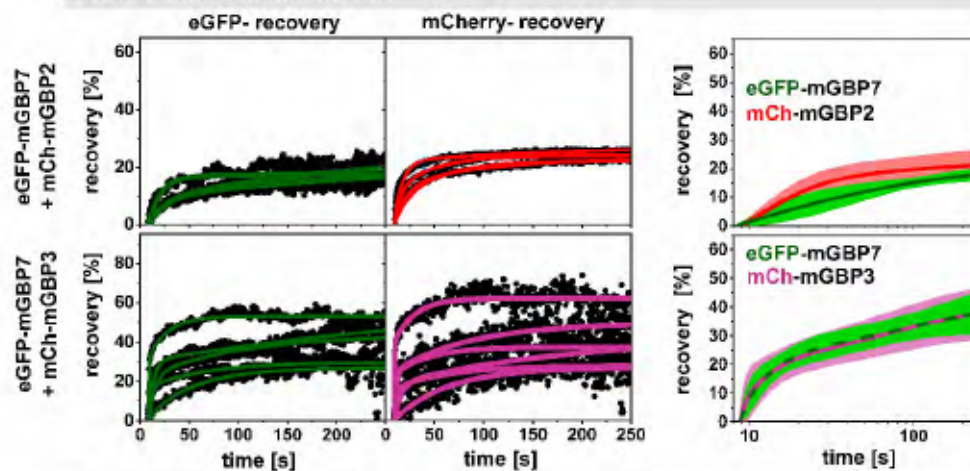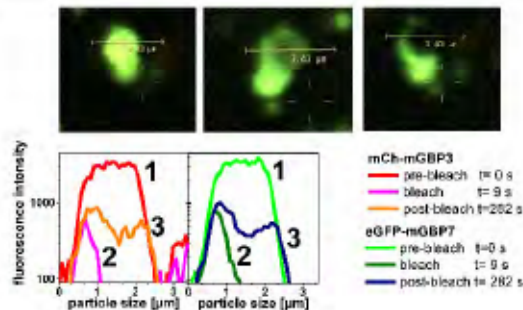
Figure 5-figure supplement 2: Recorded FRAP curves of mGBPs. A) left plots: Background corrected FRAP curves of eGFP-mGBP7 (green, top panel) and mCh-mGBP2 (red, top)/ mCh-mGBP3 (purple, bottom) after INFγ treatment. Right plots: log scale presentation of data clearly shows ongoing very low recovery in mGBP7/mGBP2 and mGBP7/mGBP3 VLS. **B)** Fluorescence intensity recovery in VLS at different time steps: Fluorescence intensity profile (3.43 μm) through a VLS of eGFP-mGBP7 (right panel) and mCh-mGBP3 (left panel) at different time steps (1) prebleached, (2) bleached and (3) postbleached. A faster intensity recovery at the periphery of the VLS indicates a hindered diffusion. Note, movement of the VLS and fusion of smaller VLS with the larger one during the measurement.

LG-domain alignment of hGBP1, mGBP2, mGBP3 and mGBP7


```
                                        [Ploop]                    [Switch 1]
                                        ++++*****                  +++++***
                                           *
>hGBP1    1  MASEIHMTGPMCLIENTNGRLMANPRALKILSAITQPMVVVAIVGLYRTGKSYLMNKLAGKKKGFSLGSTVQSHTKGIWMWCVPHP
>mGBP2    1  MASEIHMSPPMCLIENTRAQLVINQEALRILSAITQP VVVAIVGLYRTGKSYLMNKLAGKRTGFSLGSTVQSHTKGIWMWCVPHP
>mGBP3    1  M------EAPICLVENWKNQLTVNLEAIRILEQIAQP VVVAIVGLYRTGKSYLMNRLAGRNNGFSLGSTVQSETKGIWMWCVPHP
>mGBP7    1  MASGPNMEAPVCLVENENEELRVNEKAINILERITQP VVVAIVGLYRTGKSYLMNRLAGQNHGFNLGTTVRSETKGIWMWCVPHP

                   [Switch 2]
                   *++ ++++ +                    *     +++++  ++
>hGBP1   87  KKPGHILVLLDTEGLGDVEKGDNQNDSWIFALAVLLSSTFVYNSIGTINQQAMDQLYYVTELTHRIESKSSPDEHENEVEDSADFV
>mGBP2   87  KKAGQTLVLLDTEGLEDVEKGDNQNDCWIFALAVLLSSTFIYNSIGTINQQAMDQLHYVTELTDLIKSKSSPDQ--SGVDDSANFV
>mGBP3   81  TKPTHTLVLLDTEGLGDVEKGDPKNDSWIFALAVLLSSTFVYNSMSTINQQALEQLHPVTELTQLIRAKSSPRE-D-KVKDSSEFV
>mGBP7   87  SKPKFTLVLLDTEGLGDVEKGDPKNDSWIFALAVLLSSTFVYNSMSTINHQALEQLHYVTELTERIPAKSTSRS-E-EVDDSDRFV

                                                                              [Guanine Cap]
                  *                                                       *      **
                  *++++++ +++        +   +      +          +   +          +++ + +++++ *
>hGBP1  173  SFFPDFVWTLRDFSLDLEADGQPLTPDEYLTYSLKLKKGTSQKDETFNLPRLCIRKFFPKKKCFVFDRPVH-RRKLAQLEKLQDEE
>mGBP2  171  GFFPTFVWTLRDFSLELEVNGKPVTSDEYLEHSLTLKKGADKKTKSFNKPRLCIRKFFPKRKCFIFDRPAQ-HKQLSKLETLREEE
>mGBP3  165  GFFPDFIWAVRDFVLELKLNGRPITEDEYLENALKLIQGDHLKVQQSNMTRECIRYFFPVRKCFVFDRPTSDKRLLLQIENVPENQ
>mGBP7  171  SFFPDFIWTVRDFVLELKLEGRVITADEYLENALKLIIGNSIKAQKANLPRECIRHFFPHRKCFVFDRPTKDKELLVHVEEMPEDQ

             + +
>hGBP1  258  LDPEFVQQVADFCSYIFSNSKTKTLSGGIQVNGPRLESLVLTYVNAISSGDLPCMENAVL  317
>mGBP2  256  LCGEFVEQVAEFTSYILSYSSVKTLCGGIIVNGPRLKSLVQTYVGAISNGSLPCMESAVL  315
>mGBP3  251  LERNPQVESEKFCSYIPTNGKTKTLRGGVIVIGNRLGTLVQTYVNAINSGTVPCLENAVT  310
>mGBP7  257  LDHSFQVQSKEFCSYIFSNSKAKTLKEGIVVNGNRLAYLVTTYVLAINSGDVPCLENAVT  316
```

+ Dimer interface residues determined for hGBP1 (Ghosh et al., 2006)

* Catalyticly important residues determined for hGBP1 (Ghosh et al., 2006)
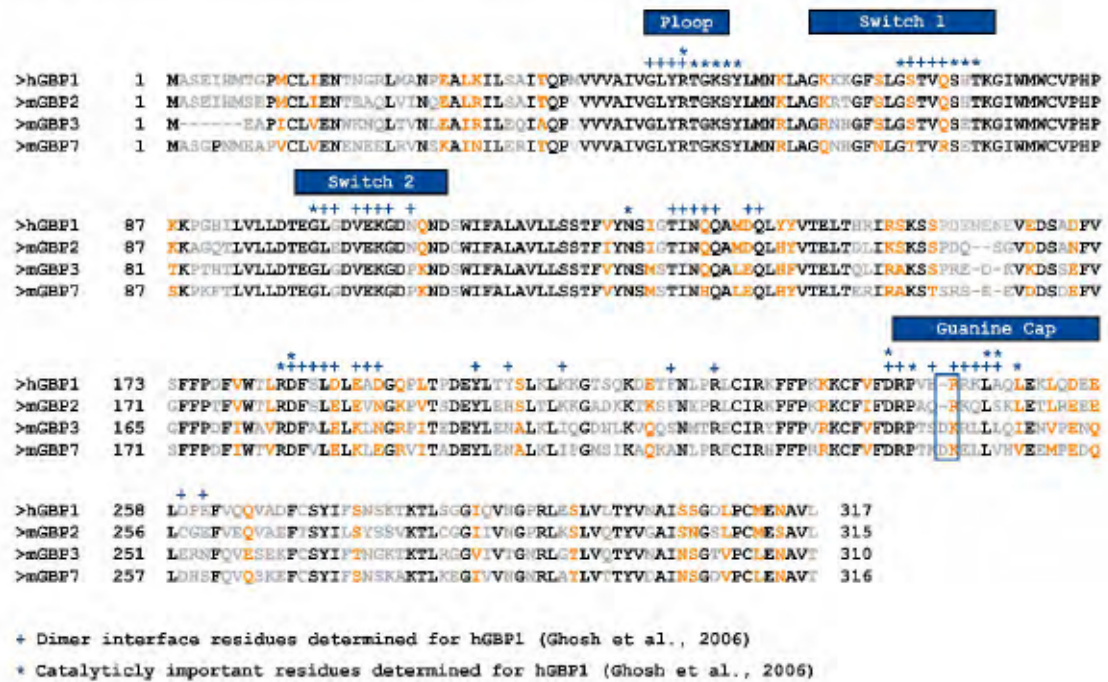
Figure 6-figure supplement 1: Sequence alignment of the LG-domains of hGBP1 (Ghosh, Praefcke, Renault, Wittinghofer, & Herrmann, 2006), mGBP2, mGBP3 and mGBP7 (Kresse et al., 2008). Important regions are highlighted above. The dimer interface residues marked as 'plus' (+), catalytically important residues marked as 'star' (*). Identical residues marked in black, similar residues in orange and different residues in grey.

Figure 6-table supplement 1: Polar contacts of the dimer interface of hGBP1 and corresponding residues of mGBP7, mGBP2 and mGBP3 extracted from PDB 2b92, 2bc9 and 2b8w. Contacts were identified by the DSSP-based (Kabsch & Sander, 1983) standard format of the software PyMOL ("The PyMOL Molecular Graphics System, Version 1.2r3pre,"). Likeliness of bond after mutation was predicted by BLOSUM substitution matrix (score lower than 0 means 'not likely') (Zomaya, 2006) and crystal structure alignment of mGBP7 model 1 (Legewie et al., 2019) and hGBP1 (PDB 1f5n). "Side" means sidechain interaction, "back" means backbone interaction.

Extracted from PDB entry:	hGBP1-hGBP1	Type	mGBP7-mGBP2	Type	mGBP7-mGBP3	Type	mGBP7-mGBP7	Type
2b92, 2bc9, 2b8w	Y47-I134	Side-back	Y47-I134	Side-back	Y47-I128	Side-back	Y47-I134	Side-back
2b92, 2b8w	S69-D188	Back-back	T69-E186	Back-back	T69-E180	Back-back	T69-E186	Back-back
2bc9	S69-D188	Side-Side	T69-E186	Back-back	T69-E180	Back-back	T69-E186	Back-back
2b92	T70-E190	Side-back	T70-E188	Side-back	T70-K182	Side-back	T70-K188	Side-back
2b92, 2b8w	V71-D188	Back-back	V71-E186	Back-back	V71-E180	Back-back	V71-E186	Back-back
2b92, 2bc9, 2b8w	Q72-D192	Side-side	R72-N190	Side-side	R72-N184	Side-side	R72-E190	Side-side
2b8w, 2b8w	Q72-D192	Side-side	R72-N190	Side-side	R72-N184	Side-side	R72-E190	Side-side
2b92, 2bc9, 2b8w	Q72-D192	Side-back	R72-N190	Side-back	R72-N184	Side-back	R72-E190	Side-back
2b92, 2b8w	Q72-Y204	Side-side	R72-H202	Side-side	R72-N196	Side-side	R72-N202	Side-side
2b92, 2bc9, 2b8w	G102-Q136	Back-back	G102-Q136	Back-back	G102-Q130	Back-back	G102-H136	Back-back
2b92, 2bc9, 2b8w	V104-R223	Back-side	V104-R221	Back-side	V104-R215	Back-side	V104-R221	Back-side
2bc9	E105-Y204	Side-side	E105-H202	Side-side	E105-N196	Side-side	E105-N202	Side-side
2b92, 2b8w	N109-D140	Side-side	P109-D140	**Not likely**	P109-E134	**Not likely**	P109-E140	**Not likely**
2b92, 2bc9, 2b8w	I134-Y47	Back-side	I134-Y47	Back-side	I134-Y41	Back-side	I134-Y47	Back-side
2b92, 2bc9, 2b8w	Q136-G102	Back-back	H136-E102	Back-back	H136-G96	Back-back	H136-G102	Back-back
2b92, 2b8w	Q137-Q141	Side-side	Q137-Q141	Side-side	Q137-Q135	Side-side	Q137-Q141	Side-side
2b92	D140-N109	Side-	E140-	Side-	E140-	**Not**	E140-P109	**Not likely**

		side	N109	side	P103	likely		
2b92, 2b8w	Q141-Q137	Side-side	Q141-Q137	Side-side	Q141-Q131	Side-side	Q141-Q137	Side-side
2b92, 2bc9	D184-R240	Back-side	D182-R238	Back-side	D182-R232	Back-side	D182-R238	Back-side
2bc9	F185-R244	Back-side	F183-R242	Back-side	F183-K237	Not likely	F183-K242	Not likely
2b8w	S186-R240	Side-side	V184-R238	Not likely	V184-R232	Not likely	V184-R238	Not likely
2b8w	S186-L247	Back-back	S186-L245	Back-back	S186-L240	Back-back	S186-L246	Back-back
2b92, 2bc9, 2b8w	D188-S69	Back-back	E186-S69	Back-back	E186-S63	Back-back	E186-T69	Back-back
2bc9, 2b8w	D188-S69	Side-side	E186-S69	Side-side	E186-S63	Side-side	E186-T69	Side-side
2b92, 2b8w	D188-V71	Back-back	E186-V71	Back-back	E186-V65	Back-back	E186-V71	Back-back
2b8w	D188-K246	Side-side	E186-Q244	Side-side	E186-L239	Not likely	E186-L245	Not likely
2b92, 2bc9, 2b8w	D192-Q72	Side-side	E190-Q72	Side-side	E190-Q66	Side-side	E190-R72	Side-side
2b92, 2bc9, 2b8w	D192-Q72	Back-side	E190-Q72	Back-side	E190-Q66	Back-side	E190-R72	Back-side
2b92, 2b8w	Y204-Q72	Side-side	N202-Q72	Side-side	N202-Q66	Side-side	N202-R72	Side-side
2bc9	Y204-E105	Side-side	N202-E105	Side-side	N202-E99	Side-side	N202-E105	Side-side
2b92, 2bc9	R223-V104	Side-back	R221-V104	Side-back	R221-V98	Side-back	R221-V104	Side-back
2bc9	D239-R244	Side-side	D237-R242	Side-side	D237-K237	Not likely	D237-K243	Not likely
2b92, 2b8w	D239-R245	Side-side	D237-K243	Side-side	D237-R238	Not likely	D237-E244	Not likely
2b92	D239-R245	Side-side	D237-K243	Not likely	D237-R238	Not likely	D237-E244	Not likely
2b92, 2bc9	R240-D184	Side-back	R238-D182	Side-back	R238-D176	Side-back	R238-D182	Side-back
2b92, 2bc9	R240-S186	Side-side	R238-S184	Side-side	R238-A178	Not likely	R238-V184	Not likely
2b92	H243-R244	Side-side	K241-R242	Not likely	K241-K237	Not likely	K241-K243	Not likely
2bc9	R244-F185	Side-back	K243-F183	Not likely	K243-F177	Not likely	K243-F183	Not likely
2bc9	R244-D239	Side-side	K243-D237	Not likely	K243-D237	Not likely	K243-D237	Not likely
2b92	R245-D239	Side-side	E244-D237	Not likely	E244-D231	Not likely	E244-D237	Not likely
2b8w	R245-D259	Side-side	E244-C257	Not likely	E244-E252	Not likely	E244-D258	Not likely
2b92	R245-E261	Side-side	E244-E259	Not likely	E244-N254	Side-side	E244-S260	Side-side
2b8w	L247-S186	Back-back	L246-S184	Back-back	L246-A178	Back-back	L246-V184	Back-back
2b8w	D259-R245	Side-	D258-	Side-	D258-	Not	D258-E244	Not likely

		side	K243	side	R238	likely		
2b8w	D259-R245	Side-side	D258-K243	**Not likely**	D258-R238	**Not likely**	D258-E244	**Not likely**
2b8w	D259-R245	Side-side	D258-K243	**Not likely**	D258-R238	**Not likely**	D258-E244	**Not likely**

SI references:

Ghosh, A., Praefcke, G. J., Renault, L., Wittinghofer, A., & Herrmann, C. (2006). How guanylate-binding proteins achieve assembly-stimulated processive cleavage of GTP to GMP. *Nature, 440*(7080), 101-104. doi:10.1038/nature04510

Kabsch, W., & Sander, C. (1983). DICTIONARY OF PROTEIN SECONDARY STRUCTURE - PATTERN-RECOGNITION OF HYDROGEN-BONDED AND GEOMETRICAL FEATURES. *Biopolymers, 22*(12), 2577-2637. doi:10.1002/bip.360221211

Kang, M., Day, C. A., Kenworthy, A. K., & DiBenedetto, E. (2012). Simplified Equation to Extract Diffusion Coefficients from Confocal FRAP Data. *Traffic, 13*(12), 1589-1600. doi:10.1111/tra.12008

Kresse, A., Konermann, C., Degrandi, D., Beuter-Gunia, C., Wuerthner, J., Pfeffer, K., & Beer, S. (2008). Analyses of murine GBP homology clusters based on in silico, in vitro and in vivo studies. *BMC Genomics, 9.* doi:10.1186/1471-2164-9-158

Legewie, L., Loschwitz, J., Steffens, N., Prescher, M., Wang, X., Smits, S. H. J., . . . Pfeffer, K. (2019). Biochemical and structural characterization of murine GBP7, a guanylate binding protein with an elongated C-terminal tail. *Biochemical Journal, 476*(21), 3161-3182. doi:10.1042/BCJ20190364

. The PyMOL Molecular Graphics System, Version 1.2r3pre. Schrödinger, LLC.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods, 9*(7), 671-675. doi:10.1038/nmeth.2089

Soumpasis, D. M. (1983). THEORETICAL-ANALYSIS OF FLUORESCENCE PHOTOBLEACHING RECOVERY EXPERIMENTS. *Biophysical Journal, 41*(1), 95-97. doi:10.1016/s0006-3495(83)84410-5

Steffens, N., Beuter-Gunia, C., Kravets, E., Reich, A., Legewie, L., Pfeffer, K., & Degrandi, D. (2020). Essential Role of mGBP7 for Survival of Toxoplasma gondii Infection. *mBio, 11*(1). doi:10.1128/mBio.02993-19

Zomaya, A. Y. (2006). *Handbook of Nature-Inspired and Innovative Computing* (1 ed.): Springer.

Chapter 4: Human KDM6A/UTX is a multifaceted nuclear protein, modifying and interacting at chromatin structures and beyond

Chapter digest

Aims	Methods	Outcome
Analyze demethylase activity, interaction ability, cellular localization, protein stability of KDM6A variants and observe the relation to cellular phenotypes.	ELISA, IP/western blot, FACS, MFIS-FRET, SR/confocal imaging & image analysis	Truncated variants of KDM6A cause significant levels of cellular damage and apoptosis, while substitution variants only show mild effects.

Contributions

This chapter includes multiple sections taken from **Koch et al (2021)**. Affected sections are highlighted accordingly. See **Publications** section (Page 200) for a detailed overview on contributions to the published sections.

Additional, unpublished experiments and data analysis were done by *Julian Koch* and *Annemarie Greife*.

Abstract

In the past decade, multiple deep sequencing and pan-cancer studies revealed high mutation frequencies in genes coding for histone-modifying proteins, with the KDM6A gene being one of the most prominent targets, particularly in urothelial cancer. The main-functionality of KDM6A is the demethylation of H3K27me2/3 and interaction with transcription factors at chromatin structures. However, there might be additional demethylase independent functions, reported in literature and by our own recent findings. Structurally, KDM6A has a demethylase domain (JmjC) and a N-terminal multi-interaction domain (TPR) connected by a largely disordered linker region (IDR). Due to the high pan-cancer occurrence of mutations in KDM6A, many cancer related studies exist but very few stand out by investigating the structure-function relationship. Consequentially, we aimed to establish a clean correlation between the molecular features: We systematically characterized substitution and truncation variants of KDM6A in respect to their demethylase activity, interaction ability, cellular localization, protein stability and the observed cellular phenotypes. Demethylase activity depends on the mutation status of the JmjC-domain. Interaction ability with Retinoblastoma-binding protein 5 (RBBP5) depends on the presence of the TPR-domain and to a lesser amount to the presence of the other domains. Cellular localization was mainly nuclear for all variants, with the tendency of the truncated variants to accumulate at extranuclear DNA

together with Nucleophosmin (NPM1). Protein stability was lower in truncated variants. Quantification of the resulting cellular phenotypes revealed severe effects caused by whole-domain deletion tends to cause the most severe effects, resulting in cellular damage and apoptosis. Substitution variants showed only mild effects. These findings hint towards a strong interplay between the two functional domains and possibly even an involvement of the linker region. By further expanding our findings, we hope to increase biomolecular understanding of KDM6A in its environment and contribute valuable knowledge towards medical research, especially in the field of targeted therapy of urothelial cancer.

1. Introduction

The tightly controlled and selective activation and deactivation of genes is a crucial mechanism maintaining basic cellular functions as well as cell cycle [1], proliferation [2] and tissue specificity [3]. While regulating genes, preventing uncontrolled spreading of anomalous cells is crucial, since this sequence of events will fatally lead to cancer development and growth [4, 5]. Biochemically, regulation of genes is largely controlled by reversible modifications at the histone tails, which will tune the nucleosomal DNA towards a desired state [6, 7]. The states are described as either active, reversibly inactive or irreversible repressed. These modifications occur at all existing histone tails and consist of methylation, acetylation, phosphorylation, ubiquitination and citrullination [8]. Historically, KDM6A was discovered and introduced as a histone H3K27-specific demethylase [9-11]. Related research was mostly focused on its enzymatic activity, although some findings already suggested that KDM6A contains an interaction domain and might also exert demethylase-independent functions [12-14]. As of today, only the catalytically active **JmjC-domain** is available as crystal structure [15]. Intriguingly, recent advances in deep learning algorithms [16] enabled the prediction of the full structure (**Fig. 1A**) based on the sequence (**Fig. 1B**) of KDM6A [11]. The C-terminal part of KDM6A contains the JmjC demethylase domain [15]. The structure of the JmjC-domain (PDB: 3AVR) [15] and the displayed AlphaFold-structure (**Fig. 1A**) is concurrent. The JmjC-domain is able to specifically recognize H3K27me2 or me3 marks [17] and demethylates them towards lower methylated states [18]. The H3K27me2/3 marks are repressive chromatin modifications [7]. Disposal of this modification will enhance gene accessibility, mostly in combination with gain of the K27ac and K4me3 marks [6, 19-21]. Concurrently, the JmjC-domain acts as interaction interface for Histone H3 binding, no further JmjC-specific interactions were proven so far. The few published physical interactions focus mainly on the N-terminal part of KDM6A [22, 23], consisting of an eight helices long tetratricopeptide repeat **TPR-domain** [15] and a preceding low-complexity region. The TPR forms a grooved structure as interaction interface, forming a highly specific topology that usually enables multiple binding sites in one domain [24, 25]. As of today, there is no crystal structure available for the TPR of KDM6A. The

displayed TPR (**Fig. 1A**) predicted by AlphaFold is in good agreement with a prototypic TPR-8 structure [26]. Topologically, it is predicted to be in weak contact with the JmjC. It aligns next to it if the molecule is in an unbound, collapsed state. The JmjC and TPR domains are connected by an unstructured and intrinsically disordered **IDR-domain**. There is no structural or dynamical information available for the IDR. The IDR-structure displayed by AlphaFold (**Fig. 1A**) has a very low per-residue confidence score (pLDDT) [16]. The exact topology of the assembly of the IDR around the structured domains was determined by a few weak contacts and the rigidity displayed is highly misleading. In the unbound, collapsed state of KDM6A, the IDR will be under constant dynamic motion as indicated by the very low pLDDT. Although there are no apparent functionalities assigned to this linker other than an increased solubility and domain connectivity, linker behavior and post-translational modifications at the TPR could have an influence on the rest of the protein.

While publications on KDM6A addressing medical or genetic issues usually excel in genome analysis [27], cellular [28] or animal models [14, 29] and therapeutic options [28, 30], they often disregard the biomolecular origin of the results. Additionally, some studies have been conflicting, characterizing KDM6A both as a tumor suppressor [31, 32] and oncogene [33-35]. KDM6A variants in urothelial cancer feature nonsense and missense mutations distributed over the whole gene [27, 36], although there are mutational hotspots found in COSMIC v92 (GRCh 38, November 2020). COSMIC (Catalogue Of Somatic Mutations In Cancer) is a comprehensive online database of somatically acquired mutations in cancer. The three most frequent point mutations across all tissues are the synonymous mutation Q1037= (c.3111G>A) in the JmjC domain, the missense mutation T726K (c.2177C>A) and the truncating Q555* (c.1663C>T), both in the intrinsically disordered region. Other than that, there is no particular preference for any domain in any type of cancer or the hereditary Kabuki Type II syndrome.

We therefore generated a broad selection of 'representative' variants to include every part of the protein and address all 'core functions' of KDM6A (**Fig. 2A**). We selected two urothelial cancer cell lines, T-24 (KDM6A mut.) [37-40] and SW-1710 (KDM6A WT) [41] as systems since the cell lines might cope differently with the overexpression of either KDM6A WT or the variants. We subsequently established methods (**Fig. 2B**) to explore multiple aspects of the variants possibly connected to a malignant function of KDM6A in cancer, namely localization [42], interaction [43, 44] and activity [45].

Since nuclear import and localization is crucial for any predominantly nuclear protein, it can be easily disturbed if the import signal is missing and the protein is unable to pass the nuclear pore complex [46]. This can lead to damage by accumulation of the protein in wrong compartments and its absence at the destination [47].
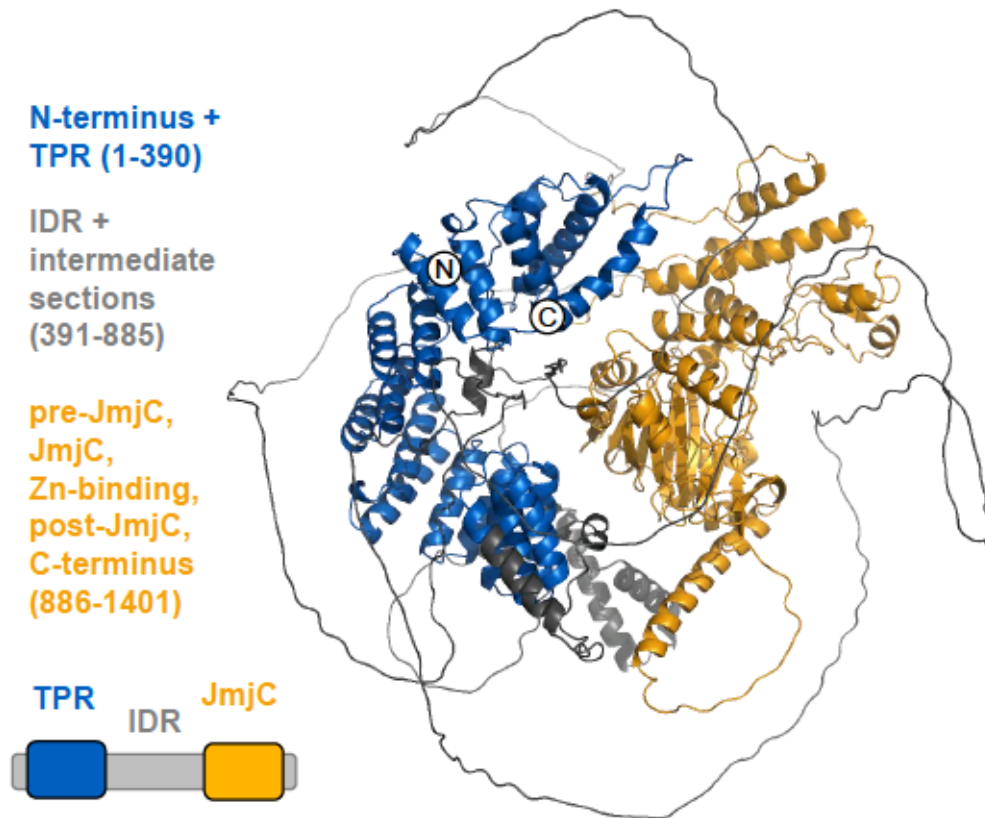
N-terminus + TPR (1-390)

IDR + intermediate sections (391-885)

pre-JmjC, JmjC, Zn-binding, post-JmjC, C-terminus (886-1401)

TPR   IDR   JmjC

Figure 1: Structure of KDM6A. A. Structure of KDM6A (AF-O15550-F1-model-v1) predicted by AlphaFold [16], introducing the domains with their respective color code used in this work.

We have recently shown that nuclear import of KDM6A is connected to its interaction and co-transport with proteins from the KMT2D family [48] and mutations in the TPR have been implied to hinder this co-transport [23]. Considering the impact of interaction on our study, decreasing or even enhancing protein-protein or protein-DNA/RNA interactions can easily promote cancer development [43, 49-52]. Concurrently, database search shows an increase in publications studying protein interaction domains in histone-modifying enzymes [53, 54] and eventually for KDM6A as well [22, 23]. Prominent interactors for KDM6A are the COMPASS complex [22, 36, 48] or CBP/p300 [19-21, 35, 55, 56]. In synergy with the activating H3K27me2/3-demethylation by KDM6A, the COMPASS-core components histone-lysine N-methyltransferase 2C or 2D (KMT2C/D) will methylate H3K4 and the CREB-binding protein/Histone acetyltransferase p300 (CBP/p300)-complex will acetylate the fully demethylated H3K27 to activate transcription [21]. These histone-modifying enzymes can harbor multiple and multivalent protein interaction domains that are crucial for the complex interaction and recruitment schemes [54, 57]. For instance, the two functional domains of KDM6A are both able to interact multiple times (TPR with e.g. COMPASS and JmjC with Histone H3) and the TPR can multivalently interact with at least two COMPASS proteins, Retinoblastoma-binding protein 5 (RBBP5) [48] and Set1/Ash2 histone methyltransferase

complex subunit (ASH2L) [23]. In respect to these vast interactions, we analyzed the changing ability of KDM6A variants to pull RBBP5 in an immunoprecipitation (IP). RBBP5 is a core component of the COMPASS-complex, binding specific chromatin sequences, stabilizing and activating the methyltransferase activity of KMT2 proteins [58]. Additional components of the COMPASS-complex are the PAX-interacting protein 1 (PAXIP1), WD-repeat protein 5 (WDR5) and the protein dpy-30 homolog (DPY30). Lastly, we covered the demethylase activity KDM6A variants. Malignancies caused by disruption of processes at developmental genes (e.g. HOXB1) were connected with the demethylase activity of KDM6A in multiple studies [10, 11, 18, 45, 59]. Therefore, we developed an ELISA-based activity assay to evaluate and quantify the demethylase activity for each variant used in our experiments. In addition, as a readout for the changes in cellular phenotype, we established a qualitative and quantitative imaging approach as well as an apoptosis assay. We then connected our observations at the cellular level with the mutational changes.
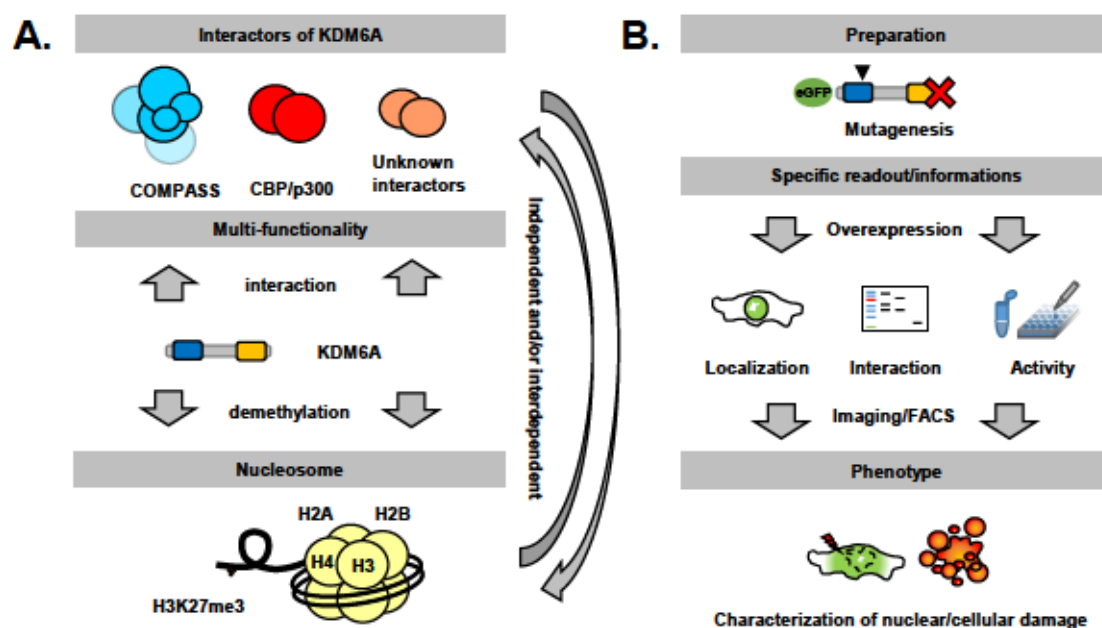

Figure 2: Core functions of KDM6A and workflow used for comprehensive analysis. A. The core functions of KDM6A being either TPR-mediated interaction or JmjC-dependent demethylase activity. Demethylation specifically occurs at the H3K27me3. Reported interactions of KDM6A with COMPASS complex and CBP/p300. The interaction functions of KDM6A can be interdependent and or independent from the demethylase functionality. **B.** Workflow used in our approach. KDM6A variants were generated by site-directed mutagegenesis and transiently transfected (overexpressed). Specific readouts were interaction by Co-IP and Western Blot, activity by ELISA and localization by fluorescence imaging and scatterplot analysis. We then linked our specific information on variants to the cellular phenotype observed via imaging and FACS.

By elucidating these structure-function relationships, we aim to increase the understanding for the different functionalities of KDM6A and how they influence and interact with the cellular environment. Gaining this kind of knowledge is key in finding approaches in targeted cancer therapy, which is fundamental for the treatment of urothelial cancer but also for other types of this malignant disease.

2. Results

Selection and generation of substituted and truncated KDM6A variants based on the mutational landscape of KDM6A in urothelial cancer and bio-informatics evaluation.
(*Text and figures taken or adapted from manuscript 1. (see **Publications** section, page 187)*

Based on 64668 tested samples from 44 tissue types, 2496 unique mutations are listed for *KDM6A* in COSMIC v92 (GRCh 38, November 2020). Among the cancer tissues, meninges and the urinary tract exhibit the highest mutation frequency for KDM6A with more than 30% of the tested cancer samples, respectively (411/1336 cases for UC). We selected T726K as a hotspot substitution variant as well as substitution mutations typically found in urothelial cancer cell lines and tissues: E315Q (located in TPR helix 6) and D336G (TPR helix 7), which shows reduced affinity to ASH2L[23], P966R (turning loop in the ribcage-like beta sheet cavity containing the active center), V1338F (zinc-binding domain) and C1361Y (zinc-chelator). Positions P966, Q1133, V1338 and C1361 are conserved in the KDM6A zebrafish orthologue and the closest functional KDM6A-paralogs KDM6B and KDM6C, giving them high probability for an important role. To reasonably compare the impact of these mutations, we additionally selected the variants Q1133A (interaction with Histone H3 tail) and H1329A (hydrophobic patch near active center), known to be catalytically inactive [15]. We created all substitution variants (**Fig. 3A**) by site-directed mutagenesis using eGFP-KDM6A as a template (1401 aa full-length KDM6A, isoform 1, 2 aa eGFP linker length). Nonsense mutations make up nearly one quarter of all point mutations and generate C-terminally truncated variants with full and/or partial losses of the IDR and JmjC. To understand the domains even better, one could also generate artificial N-terminally deleted constructs, which will not exist in a physiological context. Consequentially, we established a set of truncated variants, each with an eGFP-tag fused to their N-terminus, analogous to eGFP-KDM6A WT: ΔTPR, ΔIDR, ΔJmjC, TPR (= ΔIDR/ΔJmjC) and JmjC (= ΔTPR/ΔIDR) (**Fig. 3B**). We used a segmental cloning approach, treating the domains as building blocks, which can be combined individually. All KDM6A substitution and deletion variant proteins were detectable at the expected sizes by using a-KDM6A and a-GFP antibodies (**Fig. 3C/D**).
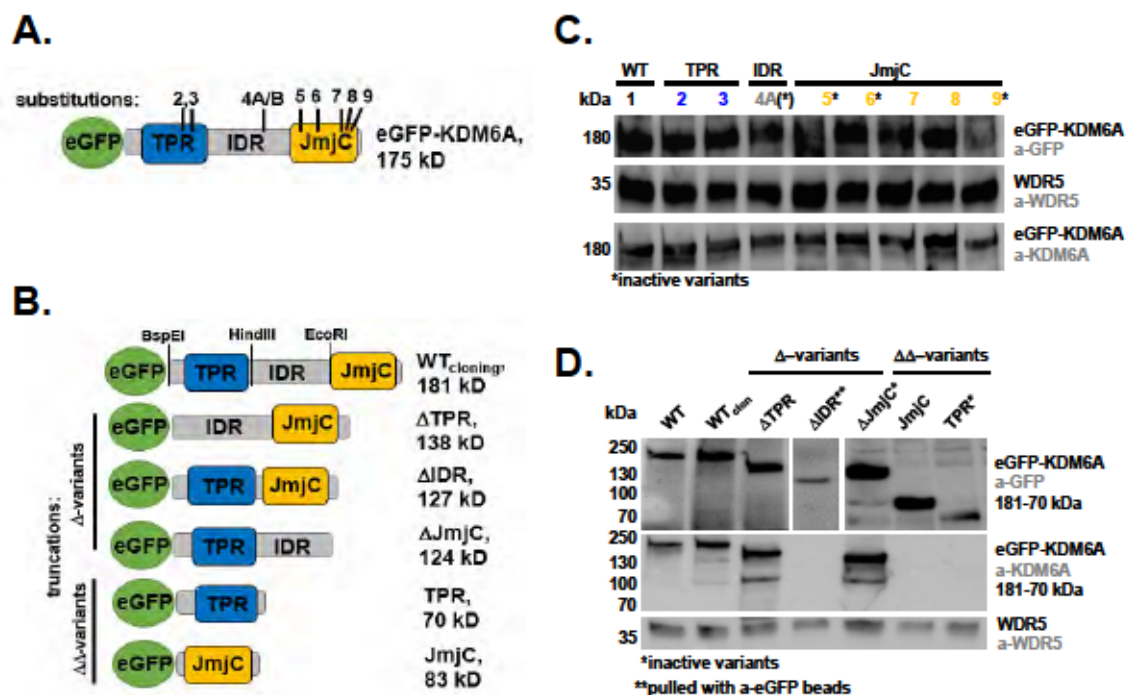
Figure 3: KDM6A variants, composition and expression. Code: 1 = WT, 2 = E315Q, 3 = D336G, 4A = T726K, 4B = T726V, 5 = P966R, 6 = Q1133A, 7 = H1329A, 8 = V1338F, 9 = C1361Y. **A.** Graphic display of the mutated position in the substitution variants. **B.** Graphic display of KDM6A truncation variants. **C./D.** Western blot showing all substitution and deletion variants 48 h after transient transfection, detected with a-eGFP and a-KDM6A antibodies, as well as WDR5 as lysate control, detected with a-WDR5 antibodies. Variants with IDR deletions are not detectable with the a-KDM6A antibody, as it recognizes epitopes within the IDR. The ΔIDR variant was enriched via IP. Color code: Blue-grey-yellow according to the site of mutation (as depicted in **Fig. 1**).

To get comprehensive biochemical data for the generated variants, we predicted protein solubility and pI [60], secondary structure [61] and thermodynamic stability [62]. As expected, single substitution variants were predicted (**Table 1**) as mostly unchanged compared to WT with small differences in thermodynamic stability, small changes in secondary structure (especially P966R) and change in pI, but not in solubility. The truncated variants were not eligible for thermodynamic stability and secondary structure prediction, but generically we expect a decrease in structural stability when removing large parts of the protein.

Fittingly, the truncated variants showed considerable differences in pI and solubility, which could alternate protein characteristics in vivo and in vitro. Deleting either TPR or JmjC increased predicted solubility. The single IDR is the most soluble part but this construct was not considered meaningful for our experiments, since the single domain has no apparent functionality yet. Accordingly, prediction for IDR-containing variants have a boost in their solubility prediction and, vice versa, lack of the IDR strongly lowers the solubility. Other nuclear proteins from the broader interactome of KDM6A all have a higher solubility prediction (**Table 1**). Notably, RBBP5 has a higher solubility than WDR5, although they both majorly consist of the solenoid WD40-domain [63, 64]. Here, the N-terminal disordered tail of RBBP5 apparently makes a large difference in solubility between both proteins, in analogy to KDM6A's highly soluble IDR-domain and to the disordered Nucleophosmin (NPM1).

Table 1: Prediction of solubility and pI of KDM6A variants. (From Koch et al, 2021) Solubility predictions[60] for KDM6A variants and nuclear proteins as measure for comparison. High pI (meaning high positive net charge at physiological pH) could point towards an abundance of DNA-binding sites or strong tendency towards negatively charged DNA[65].

KDM6A variant	pI	Predicted solubility
KDM6A WT	7.68	0.278
ΔTPR	7.12	0.336
ΔIDR	6.98	0.167
ΔJmjC	9.23	0.343
TPR	6.10	0.209
JmjC	9.06	0.266
IDR	9.62	0.483
T726K	7.88	0.278
T726V	7.68	0.277
Nuclear proteins	**pI**	**Predicted solubility**
NPM1	4.61	0.866
KMT2C	6.33	0.412
KMT2D	5.47	0.498
WDR5	9.40	0.396
RBBP5	4.94	0.654

KDM6A demethylase activity is strongly affected by substitutions and deletions within the JmjC domain.

*(Text and figures taken or adapted from manuscript 1. (see **Publications** section, page 187)*

To assess the demethylase activity of KDM6A variants, we established an ELISA-based H3K27me3 demethylation assay. **Fig. 4A** summarizes the results and shows all measured datapoints. Upon mixing with an H3K27me3 peptide (minimum residues required: H3 aa 15-34), KDM6A binds to the peptide. The demethylation reaction (**Fig. 4B**) occurs by oxidation of one of the ε-methyl groups of H3K27me3 in the presence of Fe(II) and under consumption of O_2 and α-ketoglutarate. Apart from the demethylated H3K27, succinate, CO_2 and formaldehyde are produced by the reaction. The assay conditions were first evaluated with commercially available recombinant full-length KDM6A. After initial evaluation, the expressed GFP-tagged KDM6A variants were pulled from lysates, bound to GFP-dynabeads (anti-GFP nanobodies coupled to magnetic beads). The elution of the protein proved to be impractical since acidic elution and heat/SDS would inactivate the protein. Therefore, the protein was left on the dynabeads, the washed dynabeads were subsequently incubated. We used fluorescence emission spectra of the eGFP-tag to determine the respective amount of protein bound onto dynabeads and normalized it throughout the experimental replicates (**Fig. 5A/B**). The activity of the bound protein was determined with primary antibodies against the demethylated products (a-H3K27me2 antibody with affinity towards H3K27me2 and me1 but not against me3). The secondary antibodies used were conjugated with alkaline phosphatase (ALP). The colorimetric readout measured the absorption of para-nitrophenol (pNP) generated through de-phosphorylation of para-nitrophenolphsophate (pNPP) catalyzed by ALP. The data points achieved were subsequently fitted via 4PL-regression (**Fig. 5C/D, Eq. 1**). The eGFP-KDM6A WT species served as a reference point for the demethylase activity of the variants. Data was obtained from at least nine data points from three independent experiments, each in triplicates.
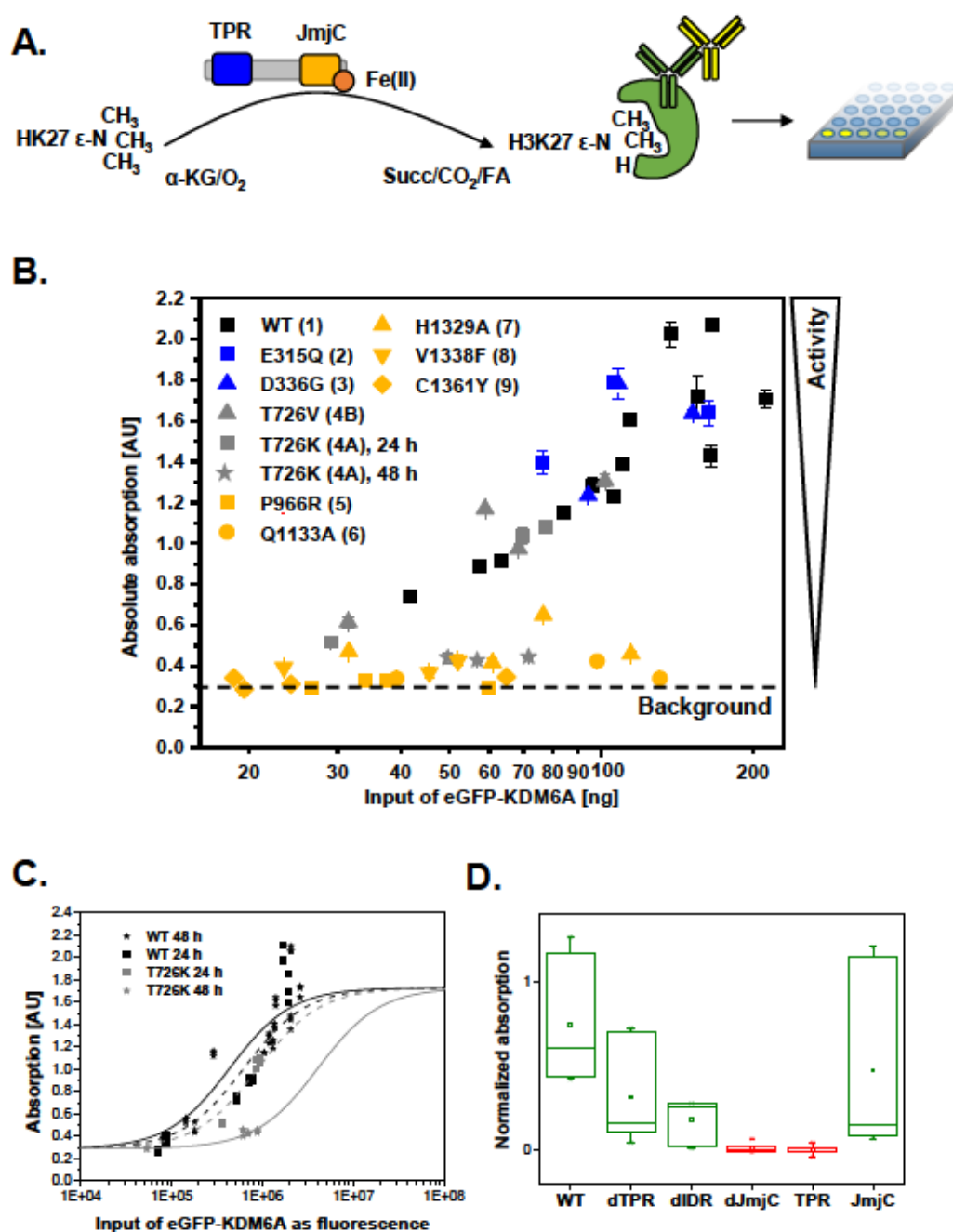
Figure 4: Demethylase activity assay results. A. Data points of all substitution variants measured, error bars display the deviation within triplicates. Active variants show an increased assay absorption with increasing amounts of input. Inactive variants remain at background assay absorption independently of the amount of protein used in the measurement. **B.** Enzymatic reaction catalyzed by KDM6A: H3K27me2/3 is demethylated to H3K27me1/2 under consumption of O_2, α-ketoglutarate and the production of succinate, CO_2, and formaldehyde. **C.** Activity of the T726K variant is dependent on post expression time. WT activity does not change significantly between 24 h (black) to 48 h (grey), whereas in T726K activity is slightly lower than WT after 24 h (blue) and strongly reduced after 48 h (cyan), as can be seen by the shift of the 4PL-fit towards a higher input. **D.** Absorption of truncated variants, three repeats with triplicates each. We qualitatively confirmed that all truncated variants with

a JmjC domain (WT, ΔIDR, ΔTPR and JmjC) are catalytically active (green boxes) since the distribution of the signal over the background is an evidence for activity. The two variants without JmjC domain (ΔJmjC and TPR) are catalytically dead (red boxes) and consequentially not exceed assay background levels.
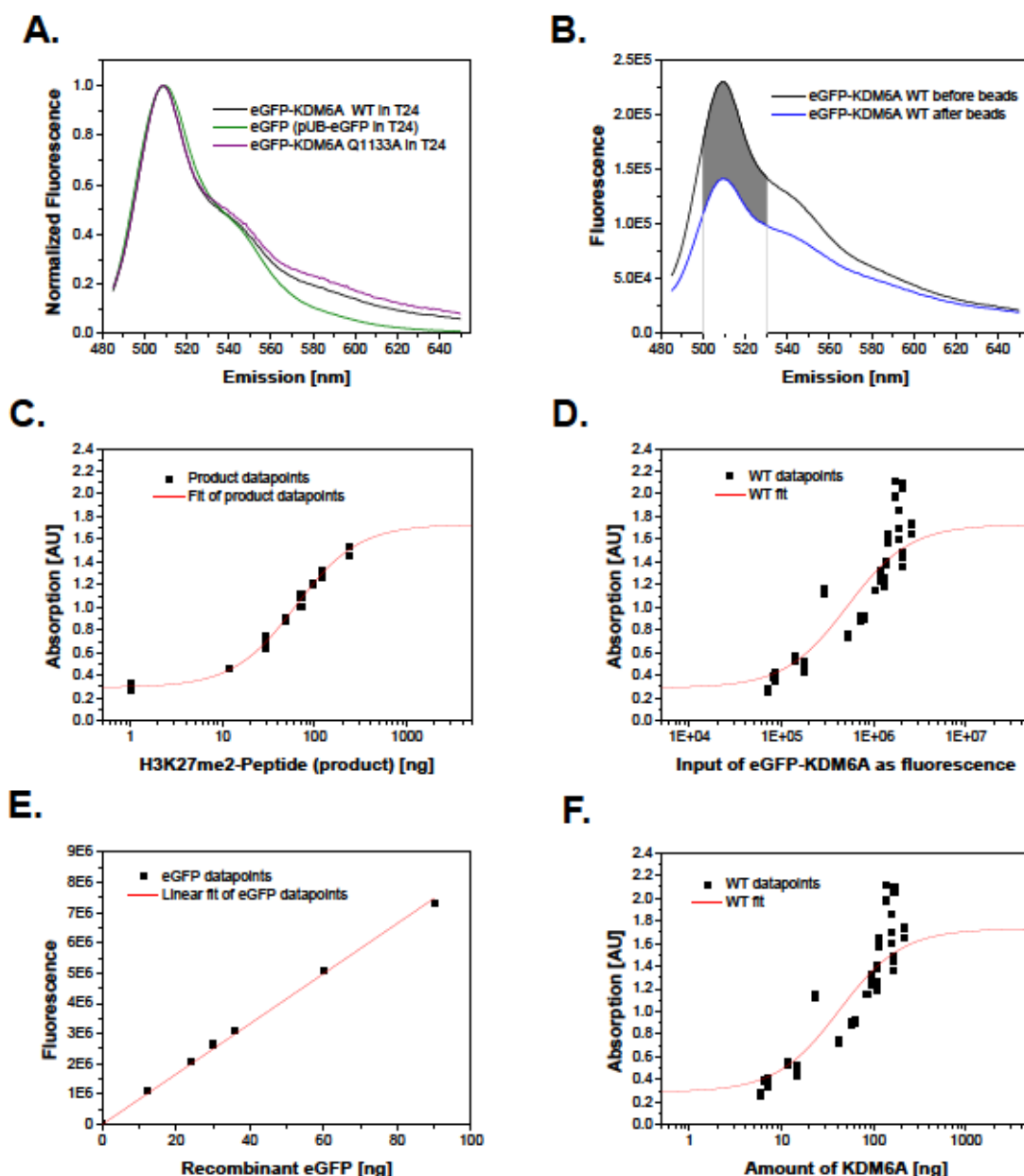

Figure 5: Demethylase activity assay principle and references. A. Normalized fluorescence emission spectra obtained from crude cell lysate of eGFP, eGFP-KDM6A WT or eGFP-KDM6A Q1133A samples, with excitation at 470 nm. All fluorescence emission spectra of the substitution variants had a similar shape (Q1133A variant shown as example). **B.** Fluorescence emission spectra of eGFP-KDM6A WT crude cell lysate before incubation (black line) and after incubation (blue line) with GFP-trap dynabeads. The amount of fluorescent protein bound to the GFP dynabeads was determined by subtracting the eGFP fluorescence emission after incubation from eGFP fluorescence emission before incubation within in the range of 500-530 nm. The integral obtained was used to

calculate the amount of fluorescent protein bound to the beads. **C.** 1-238 ng demethylated product (H3K27me2 peptide, MW = 2945.5 g/mol) served as a reference curve to determined how much demethylated product (x-axis) corresponds to the respective amount of assay absorption unit (y-axis). Fitted accordingly to *Eq. 1*. **D.** The integral value describing the amount of fluorescent protein bound to beads (as described in **B.**) was plotted against absorption (shown exemplary for KDM6A WT) and fitted accordingly to *Eq. 1*, with *c* being an open fit parameter. **E.** A defined amount of recombinant eGFP was used to correlate the fluorescent emission (calculated as described in **B.**) to the amount of protein. The displayed linear dependency was used to analogously calculate the amount of eGFP-KDM6A per fluorescent emission unit. **F.** Same as **D.** but with transformed x-axis using the dependency from **E.**, converting fluorescent emission units into the corresponding amount of eGFP-KDM6A protein. The data large data scattering in **D** and **F** can be explained by considerable variation of ELISA signal based on conditions (expression based activity-differences, incubation temperature, incubation time, signal/concentration-relationship) for biological replicates done over a month-long time span. Signal normalization (Max/Min) was not sufficient to smoothen dataset. Additionally, there was a minor variation of ELISA signal based on inaccuracy (e.g. pipetting errors) between technical replicates.

Fitting was done with a four parameters logistic regression (4-PL) according to *Eq. 1*:

$$y = d + \frac{a - d}{1 + \left(\frac{x}{c}\right)^b} \qquad Eq. 1$$

with *y* = assay absorption, *x* = amount of demethylated product (in **Fig. 5 C**), amount of KDM6A species (in **Fig. 5 D/F**), *a* = background signal, *d* = maximum signal, *b* = slope, *c* = c50, x-value at half maximum y.

The following fit results were obtained for the substrate reference curve (**Fig. 5 C**): *a* = 0.29, *b* = 1.27, *d* = 1.73 *and c* = 62.4 ng. All measurement with KDM6A protein were later normalized to these values of *a* and *b* to correct for background and maximum signal fluctuation between experiments. The exponent *b*, the slope-factor, was assumed as constant. For the KDM6A measurements the amount of protein displayed as fluorescence was plotted against assay readout (**Fig. 5 D**) and fitted accordingly with 4-PL, with *c* being the only open parameter. As an example, c50 for WT was determined with 518,000 +/- 56,000 AU. If variants were less active, one would expect an increase in c50, as the amount of protein needed to produce the same amount of demethylated product (displayed as assay absorption) would rise. Transforming the x-axis from fluorescent signal into amount of protein was done by using recombinant eGFP of known concentration as reference (**Fig. 5 E**) under the same conditions and assuming the same spectroscopic parameters (**Fig. 5 A**).

Accordingly, determined the molar integrated fluorescence (500-530 nm) from emission spectra for recombinant eGFP (MW = 27 kDa), yielding $2.34*10^{18}$ AU/mol with 1 ng eGFP =

86,900 [AU] under given conditions. Therefore, 1 ng eGFP-KDM6A WT (MW = 181 kDa) would yield an integrated fluorescence signal of 12,900 [AU]. Using this relation, we could calculate the amount of fluorescent KDM6A in the assay (**Fig. 5 F**). The *c50* for all available substitution variants was obtained in this manner, fixing all parameters except *c*, using it as a relative measure of activity.

In a second set of measurements, we analyzed all truncated variants for their demethylation activity. Since the truncated variants showed an anomalous and changing emission spectrum quantification of the protein amount based on the emission spectrum (as in **Fig. 5 B**) was impossible. Therefore, we assessed the demethylase activity of the truncated KDM6A variants was a qualitative manner (**Fig. 4D**). Inactive variants would never exceed background levels while active variants will be distributed above background. As expected, all variants with JmjC domain could be confirmed as active and all variants without JmjC domain did not show any activity.

To quantify and compare the substituted variants comprehensively, we calculated the specific activity for each variant. As calculated from **Fig. 5 C**, the *c50* for half absorption value always corresponds to 62.4 ng demethylated product. Given the fixed time of 240 min per assay, the specific activity under these conditions could be calculated for each variant, as summarized in **Table 2** below.

Table 2: Activity of KDM6A substitution variants. (From Koch et al, 2021) Specific activity of KDM6A variants depends strongly on the mutation site. Mutations affecting the JmjC domain impair catalytic activity, whereas TPR substitution mutation slightly enhanced activity. The IDR mutation T726K showed time-dependent diminution of activity selectively caused by the K726 residue.

No.	KDM6A variant, affected domain	Specific activity [10^{-3} µmol min^{-1} mg^{-1}]
1	WT	2.15 ± 0.24
2	E315Q, TPR	3.96 ± 0.58
3	D336G, TPR	3.76 ± 0.89
4a	T726K, IDR	0.69 ± 0.10
4b	T726V, IDR	1.54 ± 0.11
5	P966R, JmjC	<0.07 ± 0.03
6	Q1133A, JmjC	<0.09 ± 0.02
7	H1329A, JmjC	0.28 ± 0.05
8	V1338F, JmjC	0.29 ± 0.03
9	C1361Y, JmjC	<0.12 ± 0.03

Note that TPR substitutions E315Q and D336G both had a higher activity than WT. Interestingly, activity in T726K decreased time-dependently within two days after transfection (**Fig. 4D**). This prompted us to generate the eGFP-KDM6A T726V variant, which displayed a demethylase activity in the range of WT, narrowing the cause of the time-dependency down to the presence of the K726 residue. As expected, JmjC-mutations all had impaired demethylase activity. The eGFP-KDM6A variants H1329A and V1338F had strongly reduced activity, while P966R, Q1133A and C1361Y were non-active.

Literature meta-assessment and database mining reveal nuclear and mostly chromatin associated interactions of KDM6A.

To get an overview on putative KDM6A interaction, we used the STRING-database [66] to evaluate the ten most likely protein interaction partners, judged by text-mining, experiments, databases, co-expression, neighborhood, gene fusion, co-occurrence. The research yielded mostly COMPASS proteins (KMT2C/D, WDR5, ASH2L, PAXIP1, RBBP5) as proposed binding partners (**Fig. 6**). Additionally, Histone H3, EZH2, HDAC3 and HOXB1 were found. EZH2 is one of the main antagonists of KDM6A in its catalytic function, acting as a K27 methyltransferase [32]. HDAC3 is also affiliated with the antagonistic side of KDM6A, deactivating chromatin by deacetylation [67]. Lastly, HOXB1 is one of the most prominently identified target gene clusters of KDM6A [11].
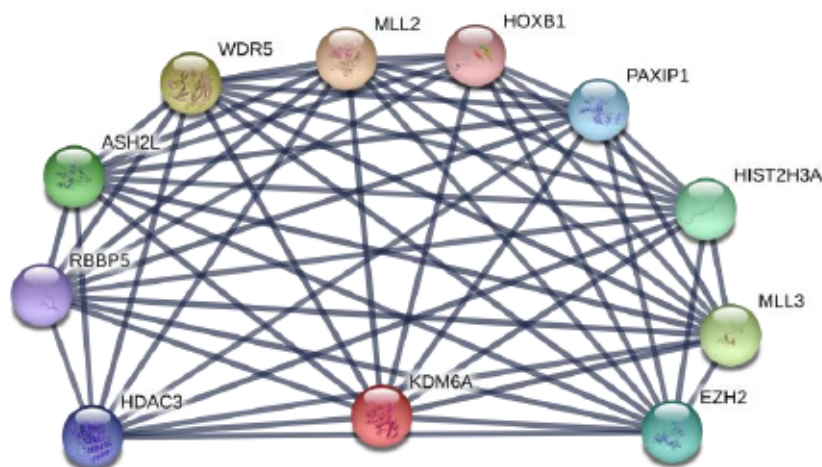

Figure 6: Predicted interactions of KDM6A. STRING network (as of 2019) with ten nodes when using all interaction sources, included text-mining, experiments, databases, co-expression, neighborhood, gene fusion, co-occurrence.

In addition to the STRING-results, we started a literature meta-assessment of articles listed with 'KDM6A' and/or 'UTX' in the title (approx. 500 articles in 2021). We allowed all species homologs but only attributed interaction assessed by Co-IP/western blot and mass

spectrometry. Here, COMPASS members had again the highest occurrence. In addition, CBP/p300 was found multiple times as well as the RNA-Pol II interactor SUPT6H, the retinoic acid receptors RXR/RAR and the tumor-suppressor p53. See **Chapter 4-Table supplement 1** for a full list.

Mass spectrometry (MS) data reveals novel interactors of KDM6A

(Text and figures taken or adapted from manuscript 1. and contains data from 2. (see **Publications** *section, page 187)*

To determine urothelial and cancer specific interaction partners of KDM6A, we performed a proteomics analysis from mass spectrometry of stably transduced KDM6A-tagGFP2 in the urothelial cancer cell lines (UCCs) RT-112 and VM-CUB1, which both exhibit KDM6A and/or KMT2C/D mutations [48]. Among the top ten proteins in KDM6A pull down and subsequent MS (see **Table 3**), we found Nucleophosmin (NPM1), Nucleolin, ribosomal subunits 40S and 60S, as well as histone variants. The identified proteins were associated with the following pathways [68] (see **Fig. 7**): cell cycle, chromatin organization, DNA repair, RNA metabolism (including rRNA processing) and protein metabolism (including ribosome biogenesis and post-translational modification of histones). This will open up to future research by confirming the interactions and performing initial characterization. Notably, COMPASS components were undetectable, although previous Co-IP experiments suggested assembly with RBBP5 and Ash2L [23, 48].

Since the origin of the activity's time-dependence was unclear, we tested T726K in a proteomics experiment for post-translational modifications. Concurrent, we predicted [69] PTMs of all mutated sites in our substitution variants before and after mutation. PTM-prediction differs multiple times between T726 and K726. Unfortunately, upon MS-analysis, no modification of the K726 residues was detectable, leaving the question of the K726's role unanswered for the moment. See **Chapter 4-Table supplement 2** for detailed values.
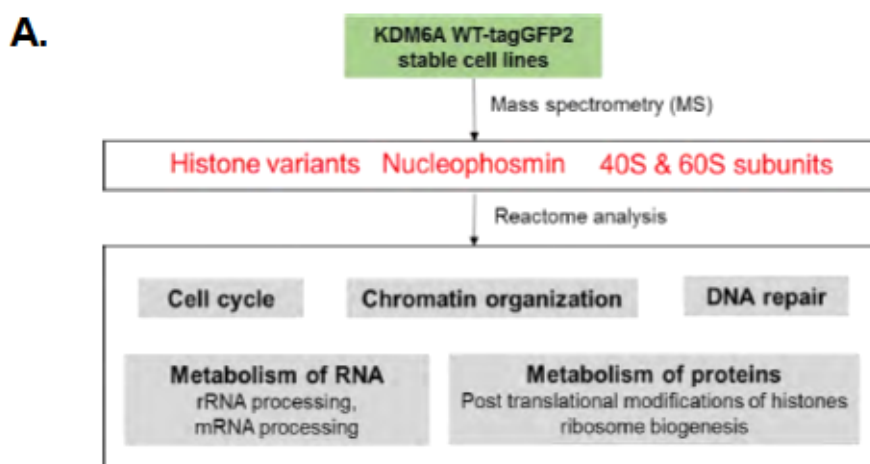
A.

Figure 7: Pathway enrichment analysis of MS data from KDM6A-tagGFP2 cell lines. A. Data from quadruplicates of three independent experiments of two different KDM6A-tagGFP2 cell lines using the reactome database. MS data sets were applied to the reactome analysis database. The resulting output shows significantly enriched pathways. The five most interesting pathways are depicted: Cell cycle, chromatin organization, DNA repair, RNA metabolism including rRNA processing and protein metabolism including ribosome biogenesis and post-translational modification of histones.

Table 3: Top ten proteins identified by MS in KDM6A-tagGFP2 stable cell lines VM-CUB1 and RT-112. Among the top ten proteins in KDM6A pull down and subsequent MS, Nucleophosmin (NPM1), Nucleolin, ribosomal subunits 40S and 60S, as well as histone variants, are detectable. Mutational KDM6A and KMT2C/D status and experiments with RT-112 and VM-CUB1 have been recently published [48].

#	p value	Protein names
1	7.71E-08	KDM6A
2	6.19E-07	NPM1
3	4.31E-06	60S ribosomal protein L10a
4	5.53E-05	Histone H2A types A/B/C
5	2.66E-05	60S ribosomal protein L18a
6	7.40E-05	Histones H4
7	4.88E-04	60S ribosomal protein L14
8	4.23E-04	Histones H2A.V H2A.Z
9	3.80E-04	Histone H3.3
10	2.63E-04	Nucleolin

Full-length KDM6A is mandatory for maximal binding of the COMPASS-complex core component RBBP5.

*(Text and figures taken or adapted from manuscript 1. (see **Publications** section, page 187)*

After our initial interaction research, we settled on a selection of interaction partners to test and confirm with eGFP-KDM6A WT, our truncated variants (excluding ΔIDR due to stability issues) and the substitution variants D336G and T726K. Since all of the a-KDM6A antibodies bind to an epitope within the IDR, they would not be suitable for truncated variants lacking this domain. Consequentially, we decided to visualize eGFP-tagged species with an a-GFP antibody (see **Fig. 4 C**). Since we already confirmed RBBP5's interaction with KDM6A in earlier experiments[48] and the antibody gave better signal than the a-WDR5 antibody, we settled on RBBP5 as target interacting protein. After initially evaluating conditions for co-immunoprecipitation (washing control with eGFP see **Fig. 8E**), we tested for RBBP5 interaction with KDM6A experiments 48 h post transfection of selected KDM6A truncation variants (**Fig. 8B**) in T-24 and SW-1710. We had previously shown that RBBP5 was enriched in KDM6A-tagGFP2 Co-IPs in different urothelial cancer cell lines [48]. Interestingly, we observed that for maximum interaction with RBBP5 all KDM6A domains are needed (**Fig. 8C**). Deletion of any domain impaired RBBP5 binding. Specifically, TPR-containing variants (ΔJmjC and TPR, **Fig. 8D**) and to some extent IDR-containing variants (ΔTPR) bound RBBP5 to some degree, but the JmjC domain alone did not at all. Notably, a recent study [23] found that KDM6A TPR mutations, among them the D336G variant, predominantly localized in the cytoplasm in stably transfected HeLa cells. We performed a pull-down experiment of KDM6A WT, D336G and T726K variants with RBBP5 to test for impaired association with the KMT2C/D-complex. RBBP5 was pulled down with all three KDM6A variants to similar extents, but not with the eGFP control in both urothelial cancer cell lines (**Fig. 8C**), suggesting that they are present in the same complex.

Intriguingly, our proteomics analysis (**Table 3**) failed to prove the expected COMPASS-interaction that we could show MS- and Co-IP experiments concurrently. Instead, as stated above, we found a strong enrichment of nucleolar proteins, such as Nucleolin, NPM1 and ribosomal subunits. In our Co-IP experiments, NPM1 was only visibly enriched with the KDM6A T726K mutant in T-24 whereas other selected KDM6A variants failed to interact or gave very weak bands (ΔTPR and ΔJmjC, **Fig. 8B** white stars). Therefore, we searched for a complementary approach to detect associations of KDM6A with NPM1.
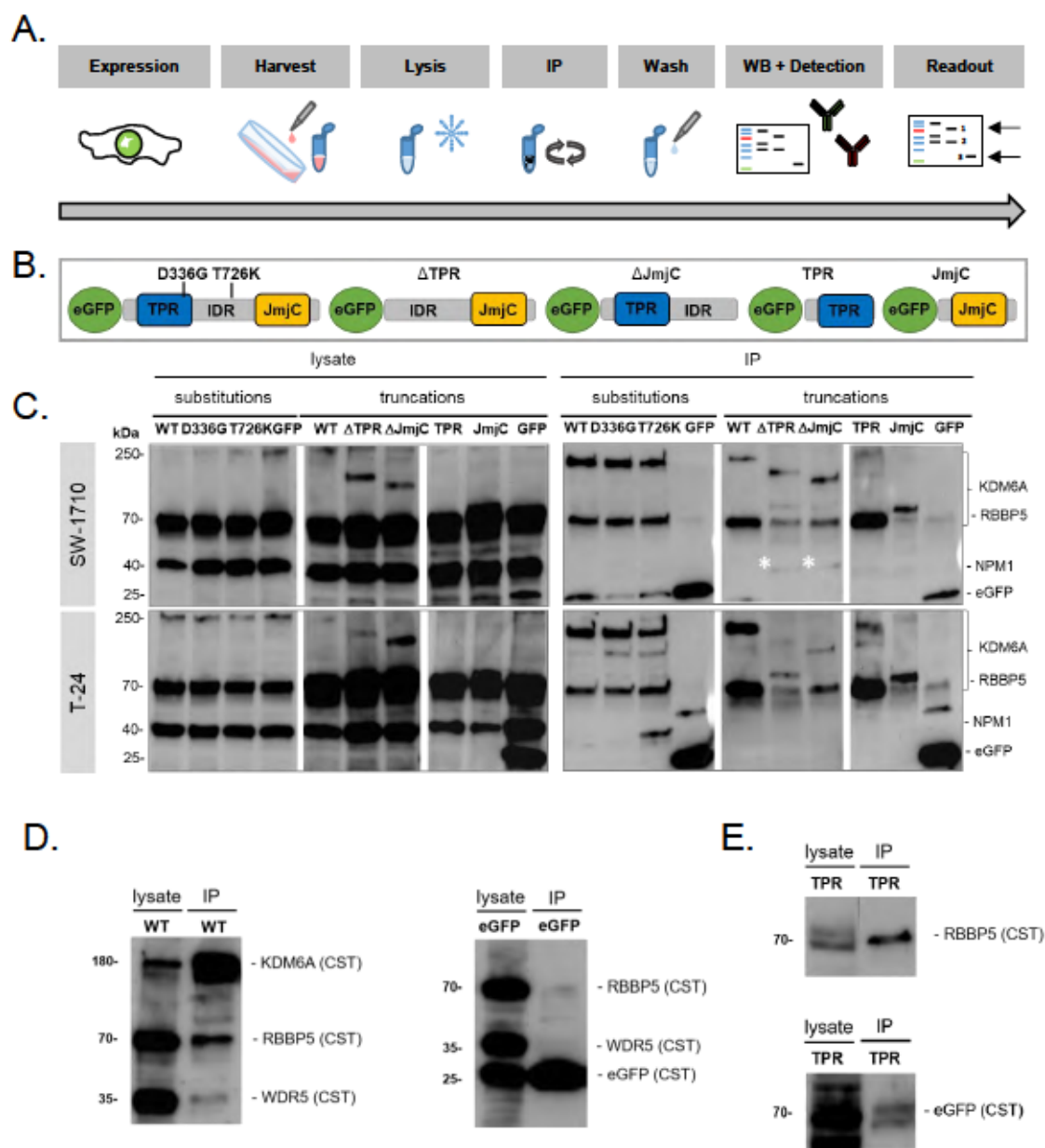
A. Expression | Harvest | Lysis | IP | Wash | WB + Detection | Readout

B. D336G T726K | ΔTPR | ΔJmjC | TPR | JmjC

eGFP TPR IDR JmjC | eGFP IDR JmjC | eGFP TPR IDR | eGFP TPR | eGFP JmjC

C.

lysate | IP

substitutions | truncations | substitutions | truncations

kDa | WT D336G T726K GFP | WT ΔTPR ΔJmjC TPR JmjC GFP | WT D336G T726K GFP | WT ΔTPR ΔJmjC TPR JmjC GFP

SW-1710

250-
70-
40-
25-

KDM6A
RBBP5
NPM1
eGFP

T-24

250-
70-
40-
25-

KDM6A
RBBP5
NPM1
eGFP

D.

lysate IP | lysate IP
WT WT | eGFP eGFP

180- KDM6A (CST)
70- RBBP5 (CST)
35- WDR5 (CST)

70- RBBP5 (CST)
35- WDR5 (CST)
25- eGFP (CST)

E.

lysate IP
TPR TPR

70- RBBP5 (CST)

lysate IP
TPR TPR

70- eGFP (CST)

Figure 8: Immunoprecipitation with eGFP-KDM6A WT and variants. A. Workflow of the IP-experiment, from expression in cells to western blot. **B.** KDM6A variants used in the interaction study. **C.** IP of KDM6A variants with RBBP5 and NMP1 in SW-1710 and T-24 (n=2-7). The cleared cell lysates (left section) indicates GFP, RBBP5 and NMP1 protein levels after transfection of KDM6A WT, variants and eGFP. For SW-1710 and T-24, IP (right section) was used to pull of RBBP5 with eGFP-KDM6A WT, D336G, T726K, ΔTPR, ΔJmjC, TPR, JmjC and the control eGFP. In SW-1710, NPM1 is slightly enriched in the IP with the in the eGFP-KDM6A ΔTPR and ΔJmjC variants, whereas in T-24, NPM1 is enriched in eGFP-KDM6A T726K variant (n=1). **D.** The TPR and RBBP5 signals were separately stained from one IP on two membranes, since the bands overlap in **C**. **E.** Dynabeads washing control on same membrane, washed with 3x Ripa like buffer (RLB). While KDM6A WT pulls a clearly visible RBBP5 band and a low WDR5 band, both are undetectable in eGFP-transfected cells alone (negative control).

Fluorescent proteins eGFP/tagGFP2 and mCherry coupled to KDM6A and RBBP5 are not within sufficient proximity *in cellulo* to confirm interaction via MFIS-FRET experiments.

We designed two concurrent confocal live cell multiparameter fluorescence spectroscopy (MFIS)-FRET experiments to complement our biochemical approaches in order to clarify i) if KDM6A is able to form homo-dimers and ii) if we can confirm the physical interaction between KDM6A and RBBP5 detected in the IP-study via MFIS-FRET. In our early work [70], we could show that FRET experiments with combinations of N- and C-terminally labeled KDM6A (eGFP and mCherry, respectively) yielded no visible FRET-readout in the fluorescence lifetime. We concurrently performed an IP-experiment where we attempted to pull mCh-KDM6A with eGFP-KDM6A and detect the pull with an anti-mCherry antibody. This experiment yielded no sign for homo-interaction as well. There remain a few resources claiming that such homo-interactions exists, albeit still lacking conclusive evidence [14]. For another MFIS-FRET experiment (**Fig. 9A**), we cross-matched all terminal combinations of RBBP5 and KDM6A (**Fig. 9B**) possible with fluorescent proteins at both C-terminus and N-terminus. The KDM6A-tagGFP2 construct substituted our KDM6A-eGFP construct, which suffered from very low expression efficiency. The constructs (C-/N-terminally tagged KDM6A and RBBP5) were transiently transfected and overexpressed in T-24 cells, all consecutive MFIS-FRET experiments were done in live cells. In the preliminary evaluation run, results show no distinct and conclusive change in species-weighted average fluorescence donor lifetime $<\tau_D>_x$ (**Fig 9C/D**) for all four combinations used. For a clear FRET readout, we would at least expect a decrease of 0.2-0.4 ns. There is a small possibility for the C/C-terminal combination KDM6A-tagGFP2 with RBBP5-mCh to harbor a minor FRET signal but the data is not conclusive enough to justify the huge commitment necessary for further measurements. Concluding, the physical interaction seen in the IP experiments could not be reproduced with MFIS-FRET experiments, indicating that the fluorescent proteins topological

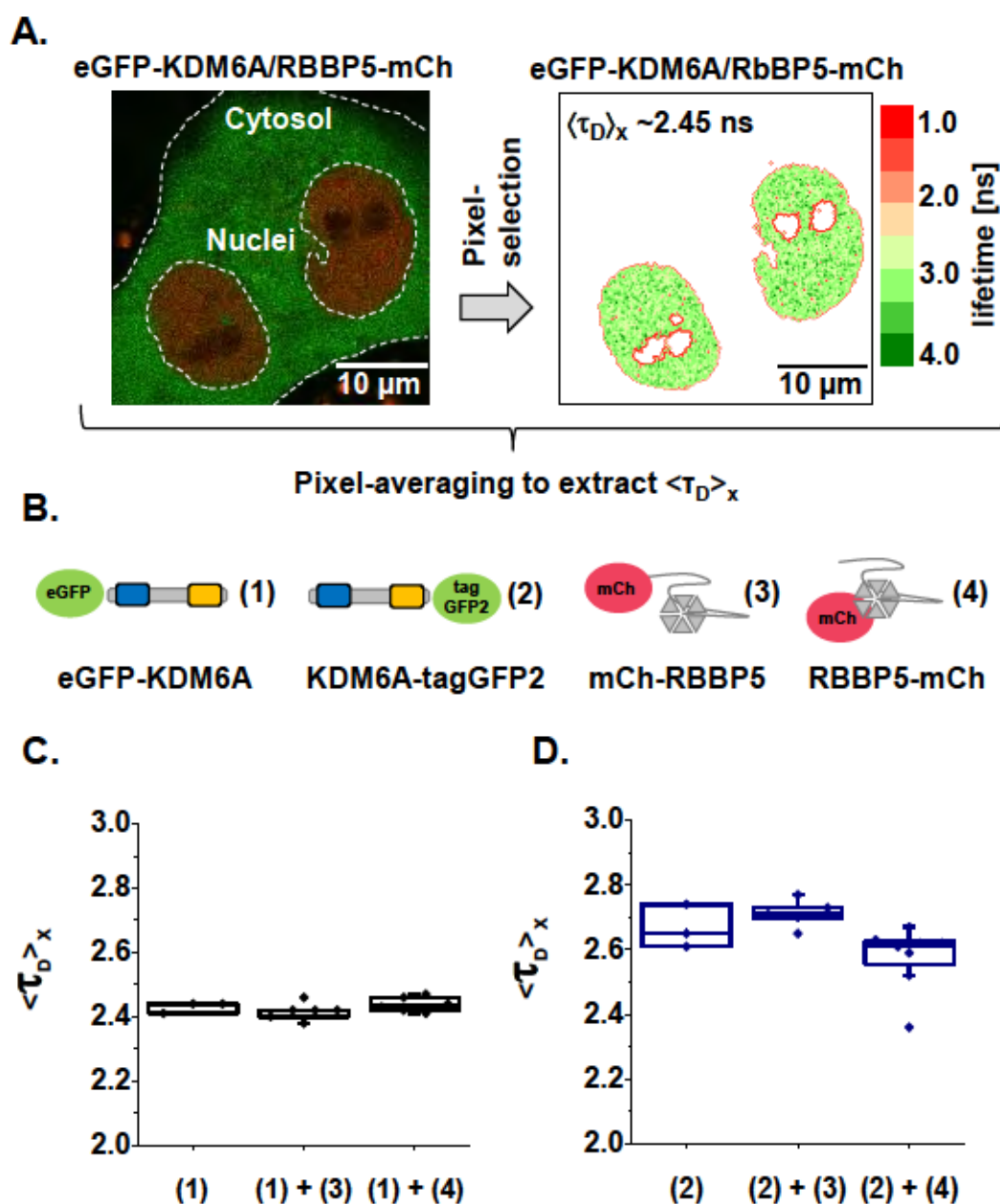distribution (protein cloud) are not at all or only to a minor fraction within a FRET distance (<100 Å).

Figure 9: MFIS-FRET with combinations of KDM6A and RBBP5. A. MFIS-FRET workflow: Pixel-wise species-weighted average fluorescence lifetime $<\tau_D>_x$ can be extracted and fitted from the FLIM-image. The distribution in the nucleus is homogenous. **B.** Constructs used in this work were either tagged N- or C-terminally with GFPs or mCherry. **C/D.** Distribution of fraction weighted average donor lifetime $<\tau_D>_x$ for Donly samples and four different combination of eGFP/tagGFP2 and mCherry with KDM6A and RBBP5 displayed as boxplots. $<\tau_D>_x$ is expected to drop below the Donly distribution if FRET occurs, scaling with the rate of FRET between donor and acceptor. Note that tagGFP2

unquenched $<\tau_D>x$ (approx. 2.6 ns) is slightly higher than eGFP (approx. 2.4 ns). For full fit parameter see **Chapter 4-Table supplement 3-7**.

KDM6A single substitutions do not alter nucleoplasmic localization.

*(Text and figures taken or adapted from manuscript 1. (see **Publications** section, page 187)*

In SW-1710 and T-24 cells, wildtype eGFP-KDM6A was located in the nucleoplasm and in the cytoplasm, sometimes with a tendency to cytoplasmic speckle formation (dependent on the dose of the transient overexpression). The localization of eGFP-KDM6A WT is similar to that of the endogenous KDM6A in these cell lines, as shown by immunostaining using two different KDM6A antibodies (sc-514859 and CST-3351, **Fig. 10**). All substitution variants showed cytoplasmic and predominantly nucleoplasmic localization with different degrees of cytoplasmic speckle formation, as well as weak accumulation around the nucleoli (**Fig. 11, Fig. 12**).
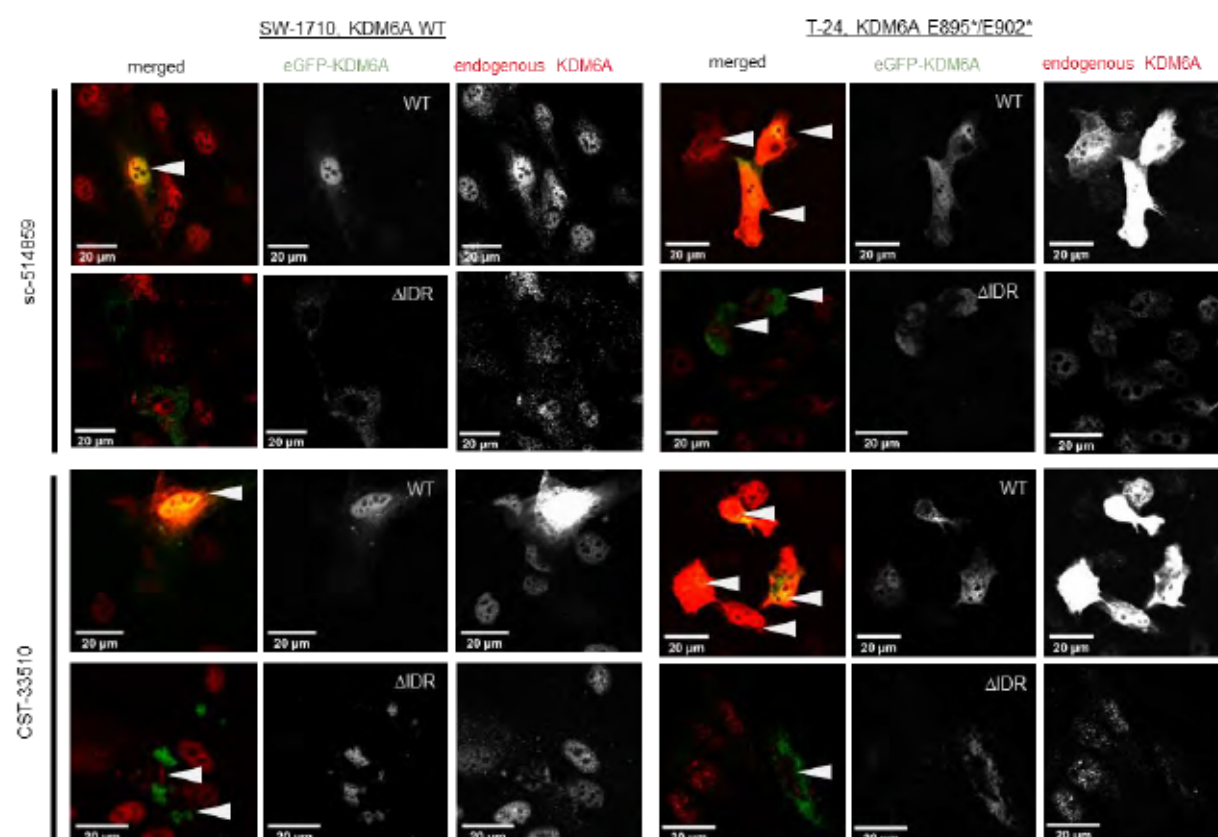

Figure 10: **Staining of endogenous KDM6A-levels in UCCs.** Control experiments for antibody staining specificity and endogenous KDM6A protein levels. SW-1710 (KDM6A WT) and T-24 (KDM6A E895*/E902* mutation) were transfected with KDM6A WT or ΔIDR variants and stained with two different antibodies (sc-514859, Santa Cruz biotechnology and CST-33510, Cell Signaling) raised against epitopes within the KDM6A central region (IDR) and 2nd antibody labeled with AbStar Red (Abberior). Clearly, both antibodies (in red channel) detect transfected KDM6A WT protein in addition

147

to the endogenous KDM6A (very strong signals in green channel), but not the ΔIDR variant since the epitope is missing in this variant.

To rule out unspecific or GFP driven localization of transfected KDM6A, we analyzed free eGFP controls in SW-1710, T-24 and HBLAK cells (**Fig. 11-13**). Free eGFP as well as eGFP-KDM6A shows both cytoplasmic and nuclear localization. The main difference between eGFP and eGFP-KDM6A is the additional presence of free eGFP in nucleoli while eGFP-KDM6A is exclusively found nucleoplasmic. As shown in one of our earlier publications[48], KDM6A transport into the nucleus depends largely on the KMT2C/D COMPASS complex. Therefore, the observation that KDM6A WT and all substitution variants localized in the nucleus, suggests functional nuclear import, as well as interaction with the COMPASS complex. An increased tendency for DNA release was identified for KDM6A truncation variants, which colocalized at cytoplasmic DNA sequences and prompted us to evaluate the nuclear damage caused by these variants.
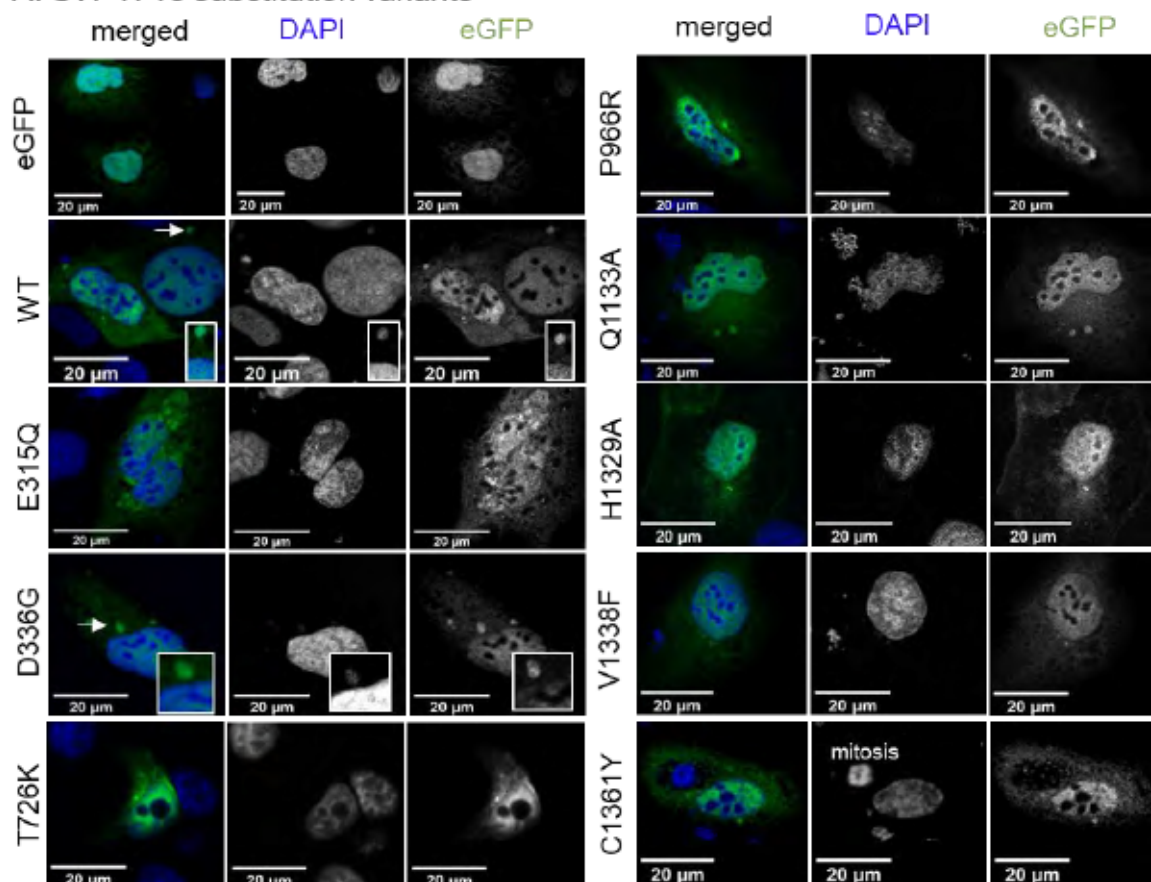
KDM6A truncated variants cause severe nuclear damage.

*(Text and figures taken or adapted from manuscript 1. (see **Publications** section, page 187)*

Compared to the eGFP and eGFP-KDM6A WT controls, deletion of any functional domain resulted in a heterogeneous cellular response with respect to subcellular distribution, localization and nuclear integrity. For further comparison, we transfected the KDM6A TPR and JmjC variants into HBLAK cells, non-transformed urothelial cell lines, and observed similar damaging effects. To confirm the non-transformed status of our key proteins in HBLAK, we performed an ICC (**Fig. 13A**). Via structured illumination microscopy (SIM), we confirmed correct nuclear or nucleolar localization for KDM6A, KMT2C, NPM1, a prominent shuttling and chaperone protein found especially in nucleoli, and SC-35, a serine/arginine-rich splicing factor. After transient transfection, KDM6A variants displayed either a weak cytoplasmic or nucleoplasmic localization in all tested cell lines. In some variants, they presented a speckled, perinuclear distribution of the transfected protein associated with nuclear defects, as shown in **Fig. 11-13**. Cellular responses in all cell lines were comparable. Nuclear DNA, which was excessively released into the cytoplasm, colocalized with KDM6A truncation variants. These results indicate the intrinsic ability of all variants to bind to chromatin, either directly through the JmjC or indirectly via other protein-protein interactions by TPR-containing variants. All variants in principle localized to the nucleus, but (to various extents) showed anomalies like partial nuclear redistribution to nucleoli or accumulation in perinuclear DNA-associated speckles. These observations raise two further questions, namely (1) how deletions of one or two KDM6A domains impair functional interactions with known interacting proteins such as RBBP5, the KMT2C/D (COMPASS) complex and (2) whether NPM1 may be involved in KDM6A interactions. Addressing these questions, we

performed protein immunoprecipitation (Co-IP) (**Fig. 8**) and co-staining of transiently transfected KDM6A variants with NPM1 and DAPI (**Fig. 14**).
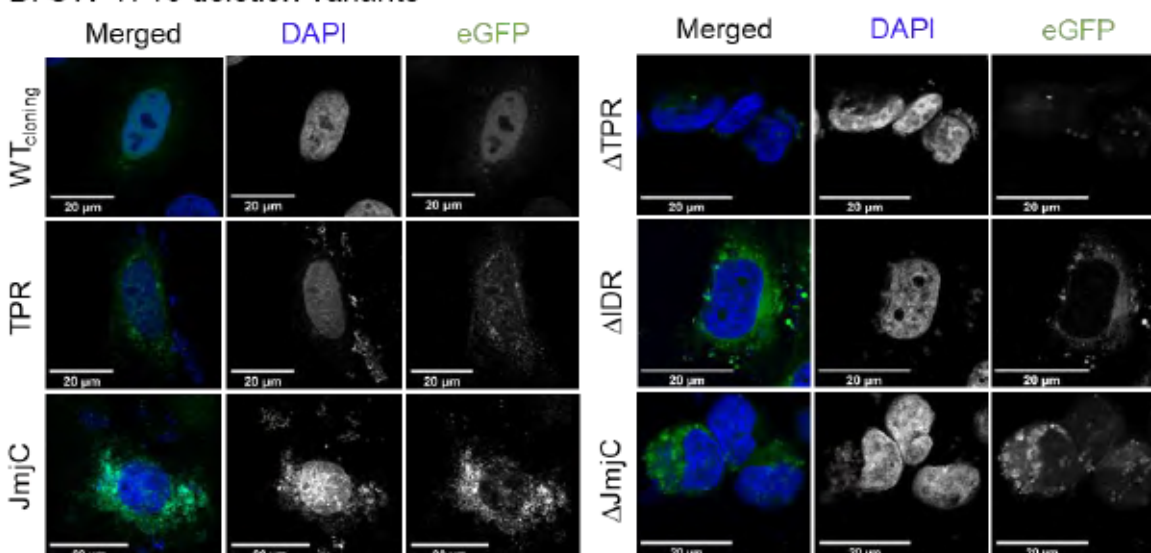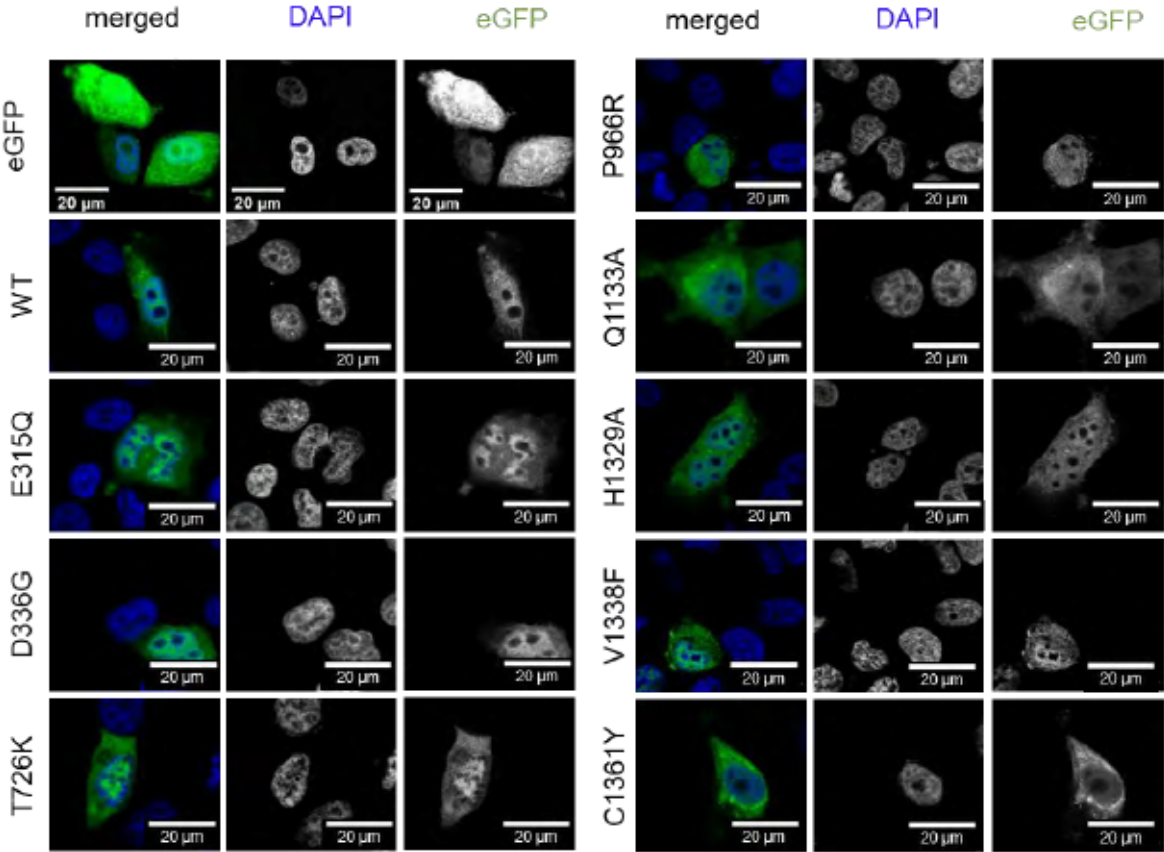

Figure 11: Transient expression of eGFP-KDM6A variants in SW-1710 cells. A. Localization of transiently transfected KDM6A substitution variants in SW-1710 using confocal microscopy and post-processing with HuygensPro 20.08. All variants localized to the nucleoplasm and cytoplasm. Occasionally, KDM6A-positive micronuclei and cytoplasmic DNA were observed for all variants

including wildtype (arrows). **B**. Abnormal localization pattern compared to the eGFP-KDM6A WT, WT_{cloning} and substitution variants: Leakage of DNA from the nucleus and subsequent extranuclear DNA patches are common features with all truncated variants in SW-1710.

A. T-24 substitution variants

B. T-24 deletion variants

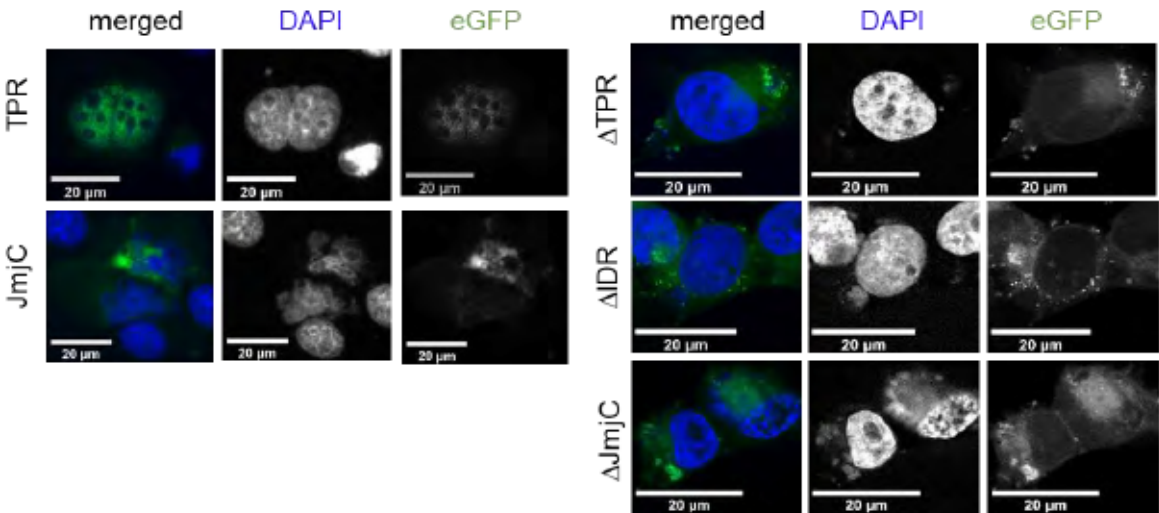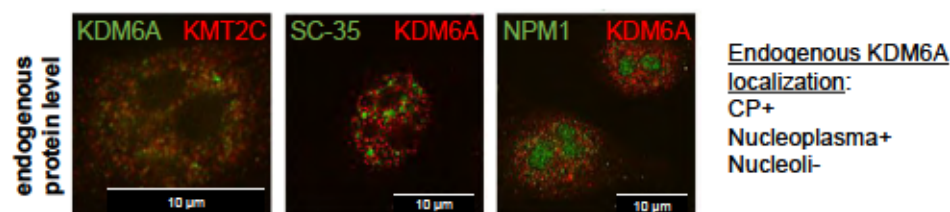**Figure 12: Transient expression of eGFP-KDM6A variants in T-24 cells. A.** Localization of transiently transfected KDM6A substitution variants in T-24 using confocal microscopy and post-processing with HuygensPro 20.08. All variants localized to the nucleoplasm and cytoplasm. Occasionally, KDM6A-positive micronuclei and cytoplasmic DNA were observed for all variants

including wildtype. **B**. Abnormal localization pattern compared to the eGFP-KDM6A WT and substitution variants: Leakage of DNA from the nucleus and subsequent extranuclear DNA patches are common features with all truncated variants T-24.

A. Endogenous localization of KDM6A, KMT2C, SC-35 and NPM1 in HBLAK cells

Endogenous KDM6A localization:
CP+
Nucleoplasma+
Nucleoli-

B. Free eGFP, WT and deletion variants in HBLAK

eGFP localization:
CP+
Nucleoplasma+
Nucleoli+

KDM6A WT localization:
CP+
Nucleoplasma+
Nucleoli -

KDM6A mutant localization:
CP +
Nucleoplasma +
Nucleoli –
Majority population

CP +
Nucleoplasma (+)
Nucleoli –
Minority population

CP +
Nucleoplasma +
Nucleoli -
Majority population

CP +
Nucleoplasma -
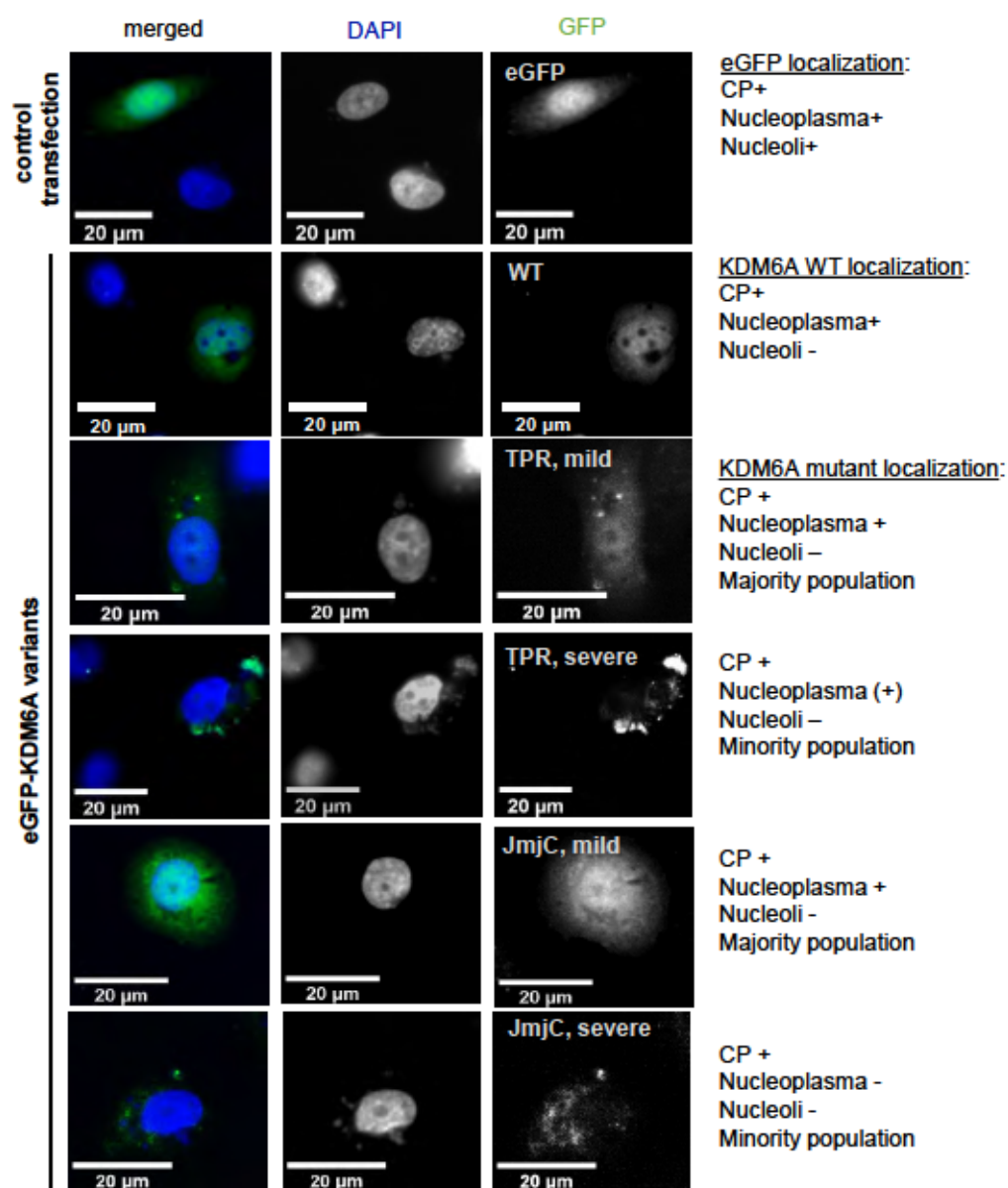Nucleoli -
Minority population

Figure 13: Endogenous KDM6A and variants in untransformed HBLAK cells. A. Localization of endogenous KDM6A, KMT2C, SC-35 and NPM1 appears natural. Imaged with SIM. **B.** Abnormal

localization pattern of eGFP-KDM6A TPR and JmjC compared to the WT, both severe and mild effects. As for T-24 and SW-1710, leakage of DNA from the nucleus and subsequent extranuclear DNA patches are common features for the tested variants in HBLAK.

KDM6A WT and NPM1 form nucleoplasmic populations while truncated KDM6A variants likewise form complexes with NPM1 at extranuclear DNA segments.

*(Text and figures taken or adapted from manuscript 1. (see **Publications** section, page 187)*

Images of eGFP-KDM6A WT, T726K and JmjC were further analyzed via scatterplots, cellular backmapping and line profiles. For eGFP-KDM6A WT and NPM1 (**Fig. 14A, left panel**), two unique fractions appeared at the y- or x-axis, representing signals in only one of the two channels. The corresponding eGFP-KDM6A WT-only fraction is shown in green and the NPM1-only fraction in red. A third fraction with green and red signals of different intensities is shown in orange and named from here on 'intermediate' fraction from here on. Cellular back mapping of the selected populations (**Fig. 114A, middle panel**), clearly showed the KDM6A WT-only signal predominantly in the cytoplasm and only a minor fraction in the nucleoplasm. As expected, the NPM1-only signal was present in nucleoli and in nucleoplasm. The intermediate fraction was always associated with the NPM1-only fraction at the nucleoli rims and within the nucleoplasm, indicating a dynamic exchange of NPM1-only, mixed complexes and KDM6A WT-only fractions at specific sites within the nucleus. The line profile through the nucleus with DAPI as a DNA indicator (**Fig. 14A, right panel**) corroborates these findings: High NPM1 (red) signal intensities were exclusively found in nucleoli, whereas green-red overlapping signals might represent the KDM6A WT assemblies with NPM1. As described above, truncated KDM6A variants elicited more severe nuclear DNA release and nuclear damage (**Fig. 11-13**). Minimizing our experimental workload, we reduced our further analysis to KDM6A JmjC and KDM6A TPR. Scatterplot analysis and line profiling of the truncated variants with NPM1 clearly indicated an enriched, colocalizing intermediate fraction at extranuclear DNA segments. As observed before, KDM6A JmjC and TPR variants were also localized to the cytoplasm. In contrast to KDM6A WT, these variants were always associated with DNA. We identified KDM6A and NPM1 compositions of variable stoichiometry and different localization, which were hardly detectable by Co-IP/WB analysis (**Fig. 8C**). Importantly, scatterplot analysis and the cellular back mapping approach highlighted differences among the wildtype, substitution and truncated KDM6A variants. The T726K variant represents a non-wildtypic mild phenotype with a low amount of DNA release. Even more so, the severe phenotype of the truncated variants is characterized by a massive DNA release and forms complexes of nearly equal stoichiometry with NPM1.
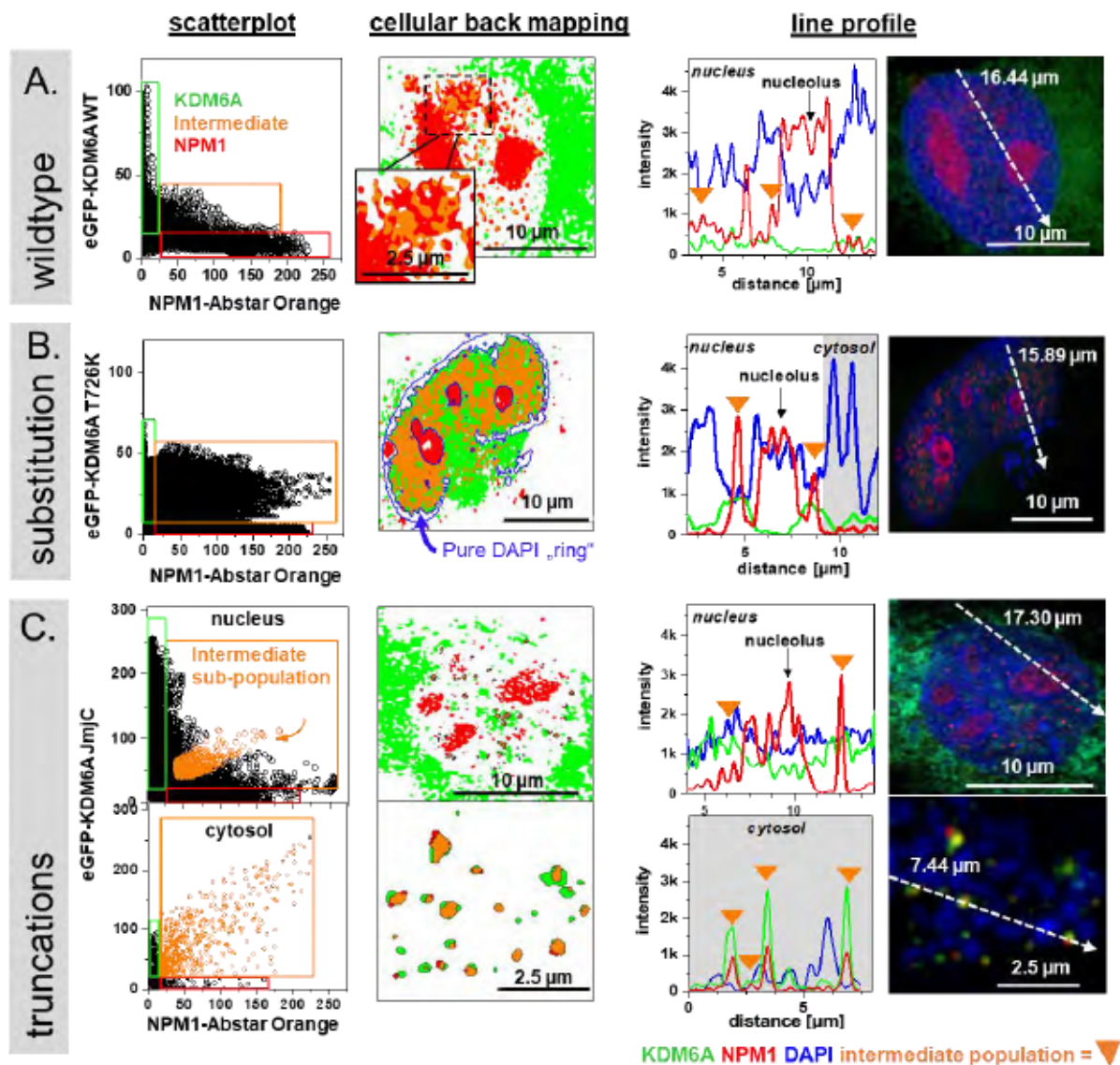
Figure 14: Cellular back mapping, scatterplots and line profile of eGFP-KDM6A variants.
Cellular back mapping identifies co-localizing NPM1-KDM6A protein populations in nucleoplasm and at extranuclear DNA. This effect has a high occurrence in JmjC variant (**C.**) and lower occurrence in WT and T726K (**A.** and **B.**). **Left panel:** Scatterplot presentation of KDM6A (green box) and NPM1 (red box) and one intermediate population representing a mixture of KDM6A-NPM1 complexes (orange box). **Middle panel:** All three populations were back-mapped to the cellular compartment. **Right panel:** Line profile through the nucleus as indicated by an arrow in the image. DAPI (blue), NPM1 (red) and KDM6A variants (green) signals were overlaid. The orange triangles indicate the intermediate populations from the scatterplot.

153

The severe phenotype of KDM6A truncation variants is characterized by mitotic defects, DNA release and significantly decreased cell viability.
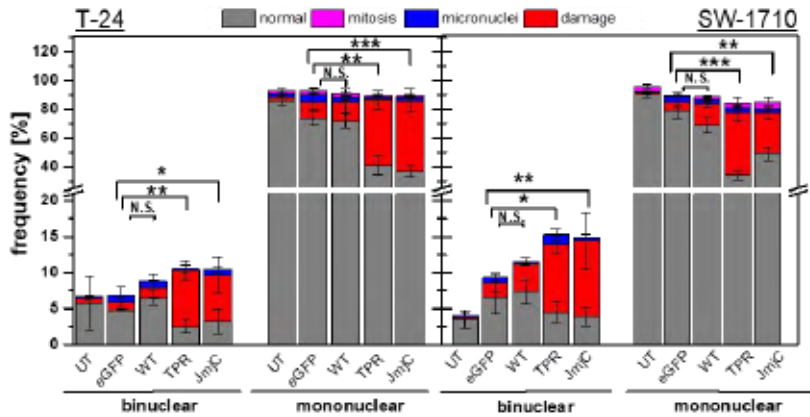
*(Text and figures taken or adapted from manuscript 1. (see **Publications** section, page 187)*
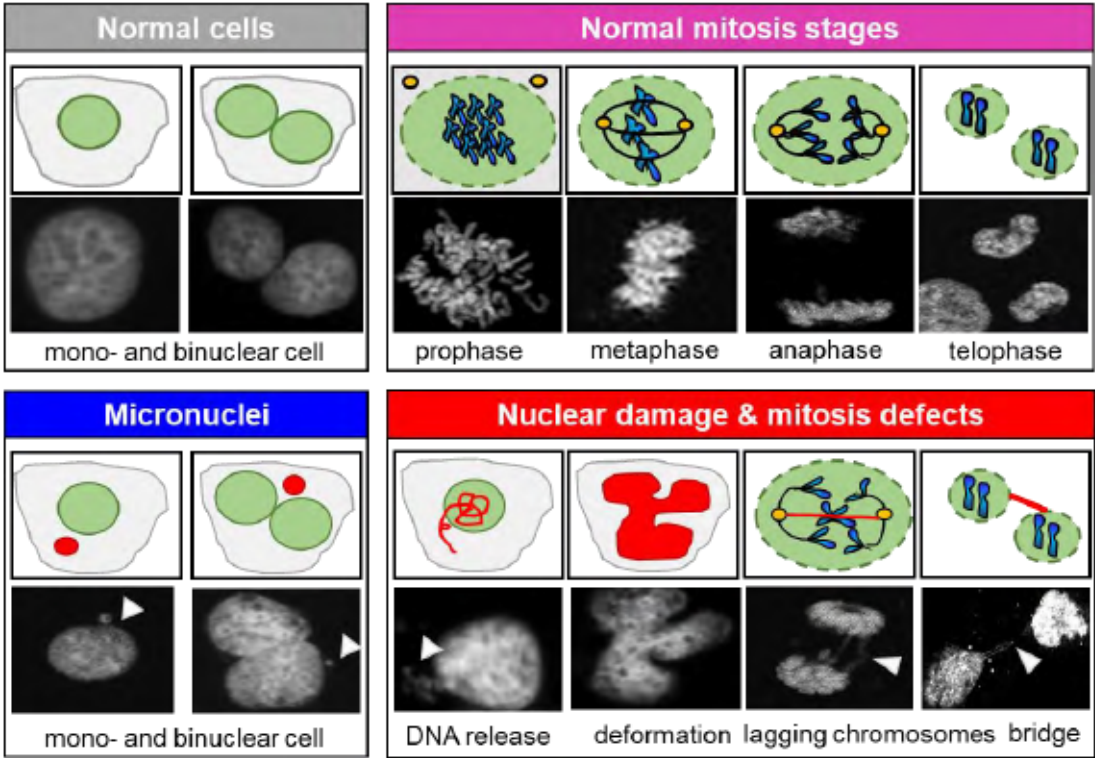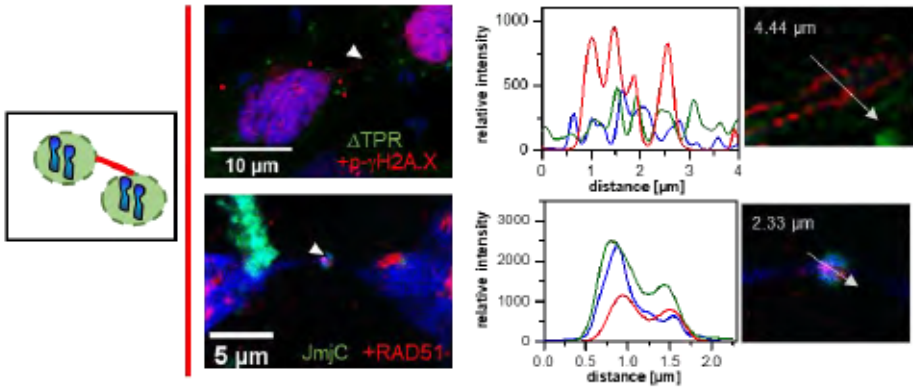
Figure 15: Quantification of severe effects observed for truncated eGFP-KDM6A variants in UCCs. A. Scoring criteria used to quantifiy the occurrence of nuclear damage. **B.** Truncated KDM6A variants promote a significant decrease in cells with normal phenotypes in both T-24 and SW-1710. Frequency of different phenotypes scored as normal interphase (grey) and mitosis (pink), micronuclei (blue) and damage (red) in bi- or mononuclear T-24 and SW-1710 cells. In mononuclear cells, normal cells decreased significantly from eGFP control to both eGFP-KDM6A TPR and JmjC, but not WT. In binuclear cells, damaged cells increased significantly from eGFP control to eGFP-KDM6A TPR and JmjC variants, but not WT. Cells with micronuclei were not significantly enriched. Overall, we observed a non-signifcant trend towards an increased number of binucleated cells. See **Chapter 4-Table supplement 8** for full statistics. T-test using two-tailed hypothesis, significance levels: * = $P \leq 0.05$, ** = $P \leq 0.01$, *** = $P \leq 0.001$, see **Chapter 4-Table supplement 9** for detailed *P*-values. **C.** Line profiles through lagging chromosomes and the "knot-like structures" of the chromatin bridges found in samples transfected with eGFP-KDM6A TPR or JmjC indicating overlapping signal intensities of KDM6A variants (green) with DNA damage markers RAD51 and p-γH2AX (red).

As DNA release has been observed earlier, we analyzed whether the typical indicators for a DNA damage response are activated by KDM6A truncation variants, namely accumulation of phospho-γH2A.X and RAD51. Visual inspection of phospho-γH2A.X in cells with truncated variants revealed defects in mitosis (**Fig. 15C**). These defects occurred in anaphase as lagging chromosomes, multiple fragmentation events and in telophase and cytokinesis by persisting chromosome bridges and accumulation of DNA damage sites at chromosome bridges. To quantify our observations, we established a scoring system to determine the amount of cells with severe phenotypes per variant (**Fig. 15A**). Here, we discriminated between mono- and binucleated cells and quantified cells with damages (cytoplasmic DNA release, extreme nuclear deformation, lagging chromosomes and chromosome bridges), micronuclei, normal interphase or mitotis. Both KDM6A deletion variants, TPR and JmjC, elicited a significant increase in DNA damage in mono- and binucleated cells. We observed a trend for cells transfected with eGFP-KDM6A WT, TPR and JmjC to be detected in binuclear cells. To evaluate whether the nuclear damage promotes apoptosis, we performed an Annexin V- based apoptosis assay 48 h post transfection. Up to 50,000 transiently transfected cells per variant, stained with Annexin-V allophycocyanin (APC) and propidium iodide (PI), were fixed and analyzed with fluorescence activated cell sorting (FACS). The measurements indicate that cell viability significantly decreases in eGFP-KDM6A TPR and JmjC variants in both urothelial cancer lines (**Fig. 16A/B**). This mirrors the results of the scoring experiments very well, pointing towards a direct connection between the nuclear damage and apoptosis induction.
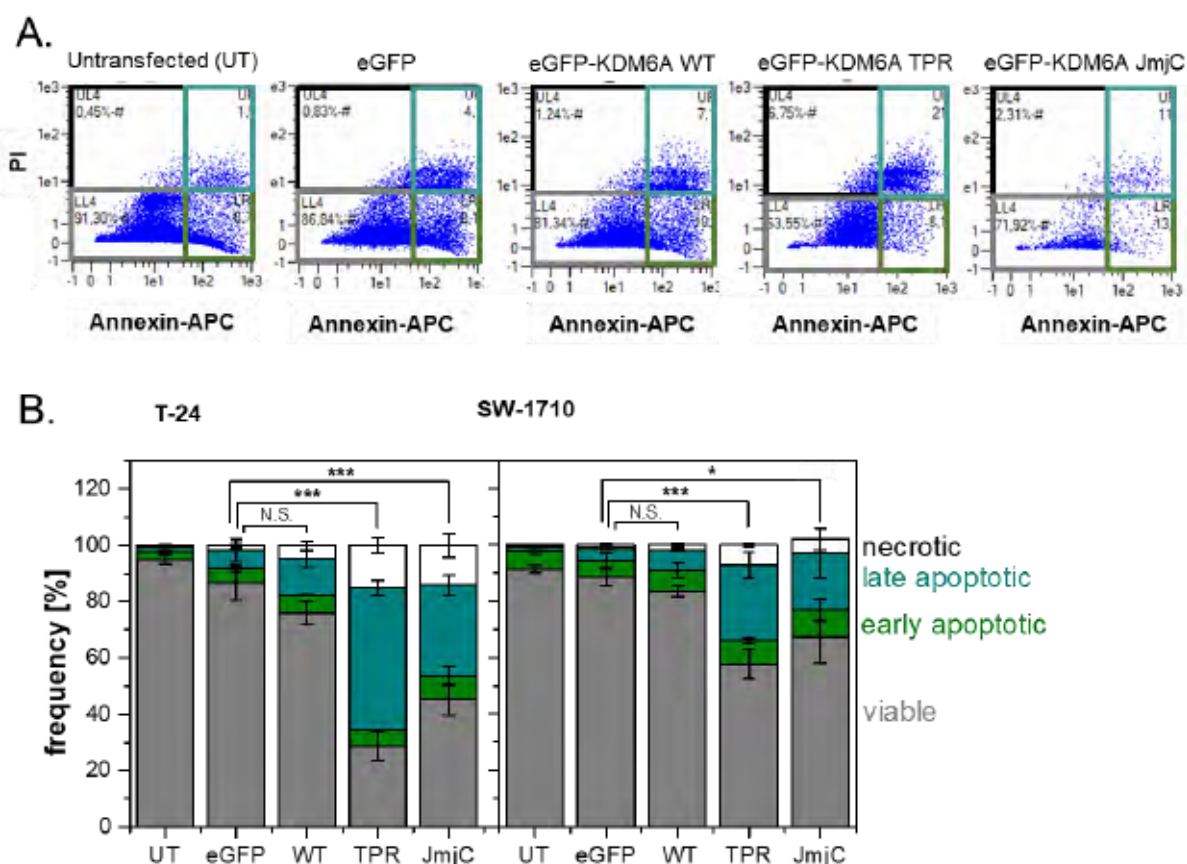
A.

Untransfected (UT) | eGFP | eGFP-KDM6A WT | eGFP-KDM6A TPR | eGFP-KDM6A JmjC

B. T-24 | SW-1710

Figure 16: Truncated KDM6A variants decrease the amount of viable in T-24 and SW-1710 cells and exhibit mitosis errors. **A**. FACS analysis in Annexin V based apoptosis assay. Transfected eGFP positive cells were gated based on the threshold obtained from untransfected cells, which were then used to plot PI (membrane permeability) against Annexin-V-APC (apoptosis marker). The plot was divided into four quadrants, representing the viable population (lower left), early apoptosis (lower right), late apoptosis (upper right) and necrosis (upper left). **B**. Statistics derived from triplicate measurements. eGFP-KDM6A TPR and JmjC show a significant decrease in cell viability in comparision to the eGFP control in both cell lines. T-test using two-tailed hypothesis, significance levels: * = $P \leq .05$, ** = $P \leq 0.01$, *** = $P \leq 0.001$, see **Chapter 4-Table supplement 10** for detailed *P*-values. **C**. Graphic summary of cellular phenotypes observed with KDM6A mutation variants, depicting the impact on localization, mitosis, apoptosis and protein assemblies.

Further directions: Initial design and testing of a new generation of eGFP-KDM6A constructs utilizing the Spot-Tag® and enhanced solubility.

In an attempt to enhance our expression quality, we first optimized the TPR-construct. We added a soluble linker region (Sol.-tag, 5x Lys), elongated the construct into the IDR-region (aa 391-426) and added a C-terminal SPOT (see **Fig. 17A**). The SPOT-tag is an affinity-body which is specifically binding a PDRVRAVSHWSS peptide sequence. It enabled peptide-based elution from beads and enables (in contrast to eGFP) experiments with

purified SPOT-tagged KDM6A in solution. The SPOT and eGFP tags can be pulled equally good with their respective beads (**Fig. 17B**). The lane from the SPOT-bead pull appears even cleaner. There is no residual eGFP-TPR-SPOT left in the supernatant (SN). Imaging (**Fig. 17C**) reveals a higher and more homogenously distributed expression. The protein strongly localizes into the nucleus. The 5x Lys Sol.-tag might increase nuclear import [71]. The amount of cellular damage needs to be assessed further in apoptosis assays and image-quantification.
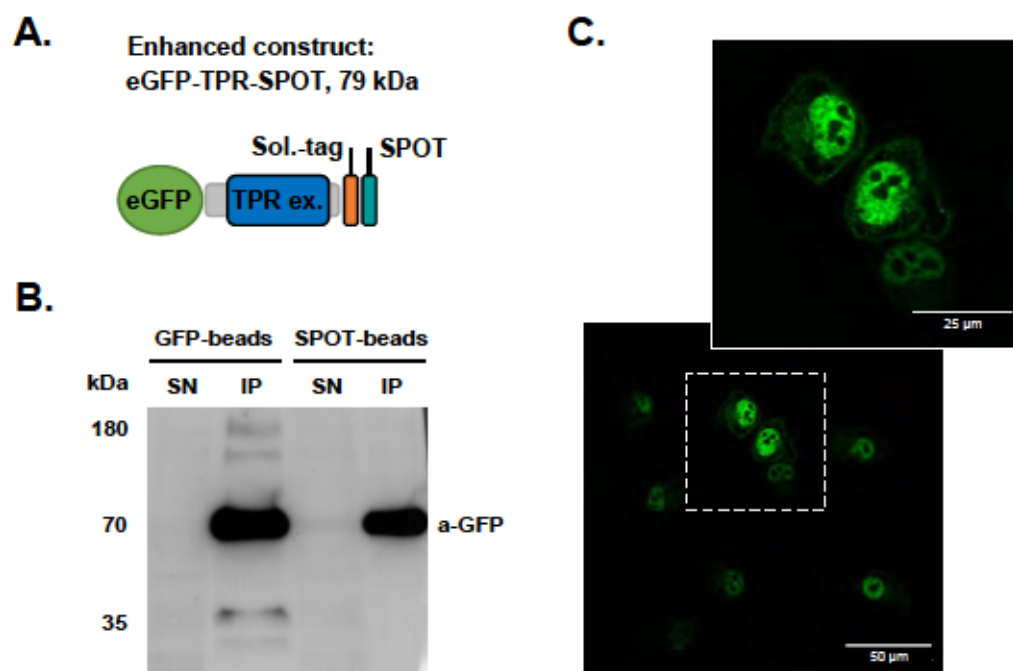
Figure 17: Optimized TPR expression construct. A. eGFP-TPR-SPOT construct. TPR ex. consists of aa 1-426, Sol.-tag consists of 5x Lys. **B.** Western blot of SPOT-beads vs GFP beads, pulling eGFP-TPR-SPOT. Detection with a-GFP antibody. **B.** Imaging of T-24 cells expressing eGFP-TPR-SPOT.

3. Discussion

(*Text taken or adapted from manuscript 1. (see **Publications** section, page 187)*

Establishing a comprehensive tool kit for the characterization of KDM6A

Starting this project, we found next to no comprehensive and clear attempt to establish a structure-function relationship between the phenotypes observed, the positions mutated or even the domains deleted while considering the characteristics of the cell line used in the experiment. To improve this situation, we developed and utilized a comprehensive analysis tool kit to understand the relationship and interplay of known and predicted regulatory features of the multi-domain protein KDM6A. The tool kit consisted of a demethylase assay, imaging techniques, immunoprecipitation and Western blot as well as FACS. The eGFP-tag

was utilized as a universal affinity and fluorescent tag. We demonstrate how single substitution mutations and deletions of the main three functional and regulatory domains, TPR, IDR and JmjC, affect the intrinsic properties of the target protein and its interactions with the cellular environment. We could show that mutations within the JmjC domain affected the catalytic activity. Although nuclear localization changed little, interactions with RBBP5 and NPM1 were affected, especially in truncated variants, where we observed a cellular mechanism to dispose of harmful KDM6A variants, namely by release of DNA-KDM6A complexes into the cytoplasm. Harmful KDM6A variants cause mitosis defects and DNA damage which promotes cell death.

Selection of KDM6A variants, biochemical quality control and evaluation of specificity of the phenotypes observed

When selecting variants for mutagenesis-based experiments there are two main considerations to make: First, what effect on the experimental readout do we deduce for the variant and, secondly, can the mutation influence the variant in any uncontrollable way outside from the readout we want to achieve. Considering this, we would think of some visible effect for all cancer-derived substitution variants by design, since they likely contributed to the cancer pathogenicity. A single point mutation is not necessarily changing the protein as a whole, especially at the size and multi-domain structure of KDM6A. Still, even single substitutions could cause unspecific effects, although unlikely. The truncated variants on the other hand are a lot harder to judge. While the functional effects might be a lot stronger here, it is challenging to pinpoint the source of an effect to a specific structural feature and at the same time exclude unspecific effects like folding issues or solubility decrease. While the substitution variants all looked stable, the truncated variants had aberrancies in the fluorescence emission spectra, cellular expression levels and even, in the case of the ΔIDR variant, problems to be visualized on western blot. The latter event prompted us to exclude the ΔIDR variant for further experiments, judging its solubility as too low to be isolated properly. Overall, we see the phenotype encountered with the truncated variants as too specific to be mainly caused by solubility or stability issues. Still, the phenotype could be influenced and blurred by the lowly performance of the truncated variants.

Changing demethylase activity of KDM6A via mutagenesis is not just one-dimensional, but harbors structural implications and post-transfection time dependency

First, we developed an ELISA-based demethylase assay suited for use with eGFP-tagged KDM6A variants. Our experiments revealed, that substitution variants within the JmjC (catalytic domain) possessed reduced or abolished demethylase activity, especially if amino

acids involved in peptide recognition, peptide binding or stabilizing were changed. Remarkably, the demethylase activity was also reduced in the IDR-variant T726K, but not in T726V, in a post-expression time-dependent manner. Therefore, we predict a unique functionality for the amino acid position K726 connected to maybe a faster, possibly partial, turnover of the protein after expression. Among the selected cancer-associated point mutations, the hotspot mutation T726K is the third most frequently listed mutation in Cosmic v92 across all tissues, further highlighting the potential. We predicted and tested K726 as a possible methylation site, but mass spectrometry analysis did not show any PTM at K726. There might also be some kind of recognition site involved. More comprehensible, the demethylase activity of truncated KDM6A variants was determined by the presence or absence of the JmjC domain. As expected, KDM6A JmjC, ΔTPR and ΔIDR exhibited demethylase activity, whereas TPR and ΔJmjC did not. However, removal of other domains negatively affected protein stability and solubility, especially for KDM6A ΔIDR. KDM6A JmjC and the flanking zinc-binding domain are known to recognize and bind several amino acids between H3R17-H3T32 of the H3K27 di- and tri-methylated N-terminal histone tail to ensure substrate specificity [15]. Consequently, mutations in the JmjC and flanking domains have a high potential to impede or even abolish the catalytic activity. As an example, the P966R mutation localized in the early JmjC is situated at a loop connecting the ribcage-like structure made of β-sheets, which surrounds the active center of KDM6A. Replacing the proline might destabilize the loop's turning angles and weaken the whole structure, ultimately abolishing the demethylase activity. However, it is unknown to what extent mutations in the TPR and IDR might contribute to KDM6A demethylase activity. While those variants with N-terminal or central truncations are mostly artificial, C-terminally truncated variants caused by nonsense mutations make up almost a quarter of all listed *KDM6A* mutations in COSMIC v92. A prevalent nonsense mutation is Q555*, which may be partly explained by the observation that it represents a hotspot for APOBEC3-mediated mutations, frequently appearing in multiple cancer types including bladder cancer. Functionally, the Q555* fully lacks the JmjC domain and has a partial deletion of the IDR. Such IDR/JmjC nonsense mutations would abolish demethylase activity by truncation or deletion of JmjC. Another example is the KDM6A fragment found in the urothelial cancer cell line T-24 with heterozygous mutations at E895* and E902*. These similar-sized variants are still expressed endogenously and can be detected as a ~97 kDa band on WB[48]. It remains speculative whether these fragments display dominant negative effects and are actively involved in generating the cancerous T-24 phenotype. Other variants, like the moderately frequent Q333*, lack both IDR and JmjC and could potentially impair TPR8 functionality.

Our truncation variants help to reveal and map different binding modes of KDM6A with RBBP5

Apart from implications on demethylase activity, our variants have impaired interactions with other proteins. We have recently shown that KDM6A associates with RBBP5 in urothelial cancer cell lines dependent on the mutation status of KDM6A and KMT2C/D proteins [48] suggesting vital interactions of KDM6A with the COMPASS complex. Here we showed that TPR and IDR substitution variants did not affect RBBP5 binding. However, especially the KDM6A mutation D336G has been shown to be predominantly cytoplasmic [23] in HeLa cells presumably due to impaired binding to ASH2L in a pull-down experiment and consequential reduced nuclear import by the KMT2C/D complex [23]. In a previous study we observed that KDM6A nuclear import was strongly decreased after double, but not single, knock down of KMT2C and KMT2D proteins [48]. Systematic deletion of KDM6A domains clearly indicated that all domains, including TPR and IDR, are necessary for proper binding of RBBP5 independent of demethylase activity. This could indicate that the binding motif of RBBP5 stretches into the IDR or that there are multiple binding epitopes distributed throughout the IDR, possibly even assembling in three-dimensional topology. Although JmjC alone does not bind RBBP5, the presence of this domain enhanced binding in KDM6A WT compared to KDM6A ΔJmjC and KDM6A TPR, this might either be due to an overall stability-enhancement by the JmjC domain or a second mode of indirect binding to RBBP5 via the Histone H3 tail. A recently published study indicated that RBBP5, WDR5 and the KDM6A JmjC domain share similar recognition and binding motifs at the Histone H3 tail [58].

The observed uniformity in localization for all variants likewise demands multiple options for nuclear import of KDM6A

Looking at the localization of KDM6A, we have to discuss multiple possible import pathways, since all variants were located in the nucleoplasm independently of their mutation status, although to different extents. The theory we verified in our earlier paper [48], was working with a nuclear co-transport of KDM6A alongside KMDT2D/C and the surrounding WRAD complex. KO of both KDMT2s would abolish nuclear import. Now our results and others [23] show that the interaction of KDM6A with core components of the WRAD, especially RBBP5, ASH2L and WDR5, are occurring via the TPR domain. Still, we see nuclear localization of all variants. This might be due to leakage since our truncated variants are seemingly rupturing the nucleus, resulting in a general distortion of nuclear import. Alternatively or concurrent, there could be another pathway or interaction using the JmjC domain as interaction for import. The nuclear localization sequences (NLS) are o Lastly, there is always the option for passive nuclear migration via pore complexes (NUP) but KDM6A is by far too large [72] to enter via NUP.

The observed severe phenotype can be explained from different perspectives and is likely connected with multiple underlying pathways

Truncations, but not substitution variants, are characterized by eliciting i) cytoplasmic DNA release, ii) enhanced levels of RAD51 and phospho-γH2AX as indicators of DNA damage, and iii) defects of mitosis caused by missegregated chromosomes at anaphase and persisting chromatin bridges at telophase and cytokinesis. All observed effects occurred on a short time scale within 36-48 h. While transient and stable overexpression of the KDM6A WT reduces long-term cell growth and colony formation [48], we never observed effects of this kind, neither short-term nor long-term. Do note that the expression levels achieved with transient transfections were considerably higher than the low levels we encountered in stably transduced cell lines. In general, aneuploidy, replication stress and mitosis errors are common in cancers[73]. Accordingly, all cancer cell lines used in this study exhibit these features on a basic level but they are profoundly enhanced after induction of KDM6A truncation variants. Among the severe phenotypes, cytoplasmic DNA release was most commonly observed. All KDM6A truncation variants were associated (directly or indirectly) with the DNA released from the nucleus as indicated by localization analysis and the cellular back mapping approach. Nuclear DNA release is the presence of cytoplasmic DNA caused by a yet unknown mechanism. We speculate that appearance of cytoplasmic DNA could be caused by i) pulverized micronuclei or ii) chromosome fragments without envelope or iii) active nuclear release due to impaired nuclear integrity or DNA damage [74]. Our observations also point towards mitotic defects followed by apoptosis. However, we cannot rule out additional mechanisms, as we do not have conclusive data on cGAS/STING activation that is expected in response to cytoplasmic DNA accumulation [75]. Moreover, introduction of KDM6A variants, especially truncated variants, elicited elevated phospho-γH2A.X levels. Phospho-γH2A.X is activated during the DNA damage stress response [28]. Enrichment of proteins involved in DNA repair and stress response (DDR) appeared in our MS-data analysis from three different urothelial cancer cell lines with stably or transiently transfected KDM6A WT. Notably, KDM6A activity in differentiating embryonal stem cells has been linked to DNA damage response pathways by colocalization with γH2A.X positive foci [28]. In addition, as an oxygen-dependent enzyme, KDM6A serves as a sensor to control chromatin and cell fate [76]. Thus, overexpressed (with a high dose-effect) and impaired KDM6A variants as well as oxygen-related stress have the tendency to increase DNA damage. This phenomenon was also observed in diabetic kidney disease [30]. An additionally prominent feature of truncated KDM6A variants was a high degree of co-localization with NPM1 at extranuclear DNA. The combined proteins appeared as a mixed population in the nucleoplasm. As NPM1 is involved in rRNA processing, ribosome maturation and shuttling of ribosomal subunits between nucleoli, nucleoplasm and cytoplasm

[77], KDM6A might be involved in these processes aswell. However, the co-occurrence of NPM1 and KDM6A truncation variants may rather result from the role of NPM1 as a chaperone [78]. At this stage, we cannot completely rule out activation of the unfolded protein response pathway (UPR) or ER proteostasis [79] by truncated KDM6A, but consider it rather unlikely for the following reasons: i) We observed correct nuclear localization of all truncated variants, ii) Perinuclear aggregation was observable in all KDM6A WT, substitution, truncation and control (eGFP) variants and iii) in truncated variants with a severe phenotype, KDM6A protein was always associated with DNA and never freely distributed throughout the cytoplasm. It remains possible that UPR stress sensors contribute to activation of the nuclear DNA damage response [79], which is confirmed by enhanced phospho-γH2A.X levels. As KDM6A itself might act as a critical stress sensor, it is difficult to ascertain at this stage which signaling cascade might explain our observations best.

Our observations hint at possible new functions or involvement of KDM6A in the cell cycle.

Specifically, the following questions are raised: i) What is the role of KDM6A during mitosis and to which extent is any such function dependent on its catalytic activity and its interplay with RBBP5 and further components of the KMT2C/D-COMPASS complex? Notably, many lysine demethylases (KDM) have cell-cycle specific roles [80, 81]. KDM4C, KDM1A and KDM7B have already been linked to mitosis by regulation of chromosome segregation, transcriptional activation of mitotic checkpoint complex components (see refs in [80]). Moreover, WDR5 and KMT proteins, likely KDM6A interaction partners, have also been shown to be involved in mitosis [82, 83]: WDR5 is part of the midbody in the spindle apparatus [81]. Intriguingly, we already found endogenous KDM6A located along the midbody in HBLAK cells. Ii) Under which conditions and in which manner do NPM1 and KDM6A directly or indirectly interact? The multifaceted functions of NPM1 in chromatin remodeling, DNA repair, cell cycle control, apoptosis, mitotic spindle, centromere and cytoskeleton binding [78, 84] and its prominent enrichment in MS analysis suggests an important link between both proteins, that have not been described before.

4. Conclusion & Outlook

During our work,we also experienced some points, that should be critically considered in future project research:

i)   Expression and stability of truncated KDM6A variants: The considerably low expression levels complicated in vitro assays (ELISA based activity assay, Co-IP and Western Blot) a lot of and made the harvest of larger protein amounts very challenging. Furthermore, the issues with stability and solubility are surely interconnected with this problem to some degree.

ii) eGFP fusion proteins and small tags for purification: Although the eGFP-tag works pristine in affinity and imaging experiments, we have difficulties in eluting the protein in its active, folded state after binding to GFP beads (in our case GFP dynabeads from chromotek), again limiting the scope of in vitro methods. With our newly introduced SPOT-tag®-coupled and solubility-enhanced constructs we already achieved very good results and we are going to further continue optimizing it.

iii) TPR: interaction via the N-terminally located TPR is actuallythe most interesting and promising target in our system, especially with all the new pathways and putative proteins of interest. We would also like to focus more on biophysical methods and complementing modeling after achieving more structural knowledge, especially of the TPR. constructs, which are designed and optimized according to the criticial issues addressed above (stability, purification and labeling tag…) will certainly help in establishing these methods and integrating them in our interdisciplinary project.

5. Methods

(*Text taken or adapted from manuscript 1. (see **Publications** section, page 187)*)

Cell lines and cell culture

Parental T-24 [37-40], SW-1710 [41], VM-CUB1 [40, 85] and RT-112 [86, 87] urothelial carcinoma cell lines were obtained from the DSMZ (Braunschweig, Germany). Cells were cultured and treated in DMEM GlutaMAX-I (Gibco, Darmstadt, Germany) with 10% fetal bovine serum (FBS; Gibco™, Thermo Fisher Scientific) and 100 U/ml penicillin/100 µg/ml streptomycin (Sigma-Aldrich, Darmstadt, Germany), except for HBLAK cells, which were solely cultured in CnT-Prime Epithelial Culture Medium (CELLnTEC, Bern, Switzerland) without any additives. Cells were incubated at 37 °C in a humidified atmosphere with 5% CO2. STR (short tandem repeat) profiling via DNA fingerprint analysis was performed for all cell lines in this study and is available upon request.

Generation of eGFP-KDM6A substitution and deletion variants

All constructs are listed in **Chapter 4-Table supplement 11.** eGFP-KDM6A wildtype (WT) was synthesized by BioCat (Heidelberg, Germany) by cloning a codon-optimized eGFP-KDM6A (both sequences full-length, KDM6A main isoform 1 (Uniprot ID O15550) without additional linker between eGFP and KDM6A) into the pcDNA 3.1(+) vector, using NheI and NotI as flanking restriction sites. Generation of eGFP-KDM6A substitution variants was done by using site-directed mutagenesis (SDM) (**Chapter 4-Figure supplement 5**) using 10 ng eGFP-KDM6A WT plasmid and mutagenesis primers (**Chapter 4-Table supplement 12**).

163

1.25 U PrimeSTAR GXL DNA Polymerase (Takara Bio, Kusatsu, Shiga, Japan) was used in PCR reactions. Successful PCR amplification and product length was checked by gel electrophoresis. After DpnI (NEB, Ipswich, MA, USA) digestion (20 U for 1 h at 37 °C), SDM amplicons were transformed into *Escherichia Coli* XL10-Gold® (Stratagene, Santa Clara, CA, USA) and spread on LB amp plates. Colonies were picked, grown and DNA was isolated using a QIAprep Spin Miniprep Kit (QIAGEN, Hilden, Germany). After sequencing, DNA from positive KDM6A substitution clones was re-transformed into *E. Coli* XL10-Gold® (Stratagene) and purified at a large scale using NucleoBond Xtra Maxi Plus EF kit (Macherey-Nagel, Dueren, Germany). Substitutions were then re-confirmed by sequencing. Generation of eGFP-KDM6A deletion variants was done by "modularized" cloning of three inserts: TPR, res. 1-390, IDR, res. 391-885 and JmjC, res. 886-1401 from the original wildtype eGFP-KDM6A pcDNA3.1(+) into the pEGFP-C1 vector for the desired combinations. Each restriction enzyme (RE) site produces a two amino acids long linker. Constructs with one insert (eGFP-TPR, eGFP-IDR) have BspEI and HindIII as flanking RE sites. Constructs with two inserts (eGFP-KDM6A ΔTPR, ΔIDR, ΔJmjC) have BspEI and EcoRI as flanking RE sites and HindIII as middle RE site. The control construct eGFP-KDM6A$_{cloning}$ has BspEI and KpnI as flanking RE sites and HindIII and EcoRI as mid RE sites. Amplification primers (**Chapter 4-Table supplement 12**) were designed according to the desired combination with appropriate overhangs and synthesized by Eurofins Genomics (Ebersberg, Germany). PrimeSTAR GXL DNA Polymerase (Takara Bio) was used for amplification according to the manufacturer's instructions. T4 DNA ligase (NEB) was used for insert ligation (10 min at RT, 3:1 insert:vector ratio). All restriction enzymes were purchased from NEB. NucleoSpin Gel and PCR Clean-Up (Macherey-Nagel) was used to extract and clean up DNA. Cloning products were transformed into *Escherichia Coli* XL10-Gold® (Stratagene). An appropriate number of colonies were picked, grown and the DNA was isolated using a QIAprep Spin Miniprep Kit (QIAGEN). After sequencing, positively cloned DNA was re-transformed into *E. Coli* XL10-Gold® (Stratagene) and purified in a large scale using a NucleoBond Xtra Maxi Plus EF kit (Macherey-Nagel).

Transient transfection

For transient transfection, cells were seeded into 6-well plates with (imaging) or without (activity, western blot) glass cover slips. 24 h later, cells were transfected at ~70% confluence using X-tremeGENE™ 9 or HP (Roche, Basel, Switzerland; application dependent use) in a 2:1 ratio (v/w) of transfection reagent to DNA. Total DNA transfected per well (9.6 cm^2) did not exceed 2 µg. Transfection was carried out 24 - 48 h for activity assay and Western Blot applications and 36 h for localization analysis.

Cell death analysis by flow cytometry

150,000 cells (SW-1710) and 200,000 cells (T-24) per well of a six well plate were seeded and reversely transfected with XtremeGENE™ HP (Roche). After 16 h cells were split into two wells. 48h post transfection, supernatant and cells were collected, centrifuged at 1000 rpm for 5 min, washed with ice-cold 1x Annexin binding buffer (Serva, Heidelberg, Germany) and centrifuged again. The pellet was resuspended in 75 µl 1x Annexin binding buffer containing 4.5 µl Annexin V-APC (Serva) and 7.5 µl PI (1 mg/ml, Serva) and incubated for 15 min in the dark at RT. The suspension was diluted with 500 µl 1x Annexin binding buffer, centrifuged, washed and fixed with 0.5% formaldehyde (methanol-free) for 20 min on ice. The reaction was stopped with 500 µl 1x Annexin binding buffer. FACS measurements and analysis was performed using the MACSQuant Analyzer X and MACSQuantify Software (Miltenyi Biotec, Bergisch-Gladbach, Germany). In total, 50,000 cells/experiment were analyzed in three independent experiments.

Co-IP and Western blot analysis

An appropriate amount of cells were lysed by suspension in SDS-free RIPA like buffer (RLB) consisting of 50 mM Tris-HCl (pH 7.5), 0.3% CHAPS, 150 mM sodium chloride, 1 mM sodium vanadate (Na_2VO_4), 10 mM sodium fluoride (NaF), 1 mM ethylene diaminetetraacetate (EDTA), 1 mM ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N',-tetraacetate (EGTA), 2.5 mM tetrasodium pyrophosphate $Na_4O_7P_2$. One µm Dithiothreitol and 1x HALT™ protease inhibitor cocktail (Sigma-Aldrich) were freshly added. Lysis was followed by immediate freezing in liquid nitrogen, thawing on ice for 30 min and repeated mixing by pipetting for 30 s each. After centrifuging the pellet, the lysate was either directly separated by SDS-PAGE on a 4-20% gradient gel or used in the following Co-IP steps. For Co-IP, the GFP containing lysate was incubated with GFP-trap dynabeads (Chromotek, Planegg-Martinsried, Germany) for 1 h at 4 C with constantly mild agitation to pull-down eGFP-KDM6A variants and complexed proteins. Dynabeads were magnetically separated and washed three times with Co-IP buffer, resuspended in SDS-PAGE loading buffer, boiled and separated on a 4-20% gradient gel. After running, the gel was transferred onto an activated PVDF-membrane. The membrane was blocked in TBS/0.1% Tween (TBS-T) and 5% BSA for 1 h at RT and subsequently incubated with the respective primary antibody (**Chapter 4-Table supplement 13**) in TBS-T, 1% BSA overnight at 4 °C. The membrane was washed three times in TBS-T at RT. The secondary antibody (**Chapter 4-Table supplement 13**) was applied for 1 h at RT in TBS-T, 1% BSA. The membrane was washed again three times and Clarity™ ECL (Bio-Rad, Hercules, CA, USA) was used to develop the signal.

ELISA-based demethylase activity assay

Preparation and lysis of the cells was done the same way as described for Co-IP. However, before the GFP-trap dynabeads were added, the fluorescent emission signal of the lysate was measured (470 nm ex., 485-650 nm em.) in a microliter cuvette. The dynabeads were then mixed in the lysate for 1 h at 4 °C under constant, mild agitation. Beads were then magnetically separated, the fluorescent emission of the remaining lysate was measured again with the same specifications as above. The delta in the emission spectra before and after bead incubation in the range of 500-530 nm was used to calculate the amount of eGFP-KDM6A pulled out of the lysate in each run. To eliminate residual RLB buffer (**crucial step!**), beads were washed two times with the activity assay (AA) buffer (50 mM TRIS-HCl (pH 7.45), 0.02% Triton X-100, 100 µM α-ketoglutarate, 50 µM $Fe(NH_4)_2(SO4)_2*6\ H_2O$, 100 µM ascorbic acid, 1 mM TCEP, cofactors and TCEP being added freshly to avoid oxidation. H3K27me3 (ProteoGenix, Schiltigheim, France) and H3K27me2 (BioCat) peptides (**Chapter 4-Table supplement 14**) were dissolved in AA buffer and mixed with the loaded beads. The beads were incubated with the peptides for 4 h at 30 °C while maintaining constant suspension. Afterwards, the beads were magnetically separated and the supernatant containing the biotin-labeled peptide was loaded onto a streptavidin-coated 96-well plate (triplicate per variant/control, 50 µl per well). After 1 h of biotin binding at RT and removal of the solution, the wells were loaded with 50 µl of the a-H3K27me2 antibody in 0.1% TBS-T and incubated for 1 h at RT while shaking gently. The wells were then washed three times with 150 µl TBS-T. Subsequently, 50 µl of a 1:1000 alkaline phosphatase (ALP)-conjugated secondary antibody (**Chapter 4-Table supplement 13**) was added and incubated for 30 min at RT. The wells were washed four times for 5 min. For detection, 100 µl of p-nitrophenyl phosphate (pNPP, Sigma-Aldrich) was added into each well and incubated for 10 min at RT in the dark, mixing thoroughly. The reaction was then quenched by 100 µl 1 M NaOH. The signal was measured in a plate reader at 405 nm absorption, quantified, normalized and fitted. To fit the ELISA readout, we used a 4-PL-regression normalized to the standard curve. The WT fit overlaid with the standard curve was used to directly calculate the relation between the amount of fluorescence signal and demethylated product. WT 4PL-regression fit was applied to calculate c_{50} as a reference point to compare the WT value with the variants. Additionally, eGFP reference measurements were used to calculate the absolute amount of protein input and calculate a specific activity. Initial validation for the assay was done with recombinant full-length KDM6A (Active Motif, Carlsbad, CA, USA).

Immunocytochemistry

Depending on the antibody requirements two protocols were used. For ICC with primary antibody incubation overnight, cells were seeded on coverslips, transiently transfected and fixed at >80% confluence with 1% (v/v) para-formaldehyde, 0.02% (v/v) Triton X-100 for 20 min at RT. Blocking and permeabilization was done with 1% (w/v) BSA, 0.1% (w/v) saponin in PBS for 30 min at RT. After overnight night incubation at 4 °C, coverslips were washed with PBS and incubated with the secondary antibody for 1 h at RT. Slides were washed with PBS, stained with DAPI, washed again frequently with PBS and mounted. For ICC with primary antibody incubation for 1 h RT, cells were prepared as before, but fixed with 4% FA (v/v) for 10 min at RT. Permeabilization was done with 0.5% (v/v) Triton X-100 for 3 min at RT and blocking with 1% (w/v) BSA in PBS for 30 min at RT. The primary antibody was incubated with shaking for 1 h at RT, coverslips were washed with PBS and incubated with shaking with the secondary antibody for 1 h at RT. Slides were washed with PBS, stained with DAPI, washed again several times with PBS and mounted. Primary and secondary antibodies are listed in **Chapter 4-Table supplement 13.**

Microscopy and image processing

Confocal imaging with live or fixed cells was performed on a confocal laser scanning microscope FV1000 IX81 inverted microscope (Olympus, Shinjuku, Japan) using a 60x water immersion UPLSAPO NA 1.2 objective. DAPI, eGFP and Star Red were excited at 405 nm, 488 nm and 635 nm, respectively, with the internal FV10-MARAD-2 main laser unit. Star Orange was excited at 559 nm with an external Opti λ 559 diode laser (NTT Electronics, Yokohama, Japan). Internal PMT detectors (Olympus) were used for detection. Confocal laser scanning microscope (LSM) processing routine was carried out with the freely accessible Fiji and Huygens Pro 20.10 (SVI, Hilversum, Netherlands) for deconvolution. For deconvolution of images fulfilling the Nyquist criterion, we used an automatically computed theoretical point spread function based on our known microscopic parameters and a model of the Olympus IX81 provided by SVI and performed 30 iterative steps of classic maximum likelihood estimation (CMLE) on our images. Structured illumination microscopy (SIM) was performed on an ELYRA PS.1 (Zeiss) using a Plan-Apochromat 63x/1.4 Oil DIC M27 objective. For imaging of the fluorescent proteins eGFP and mCherry fused to mGBPs in fixed MEF cells, 488 nm was chosen with a BP495-575+LP750 filter for eGFP and 561 nm excitation with a BP570-650+LP750 filter for mCherry, respectively. The resulting images were 75.56 µm x 75.40 µm. During SIM a grid size of 34 µm for eGFP and 42 µm for mCherry was used. Grids were 5 times rotated. Reconstruction was performed using the ZEN structured illumination feature with noise filter of -6.5 for eGFP and -6.2 for mCherry.

Multiparameter imaging spectroscopy

The FV1000 system is additionally equipped with an external time-correlated single photon counting unit (Hydra Harp 400, PicoQuant) and external detectors for MFIS-FRET measurements. eGFP was excited at 485 nm with a LDH-D-C-485 diode laser (PicoQuant) with a linearly polarized beam and a pulse frequency of 32 MHz. mCherry was excited with an Opti λ 559 (NTT Electronics) laser at a continuous wave (CW) setting. MFIS-FRET measurements were done with laser powers of ~ 200 nW at 485 nm and ~500 nW at 559 nm, a pinhole size of 200 µm, a dwell time of 20 µs at a total of 60 frames. Separation into parallel and perpendicular signals was realized by a PBS 101 polarizing beamsplitter cube (Thorlabs). eGFP signal was detected via PD5CTC single-photon avalanche photodiodes (Micro Photon Devices) with upstream HC 520/35 bandpass filters (AHF). Detection of mCherry signal was realized via cooled HMPC-100-40 hybrid-photodetectors (Becker&Hickl) with upstream HC 607/70 bandpass filters (AHF).

In the first step of pixel-wise analysis, pixels were selectively averaged by grouping with the same characteristics. In a second step, the grouped pixels were integrated for subsequent sub-ensemble analysis. To determine fluorescence-weighted lifetimes in a pixel-wise analysis, the histograms presenting the decay of fluorescence intensity after the excitation pulse were built for each pixel with 128 ps per bin. Raw data was prepared using the software "*AnI*-3SF" developed in *Seidel* group, as part of the software package for multiparameter TCSPC/imaging. Fluorescence decay analysis was done using "ChiSurf", a global analysis platform for fluorescence data developed as open source software https://github.com/fluorescence-tools. We used the fit formula *Eq. 2* for Donly-analysis:

$$f_{D|D}^{DO}(t) = \sum_i x_D^i \exp\left(-\frac{t}{\tau_D^i}\right) \qquad\qquad Eq.\ 2$$

with *DO* representing the donor in absence of the acceptor, *D|D* representing donor emission upon donor excitation, *i* representing the number of exponents needed to fit the system, x_D representing the donor fraction of each exponential expression and τ_D representing the donor lifetime of each exponential expression.

Knowing that the donor will be quenched by FRET in addition to *Eq. 2*, we derive *Eq. 3*, describing the FRET-induced donor decay on top of the donor-only decay.

$$f_{D|D}^{DA}(t) = f_{D|D}^{DO}(t) \cdot \sum_i x_{FRET}^i \exp\left(-t \cdot k_{FRET}^i\right) \qquad\qquad Eq.\ 3$$

with *DA* representing the donor in presence of the acceptor, x_{FRET} representing the fraction of molecules undergoing FRET corresponding to each FRET-rate k_{FRET} (with *i* depending of the number of energy transfer modes described by different transfer rates).

Scatterplot generation and cellular back mapping

We used the freely available, open source Fiji plugin ScatterJ [88]. Processed images were converted into 8-bit grey scale tagged image file formats (.tiff) and opened in ScatterJ. 256x256 pixel scatterplots were saved as xy-lists (.dat) for processing in OriginPro. Within scatterplot, regions were selected using Fiji free-hand-tool and back mapped to the original image. The back-mapped image, as well as channel-wise images, were saved as text sequences (.dat) or portable networks graphics (.png) for further image and matrix analysis in OriginPro. The resulting cellular back-mapping-images represent the pixel-wise analysis of defined scatterplot populations.

6. List of abbreviations

4-PL:	Four parameter logistic
aa:	Amino acid
AA:	Activity assay
ASHXL:	Absent, small, or homeotic-like
BSA:	Bovine serum albumin
CBP:	CREB binding protein
CHAPS:	(3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate)
CMLE:	Classic maximum likelihood estimation
COMPASS:	Complex proteins associated with Set1
DAPI:	4',6-diamidino-2-phenylindole
DDR:	DNA repair and stress response
DMEM:	Dulbecco's modified Eagle's medium
ECL:	Enhanced chemiluminescence
EDTA:	Ethylene diaminetetraacetate
eGFP:	Enhanced green fluorescence protein
EGTA:	Ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N',-tetraacetate
ELISA:	Enzyme-linked immunosorbent assay
FA:	Formaldehyde
FACS:	Fluorescence activated cell sorting
FBS:	Fetal bovine serum
GO:	Gene ontology

ICC:	Immunocytochemistry
IDR:	Intrinsically disordered protein
IP:	Immunoprecipitation
KDM:	Lysine demethylase
KMT:	Lysine methyltransferase
LSM:	Laser scanning microscope
M:	Manders coefficient
MS:	Mass spectrometry
NPM:	Nucleophosmin
PCC:	Pearson coefficient
PAGE:	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
PBS:	Phosphate buffered saline
PCR:	Polymerase chain reaction
pNPP:	para-Nitrophenylphosphat
PTM:	Post-translational modification
RBBP:	Retinoblastoma-binding protein
RE:	Restriction enzyme
RLB:	RIPA-like buffer
RT:	Room temperature
SDM:	Site-directed mutagenesis
SDS:	Sodium dodecyl sulfate
STR:	Short tandem repeats
SWI/SNF:	SWitch/Sucrose non-fermentable
TBS-T:	Tris-buffered saline/tween 20
TCEP:	Tris(2-carboxyethyl)phosphine
TPR:	Tetratricopeptide repeat
TRIS:	Tris(hydroxymethyl)aminomethane
UTX:	Ubiquitously transcribed tetratricopeptide repeat, X chromosome
WDR:	WD-repeat containing protein
WRAD:	WDR5-RBBP5-ASH2L-DPY30 complex
WT:	Wildtype

7. Chapter 4 References

1.	Bertoli, C., Skotheim, J. M., de Bruin, R. A. M. Control of cell cycle transcription during G1 and S phases. *Nature Reviews Molecular Cell Biology*, 2013, **14**, 518-528. DOI: 10.1038/nrm3629.

2.	Passegue, E., Wagers, A. J., Giuriato, S., Anderson, W. C., Weissman, I. L. Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. *Journal of Experimental Medicine*, 2005, **202**, 1599-1611. DOI: 10.1084/jem.20050967.

3.	Sonawane, A. R., Platig, J., Fagny, M., Chen, C. Y., Paulson, J. N., Lopes-Ramos, C. M., et al. Understanding tissue-specific gene regulation. *Cell Reports*, 2017, **21**, 1077-1088. DOI: 10.1016/j.celrep.2017.10.001.

4.	Jiang, P., Freedman, M. L., Liu, J. S., Liu, X. S. Inference of transcriptional regulation in cancers. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, **112**, 7731-7736. DOI: 10.1073/pnas.1424272112.

5.	Bradner, J. E., Hnisz, D., Young, R. A. Transcriptional addiction in cancer. *Cell*, 2017, **168**, 629-643. DOI: 10.1016/j.cell.2016.12.013.

6.	Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 21931-21936. DOI: 10.1073/pnas.1016071107.

7.	Wiles, E. T., Selker, E. U. H3K27 methylation: a promiscuous repressive chromatin mark. *Current Opinion in Genetics & Development*, 2017, **43**, 31-37. DOI: 10.1016/j.gde.2016.11.001.

8.	Bannister, A. J., Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Research*, 2011, **21**, 381-395. DOI: 10.1038/cr.2011.22.

9.	Hong, S. H., Cho, Y. W., Yu, L. R., Yu, H., Veenstra, T. D., Ge, K. Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, **104**, 18439-18444. DOI: 10.1073/pnas.0707292104.

10.	Lee, M. G., Villa, R., Trojer, P., Norman, J., Yan, K. P., Reinberg, D., et al. Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science*, 2007, **318**, 447-450. DOI: 10.1126/science.1149042.

11.	Agger, K., Cloos, P. A. C., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., et al. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature*, 2007, **449**, 731-U10. DOI: 10.1038/nature06145.

12. Miller, S. A., Mohn, S. E., Weinmann, A. S. Jmjd3 and UTX play a demethylase-independent role in chromatin remodeling to regulate T-box family member-dependent gene expression. *Molecular Cell*, 2010, **40**, 594-605. DOI: 10.1016/j.molcel.2010.10.028.

13. Vandamme, J., Lettier, G., Sidoli, S., Di Schiavi, E., Jensen, O. N., Salcini, A. E. The C. elegans H3K27 demethylase UTX-1 Is essential for normal development, independent of its enzymatic activity. *PLOS Genetics*, 2012, **8**. DOI: 10.1371/journal.pgen.1002647.

14. Shpargel, K. B., Sengoku, T., Yokoyama, S., Magnuson, T. UTX and UTY demonstrate histone demethylase-independent function in mouse embryonic development. *PLOS Genetics*, 2012, **8**, 17. DOI: 10.1371/journal.pgen.1002964.

15. Sengoku, T., Yokoyama, S. Structural basis for histone H3 Lys 27 demethylation by UTX/KDM6A. *Genes & Development*, 2011, **25**, 2266-2277. DOI: 10.1101/gad.172296.111.

16. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 2021, **596**, 583-+. DOI: 10.1038/s41586-021-03819-2.

17. Del Rizzo, P. A., Trievel, R. C. Molecular basis for substrate recognition by lysine methyltransferases and demethylases. *Biochimica Et Biophysica Acta: Gene Regulatory Mechanisms*, 2014, **1839**, 1404-1415. DOI: 10.1016/j.bbagrm.2014.06.008.

18. Dhar, S. S., Lee, S. H., Chen, K. F., Zhu, G. J., Oh, W., Allton, K., et al. An essential role for UTX in resolution and activation of bivalent promoters. *Nucleic Acids Research*, 2016, **44**, 3659-3674. DOI: 10.1093/nar/gkv1516.

19. Tie, F., Banerjee, R., Conrad, P. A., Scacheri, P. C., Harte, P. J. Histone demethylase UTX and chromatin remodeler BRM bind directly to CBP and modulate acetylation of histone H3 lysine 27. *Molecular and Cellular Biology*, 2012, **32**, 2323-2334. DOI: 10.1128/mcb.06392-11.

20. Zha, L., Li, F. F., Wu, R., Artinian, L., Rehder, V., Yu, L. Q., et al. The histone demethylase UTX promotes brown adipocyte thermogenic program via coordinated regulation of H3K27 demethylation and acetylation. *Journal of Biological Chemistry*, 2015, **290**, 25151-25163. DOI: 10.1074/jbc.M115.662650.

21. Wang, S. P., Tang, Z. Y., Chen, C. W., Shimada, M., Koche, R. P., Wang, L. H., et al. A UTX-MLL4-p300 transcriptional regulatory network coordinately shapes active enhancer landscapes for eliciting transcription. *Molecular Cell*, 2017, **67**, 308-+. DOI: 10.1016/j.molcel.2017.06.028.

22. Rickels, R., Wang, L., Iwanaszko, M., Ozark, P. A., Morgan, M. A., Piunti, A., et al. A small UTX stabilization domain of Trr is conserved within mammalian MLL3-4/COMPASS and is sufficient to rescue loss of viability in null animals. *Genes & Development*, 2020, **34**, 1493-1502. DOI: 10.1101/gad.339762.120.

23. Kato, H., Asamitsu, K., Sun, W. D., Kitajima, S., Yoshizawa-Sugata, N., Okamoto, T., et al. Cancer-derived UTX TPR mutations G137V and D336G impair interaction with MLL3/4 complexes and affect UTX subcellular localization. *Oncogene*. DOI: 10.1038/s41388-020-1218-3.

24. D'Andrea, L. D., Regan, L. TPR proteins: the versatile helix. *Trends in Biochemical Sciences*, 2003, **28**, 655-662. DOI: 10.1016/j.tibs.2003.10.007.

25. Perez-Riba, A., Itzhaki, L. S. The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition. *Current Opinion in Structural Biology*, 2019, **54**, 43-49. DOI: 10.1016/j.sbi.2018.12.004.

26. Kajander, T., Cortajarena, A. L., Mochrie, S., Regan, L. Structure and stability of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR proteins. *Acta Crystallographica: Biological Crystallography*, 2007, **63**, 800-811. DOI: 10.1107/s0907444907024353.

27. Hurst, C. D., Alder, O., Platt, F. M., Droop, A., Stead, L. F., Burns, J. E., et al. Genomic subtypes of non-invasive bladder cancer with distinct metabolic profile and female gender bias in KDM6A mutation frequency. *Cancer Cell*, 2017, **32**, 701-+. DOI: 10.1016/j.ccell.2017.08.005.

28. Hofstetter, C., Kampka, J. M., Huppertz, S., Weber, H., Schlosser, A., Müller, A. M., et al. Inhibition of KDM6 activity during murine ESC differentiation induces DNA damage. *Journal of Cell Science*, 2016, **129**, 788-803. DOI: 10.1242/jcs.175174.

29. Smith, E. R., Lee, M. G., Winter, B., Droz, N. M., Eissenberg, J. C., Shiekhattar, R., et al. Drosophila UTX is a histone H3 Lys27 demethylase that colocalizes with the elongating form of RNA polymerase II. *Molecular and Cellular Biology*, 2008, **28**, 1041-1046. DOI: 10.1128/mcb.01504-07.

30. Chen, H., Huang, Y. X., Zhu, X. Q., Liu, C., Yuan, Y. M., Su, H., et al. Histone demethylase UTX is a therapeutic target for diabetic kidney disease. *Journal of Physiology*, 2019, **597**, 1643-1660. DOI: 10.1113/jp277367.

31. Chang, S., Yim, S., Park, H. The cancer driver genes IDH1/2, JARID1C/KDM5C, and UTX/ KDM6A: crosstalk between histone demethylation and hypoxic reprogramming in cancer metabolism. *Experimental and Molecular Medicine*, 2019, **51**. DOI: 10.1038/s12276-019-0230-6.

32. Ler, L. D., Ghosh, S., Chai, X. R., Thike, A. A., Heng, H. L., Siew, E. Y., et al. Loss of tumor suppressor KDM6A amplifies PRC2-regulated transcriptional repression in

bladder cancer and can be targeted through inhibition of EZH2. *Science Translational Medicine*, 2017, **9**. DOI: 10.1126/scitranslmed.aai8312.

33. Leng, X., Wang, J., An, N., Wang, X., Sun, Y., Chen, Z. Histone 3 lysine-27 demethylase KDM6A coordinates with KMT2B to play an oncogenic role in NSCLC by regulating H3K4me3. *Oncogene*, 2020, **39**, 6468-6479. DOI: 10.1038/s41388-020-01449-y.

34. Kim, J. H., Sharma, A., Dhar, S. S., Lee, S. H., Gu, B., Chan, C. H., et al. UTX and MLL4 coordinately regulate transcriptional programs for cell proliferation and invasiveness in breast cancer cells. *Cancer Research*, 2014, **74**, 1705-1717. DOI: 10.1158/0008-5472.Can-13-1896.

35. Xie, G., Liu, X., Zhang, Y., Li, W., Liu, S., Chen, Z., et al. UTX promotes hormonally responsive breast carcinogenesis through feed-forward transcription regulation with estrogen receptor. *Oncogene*, 2017, **36**, 5497-5511. DOI: 10.1038/onc.2017.157.

36. Schulz, W. A., Lang, A., Koch, J., Greife, A. The histone demethylase UTX/KDM6A in cancer: progress and puzzles. *International Journal of Cancer*, 2019, **145**, 614-620. DOI: 10.1002/ijc.32116.

37. Bubenik, J., Perlmann, P., Helmstein, K., Moberger, G. Cellular and humoral immune responses to human urinary bladder carcinoma. *International Journal of Cancer*, 1970, **5**, 310-+. DOI: 10.1002/ijc.2910050303.

38. Makri, D., Schulz, W. A., Grimm, M. O., Schmitzdrager, B. J. Rapid killing of urothelial carcinoma-cells by wild-type p53. *International Journal of Oncology*, 1995, **7**, 637-641. DOI: 10.3892/ijo.7.3.637.

39. Steube, K. G., Meyer, C., Drexler, H. G. Secretion of functional hematopoietic growth factors by human carcinoma cell lines. *International Journal of Cancer*, 1998, **78**, 120-124. DOI: 10.1002/(sici)1097-0215(19980925)78:1<120::Aid-ijc19>3.3.Co;2-j.

40. Grimm, M. O., Jürgens, B., Schulz, W. A., Decken, K., Makri, D., Schmitzdrager, B. J. Inactivation of tumor-suppressor genes and deregulation of the C-MYC gene in urothelial cancer cell-lines. *Urological Research*, 1995, **23**, 293-300. DOI: 10.1007/bf00300017.

41. Kyriazis, A. A., Kyriazis, A. P., McCombs, W. B., Peterson, W. D. Morphological, biological and biochemical characteristics of human bladder transitional cell carcinomas grown in tissue-culture and in nude-mice. *Cancer Research*, 1984, **44**, 3997-4005. DOI: PMID: 6744315.

42. Kau, T. R., Way, J. C., Silver, P. A. Nuclear transport and cancer: from mechanism to intervention. *Nature Reviews Cancer*, 2004, **4**, 106-117. DOI: 10.1038/nrc1274.

43. Ivanov, A. A., Khuri, F. R., Fu, H. A. Targeting protein-protein interactions as an anticancer strategy. *Trends in Pharmacological Sciences*, 2013, **34**, 393-400. DOI: 10.1016/j.tips.2013.04.007.

44. Kar, G., Gursoy, A., Keskin, O. Human cancer protein-protein interaction network: a structural perspective. *PLOS Computational Biology*, 2009, **5**. DOI: 10.1371/journal.pcbi.1000601.

45. Van der Meulen, J., Speleman, F., Van Vlierberghe, P. The H3K27me3 demethylase UTX in normal development and disease. *Epigenetics*, 2014, **9**, 658-668. DOI: 10.4161/epi.28298.

46. Dean, D. A., Dean, B. S., Müller, S., Smith, L. C. Sequence requirements for plasmid nuclear import. *Experimental Cell Research*, 1999, **253**, 713-722. DOI: 10.1006/excr.1999.4716.

47. Wang, X. H., Li, S. L. Protein mislocalization: mechanisms, functions and clinical applications in cancer. *Biochimica Et Biophysica Acta: Reviews on Cancer*, 2014, **1846**, 13-25. DOI: 10.1016/j.bbcan.2014.03.006.

48. Lang, A., Yilmaz, M., Hader, C., Murday, S., Kunz, X., Wagner, N., et al. Contingencies of UTX/KDM6A action in urothelial carcinoma. *Cancers*, 2019, **11**. DOI: 10.3390/cancers11040481.

49. Perner, F., Armstrong, S. A. Targeting chromatin complexes in myeloid malignancies and beyond: from basic mechanisms to clinical innovation. *Cells*, 2020, **9**. DOI: 10.3390/cells9122721.

50. Zhang, X., Wang, Y. L., Wang, J. Y., Sun, F. Y. Protein-protein interactions among signaling pathways may become new therapeutic targets in liver cancer. *Oncology Reports*, 2016, **35**, 625-638. DOI: 10.3892/or.2015.4464.

51. Boettcher, S., Miller, P. G., Sharma, R., McConkey, M., Leventhal, M., Krivtsov, A. V., et al. A dominant-negative effect drives selection of TP53 missense mutations in myeloid malignancies. *Science*, 2019, **365**, 599-+. DOI: 10.1126/science.aax3649.

52. Papa, A., Wan, L. X., Bonora, M., Salmena, L., Song, M. S., Hobbs, R. M., et al. Cancer-associated PTEN mutants act in a dominant-negative manner to suppress PTEN protein function. *Cell*, 2014, **157**, 595-610. DOI: 10.1016/j.cell.2014.03.027.

53. Raiymbek, G., An, S. J., Khurana, N., Gopinath, S., Larkin, A., Biswas, S., et al. An H3K9 methylation-dependent protein interaction regulates the non-enzymatic functions of a putative histone demethylase. *Elife*, 2020, **9**. DOI: 10.7554/elife.53155.

54. Morgan, M. A. J., Shilatifard, A. Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation. *Nature Genetics*, 2020, **52**, 1271-1281. DOI: 10.1038/s41588-020-00736-4.

55. Zha, L., Cao, Q., Cui, X., Li, F. F., Liang, H. J., Xue, B. Z., et al. Epigenetic regulation of E-cadherin expression by the histone demethylase UTX in colon cancer cells. *Medical Oncology*, 2016, **33**, 11. DOI: 10.1007/s12032-016-0734-z.

56. Watanabe, S., Shimada, S., Akiyama, Y., Ishikawa, Y., Ogura, T., Ogawa, K., et al. Loss of KDM6A characterizes a poor prognostic subtype of human pancreatic cancer and potentiates HDAC inhibitor lethality. *International Journal of Cancer*, 2019, **145**, 192-205. DOI: 10.1002/ijc.32072.

57. Li, X. M., Li, X. D. Interrogating interactions and modifications of histones in live cells. *Cell Chemical Biology*, 2018, **25**, 1-3. DOI: 10.1016/j.chembiol.2018.01.003.

58. Kim, D. H., Tang, Z. Y., Shimada, M., Fierz, B., Houck-Loomis, B., Bar-Dagen, M., et al. Histone H3K27 trimethylation inhibits H3 binding and function of SET1-like H3K4 methyltransferase complexes. *Molecular and Cellular Biology*, 2013, **33**, 4936-4946. DOI: 10.1128/mcb.00601-13.

59. Smith, E., Lin, C. Q., Shilatifard, A. The super elongation complex (SEC) and MLL in development and disease. *Genes & Development*, 2011, **25**, 661-672. DOI: 10.1101/gad.2015411.

60. Hebditch, M., Carballo-Amador, M. A., Charonis, S., Curtis, R., Warwicker, J. Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics*, 2017, **33**, 3098-3100. DOI: 10.1093/bioinformatics/btx345.

61. Drozdetskiy, A., Cole, C., Procter, J., Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Research*, 2015, **43**, W389-W394. DOI: 10.1093/nar/gkv332.

62. Parthiban, V., Gromiha, M. M., Schomburg, D. CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Research*, 2006, **34**, W239-W242. DOI: 10.1093/nar/gkl190.

63. Mittal, A., Hobor, F., Zhang, Y., Martin, S. R., Gamblin, S. J., Ramos, A., et al. The structure of the RbBP5 beta-propeller domain reveals a surface with potential nucleic acid binding sites. *Nucleic Acids Research*, 2018, **46**, 3802-3812. DOI: 10.1093/nar/gky199.

64. Zhang, P., Lee, H., Brunzelle, J. S., Couture, J. F. The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. *Nucleic Acids Research*, 2012, **40**, 4237-4246. DOI: 10.1093/nar/gkr1235.

65. Cherstvy, A. G. Positively charged residues in DNA-binding domains of structural proteins follow sequence-specific positions of DNA phosphate groups. *Journal of Physical Chemistry B*, 2009, **113**, 4242-4247. DOI: 10.1021/jp810009s.

66. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. STRING v11: protein-protein association networks with increased coverage,

supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 2019, **47**, D607-D613. DOI: 10.1093/nar/gky1131.

67. Niegisch, G., Knievel, J., Koch, A., Hader, C., Fischer, U., Albers, P., et al. Changes in histone deacetylase (HDAC) expression patterns and activity of HDAC inhibitors in urothelial cancers. *Urologic Oncology-Seminars and Original Investigations*, 2013, **31**, 1770-1779. DOI: 10.1016/j.urolonc.2012.06.015.

68. Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., et al. The reactome pathway knowledgebase. *Nucleic Acids Research*, 2018, **46**, D649-D655. DOI: 10.1093/nar/gkx1132.

69. Pejaver, V., Hsu, W. L., Xin, F. X., Dunker, A. K., Uversky, V. N., Radivojac, P. The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Science*, 2014, **23**, 1077-1093. DOI: 10.1002/pro.2494.

70. Pätzold, R., Characterization of functional variants of the histone-modifying enzyme KDM6A/UTX in living cells using multiparameter fluorescence imaging spectroscopy, *Molecular Physical Chemistry*. 2018, Heinrich-Heine-Universität: Düsseldorf.

71. Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., et al. Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *Journal of Biological Chemistry*, 2009, **284**, 478-485. DOI: 10.1074/jbc.M807017200.

72. Wang, R. W., Brattain, M. G. The maximal size of protein to diffuse through the nuclear pore is larger than 60 kDa. *FEBS Letters*, 2007, **581**, 3164-3170. DOI: 10.1016/j.febslet.2007.05.082.

73. Bohly, N., Kistner, M., Bastians, H. Mild replication stress causes aneuploidy by deregulating microtubule dynamics in mitosis. *Cell Cycle*, 2019, **18**, 2770-2783. DOI: 10.1080/15384101.2019.1658477.

74. Salmina, K., Bojko, A., Inashkina, I., Staniak, K., Dudkowska, M., Podlesniy, P., et al. "Mitotic slippage" and extranuclear DNA in cancer chemoresistance: a focus on telomeres. *International Journal of Molecular Sciences*, 2020, **21**. DOI: 10.3390/ijms21082779.

75. Baba, T., Yoshida, T., Tanabe, Y., Nishimura, T., Morishita, S., Gotoh, N., et al. Cytoplasmic DNA accumulation preferentially triggers cell death of myeloid leukemia cells by interacting with intracellular DNA sensing pathway. *Cell Death Disease*, 2021, **12**, 322. DOI: 10.1038/s41419-021-03587-x.

76. Chakraborty, A. A., Laukka, T., Myllykoski, M., Ringel, A. E., Booker, M. A., Tolstorukov, M. Y., et al. Histone demethylase KDM6A directly senses oxygen to

control chromatin and cell fate. *Science*, 2019, **363**, 1217-+. DOI: 10.1126/science.aaw1026.

77. Cela, I., Di Matteo, A., Federici, L. Nucleophosmin in its interaction with ligands. *International Journal of Molecular Sciences*, 2020, **21**. DOI: 10.3390/ijms21144885.

78. Szebeni, A., Olson, M. O. J. Nucleolar protein B23 has molecular chaperone activities. *Protein Science*, 1999, **8**, 905-912. DOI: 10.1110/ps.8.4.905.

79. González-Quiroz, M., Blondel, A., Sagredo, A., Hetz, C., Chevet, E., Pedeux, R. When ER proteostasis meets the DNA damage response. *Trends in Cellular Biology*, 2020, **30**, 881-891. DOI: 10.1016/j.tcb.2020.09.002.

80. Dimitrova, E., Turberfield, A. H., Klose, R. J. Histone demethylases in chromatin biology and beyond. *EMBO Reports*, 2015, **16**, 1620-1639. DOI: 10.15252/embr.201541113.

81. Somma, M. P., Andreyeva, E. N., Pavlova, G. A., Pellacani, C., Bucciarelli, E., Popova, J. V., et al. Moonlighting in mitosis: analysis of the mitotic functions of transcription and splicing factors. *Cells*, 2020, **9**. DOI: 10.3390/cells9061554.

82. Ali, A., Veeranki, S. N., Chinchole, A., Tyagi, S. MLL/WDR5 complex regulates Kif2A localization to ensure chromosome congression and proper spindle assembly during mitosis. *Developmental Cell*, 2017, **41**, 605-+. DOI: 10.1016/j.devcel.2017.05.023.

83. Bryan, A. F., Wang, J., Howard, G. C., Guarnaccia, A. D., Woodley, C. M., Aho, E. R., et al. WDR5 is a conserved regulator of protein synthesis gene expression. *Nucleic Acids Research*, 2020, **48**, 2924-2941. DOI: 10.1093/nar/gkaa051.

84. Amin, M. A., Matsunaga, S., Uchiyama, S., Fukui, K. Nucleophosmin is required for chromosome congression, proper mitotic spindle formation, and kinetochore-microtubule attachment in HeLa cells. *FEBS Letters*, 2008, **582**, 3839-3844. DOI: 10.1016/j.febslet.2008.10.023.

85. Williams, R. D. Human urologic cancer cell-lines. *Investigative Urology*, 1980, **17**, 359-363. DOI: PMID: 6244232.

86. Masters, J. R. W., Hepburn, P. J., Walker, L., Highman, W. J., Trejdosiewicz, L. K., Povey, S., et al. Tissue-culture model of transitional cell-carcinoma - characterization of 22 human urothelial cell-lines. *Cancer Research*, 1986, **46**, 3630-3636. DOI: PMID: 3708594.

87. Marshall, C. J., Franks, L. M., Carbonell, A. W. Markers of neoplastic transformation in epithelial-cell lines derived from human carcinomas. *Journal of the National Cancer Institute*, 1977, **58**, 1743-1751. DOI: 10.1093/jnci/58.6.1743.

88. Zeitvogel, F., Schmid, G., Hao, L., Ingino, P., Obst, M. ScatterJ: An ImageJ plugin for the evaluation of analytical microscopy datasets. *Journal of Microscopy*, 2016, **261**, 148-156. DOI: 10.1111/jmi.12187.
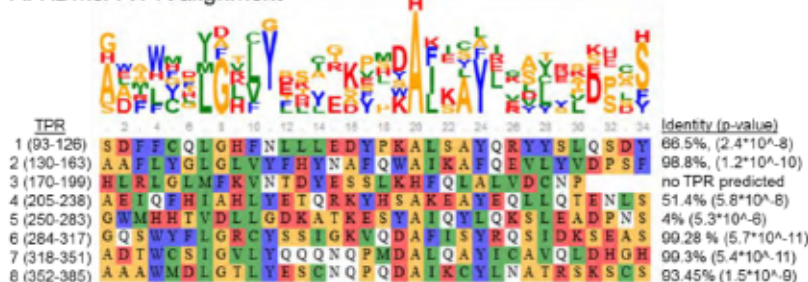
Chapter 4 Supplement

*(Text and figures taken or adapted from manuscript 1. (see **Publications** section, page 187)*

DATAFINDER:

A) For a detailed datafinder, refer to P:\SFF Projekt UTX\Methoden Ergebnisse\ List of experiment IDs (KO, KOB and KOC)

## Chapter 4-Figure supplement 1

A. KDM6A TPR alignment

TPR	sequence	Identity (p-value)
1 (93-126)	SDFFCQLGHFNLLLEDYPKALSAYQRYYSLQSDY	66.5%, (2.4*10^-8)
2 (130-163)	AAFLYGLGLVYEHYNAFQWAIKAFQEVLYVDPSF	98.8%, (1.2*10^-10)
3 (170-199)	HLRLGLMFKVNTDYESSLKHFQLALVDCNP	no TPR predicted
4 (205-238)	AEIQFHIAHLYETQRKYHSAKEAYEQLLQTENLS	51.4% (5.8*10^-8)
5 (250-283)	GWMHHTVDLLGDKATKESYAIQYLQKSLEADPNS	4% (5.3*10^-6)
6 (284-317)	GQSWYFLGRCYSSIGKVQDAFISYRQSIDKSEAS	90.28 % (5.7*10^-11)
7 (318-351)	ADTWCSIGVLYQQQNQPMDALQAYICAVQLDHGH	99.3% (5.4*10^-11)
8 (352-385)	AAAWMDLGTLYESCNQPQDAIKCYLNATRSKSCS	93.45% (1.5*10^-9)

B. TPR 3 (170-199) secondary structure prediction

AA_QUERY	1	HLRLGLMFKVNTDYESSLKHFQLALVDCNP	30
SS_PSIPRED		HHHHHH HHHHHHHHHHHH	
SS_SPIDER3		HHHHHHHHHH HHHHHHHHHHHH	
SS_PSSPRED4		EEEEEE HHHHHHHHHHHHHE	
SS_DEEPCNF		EEEEEEEE HHH EEEEEEEE	
SS_NETSURFP		HHHHHHHHH HHHHHHHHHHHHHH	

SS = α-helix β-strand

Bioinformatic analysis of KDM6A tetratricopeptide repeats. (Adapted from Koch et al., 2021) **A.** Alignment of annotated KDM6A TPRs using the *TPRprediction and alignment tool* published in [1]. All annotated KDM6A TPRs with the given amino acid range (see label) were compared for their TPR conservation to the consensus sequence as published in [4] with the most conserved amino acids W4, L7, G8, Y11, A20, Y24, A27 and P30. Identity and p-values are shown on the right hand. The 34 amino acid TPR motif could not be identified for TPR 3 (aa 170-199). Color coding according to CLUSTAL W alignment. **B**. secondary structure prediction of TPR3 using the Quick2D on the MPI Bioinformatics toolkit server [2, 3] indicated a high probability for aa 14-26 to form an α-helix, but aa 1-10 tend to form either an α-helix or a β-strand. Of note, the canonical 34 aa TPR repeating unit forms the crucial helix-turn-helix motif [4]. In general, the N-terminal TPRs 1-4 show less strict conservation towards the canonical TPR sequence compared to the C-terminal TPRs 6-8.

1.   Karpenahalli, M.R., A.N. Lupas, and J. Soeding, *TPRpred: a tool for prediction of TPR-, PPR- and SELI-like repeats from protein sequences.* Bmc Bioinformatics, 2007. **8**.
2.   McGuffin, L.J., K. Bryson, and D.T. Jones, *The PSIPRED protein structure prediction server.* Bioinformatics, 2000. **16**(4): p. 404-405.
3.   Zimmermann, L., et al., *A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core.* Journal of Molecular Biology, 2018. **430**(15): p. 2237-2243.
4.   Main, E.R.G., et al., *Design of stable alpha-helical arrays from an idealized TPR motif.* Structure, 2003. **11**(5): p. 497-508.

Chapter 4-Figure supplement 2
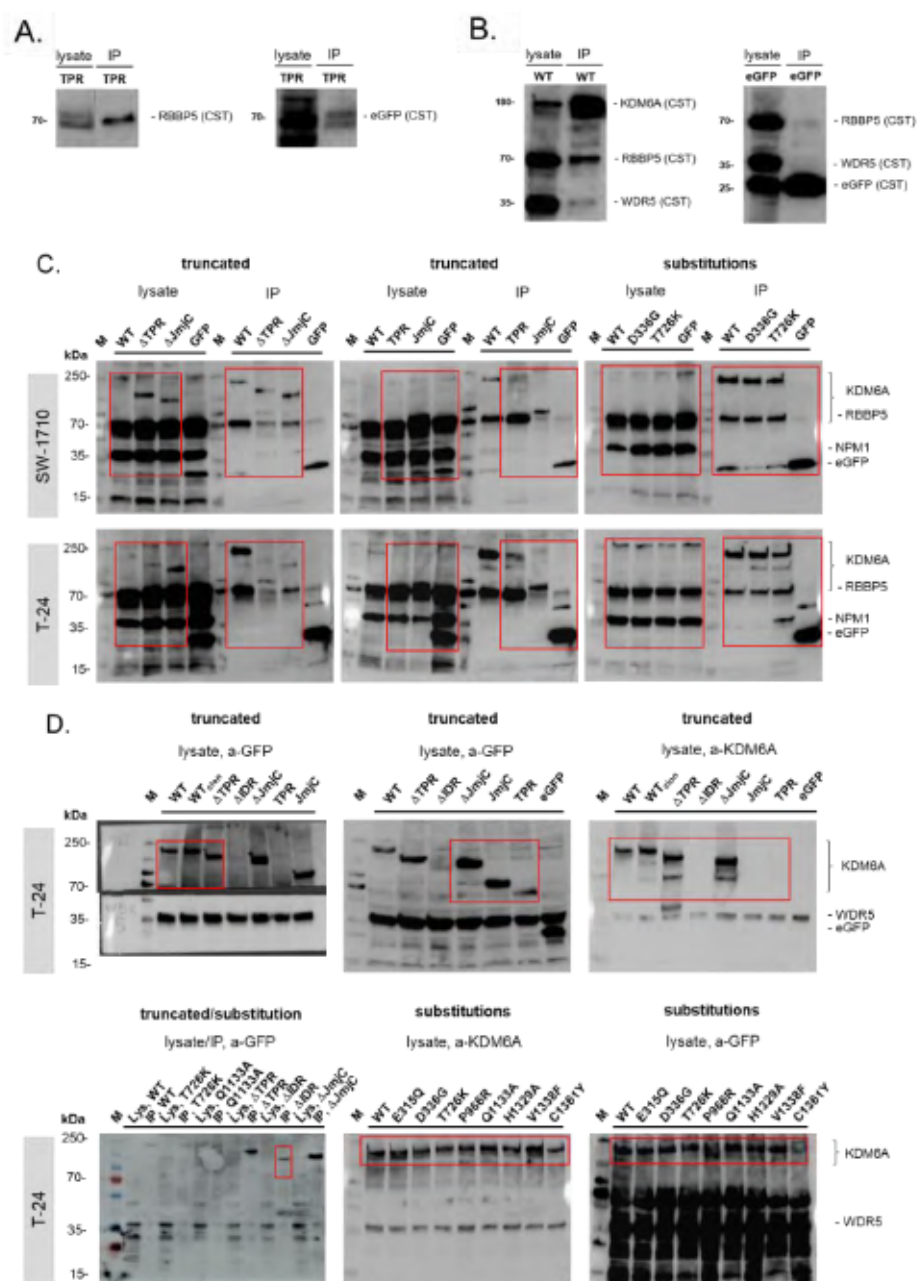
Western Blot control experiments and raw blots. A.
eGFP-KDM6A TPR whole cell lysate and the corresponding IP fraction was separated on a Western Blot and stained with eGFP (CST) or RBBP5 (CST) antibodies. RBBP5 precipitated with eGFP-KDM6A TPR. **B**. Dynabeads washing control on same membrane, washed with 3x RIPA like buffer (RLB). While KDM6A WT pulls a clearly visible RBBP5 band and a low WDR5 band, both are undetectable in eGFP-transfected cells alone (negative control). **C**. Raw blots from Figure 8. Lysate and Co-IP for selected KDM6A variants and eGFP. IP and corresponding lysate lanes were also on the same blot. All experimental conditions for the blots presented were kept constant. **D**. Raw blots from Figure 3C/D. Protein lanes shown in main figures are in red boxes. The PageRuler Plus Prestained protein ladder (Thermo Fisher) was used for orientation.
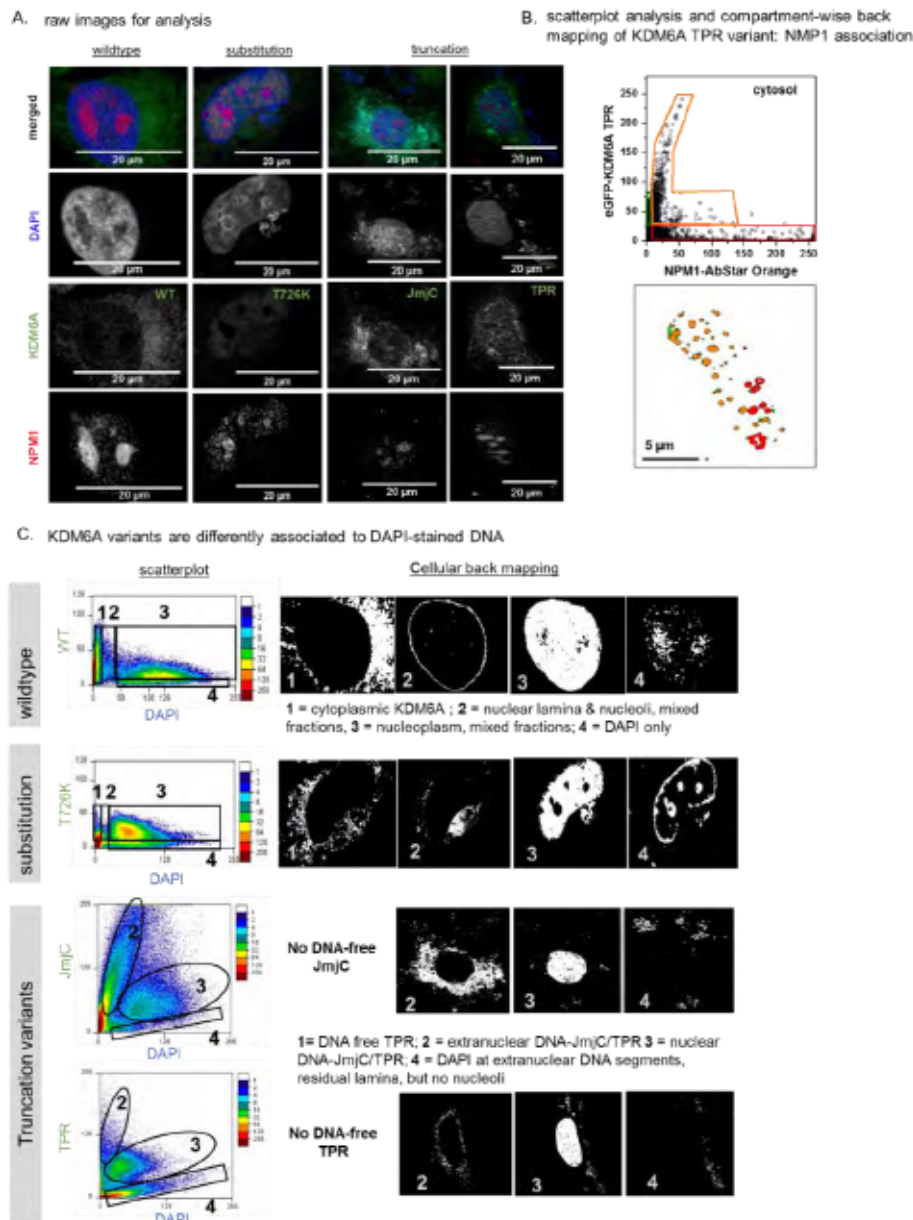
Cellular back mapping identifies co-localizing protein populations at extranuclear DNA segments. **A.** images shown in **Figure 4**, all channels. **B.** Analysis of KDM6A TPR and NMP1: Cellular back mapping identifies a colocalizing population (orange) and free KDM6A TPR (green) and NPM1 (red) at cytoplasmic DNA segments. **C.** scatterplot analysis and cellular back mapping of KDM6A variants and DAPI as indicator of KDM6A-DNA complexes. In KDM6A WT scatterplot we identified four populations: 1 = cytoplasmic KDM6A (without DAPI) and 4 = DAPI (without KDM6A), as well as, populations #2 and #3, which resemble different ratios of KDM6A-DNA complexes. Population 2 is located at the lamina and in nucleolar puncta, and population 3 is located within the nucleoplasm. Substitution variant T726K shows minor differences in populations#2 and #4, as the lamina in #2 is less clearly visible, but strongly enriched in population #4. Contour scatterplot of JmjC or TPR with DAPI identified only three populations, with population #1 being absent. Back mapping clearly indicates compartment-wise localization of different DAPI-KDM6A JmjC or TPR compositions.

Chapter 4-Figure supplement 4

A. elevated levels of p-γH2A.x
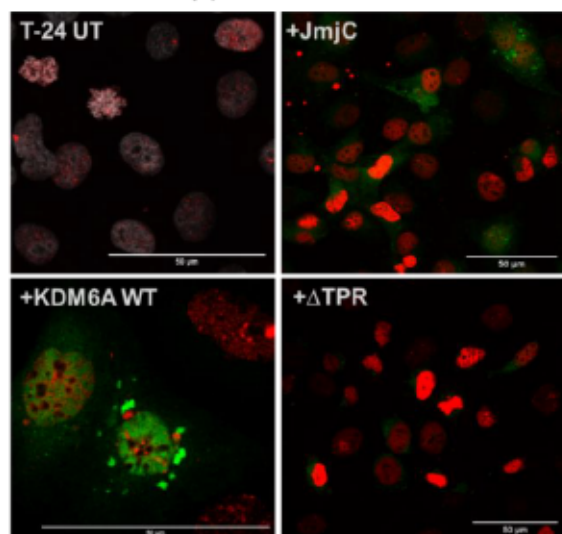
B. Endogenous distribution of KDM6A during mitosis
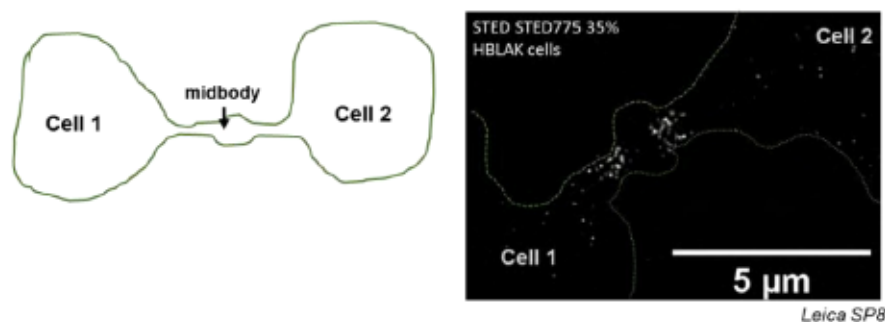
C. Localization of RAD51 after induction of KDM6AΔTPR
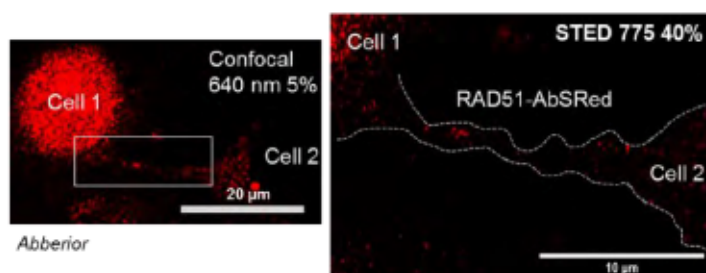
Distribution and expression levels of p-γH2A.x and RAD51 after KDM6A overexpression and endogenous localization of KDM6A during mitosis. A. p-γH2A.x protein level in untransfected and KDM6A variants transfected T-24 cells. Note, that even in strongly overexpressed KDM6A WT cells, p-γH2A.x was not increased, but strongly in KDM6A JmjC and ΔTPR cells. **B.** endogenous localization of KDM6A (CST antibody, 2nd antibody goat-anti-rabbit AbStar red) as imaged with Leica STED SP8 in HBLAK cells. For STED, the sample was depleted with 35% of 775 nm laser. **C.** RAD51 proteins were detected at a persisting chromatin bridge of two daughter cells, with cell 1 being KDM6A ΔTPR positive. The image was taken with Abberior STED, excitation laser 640 nm 5% and depletion laser 775 nm with 40%.
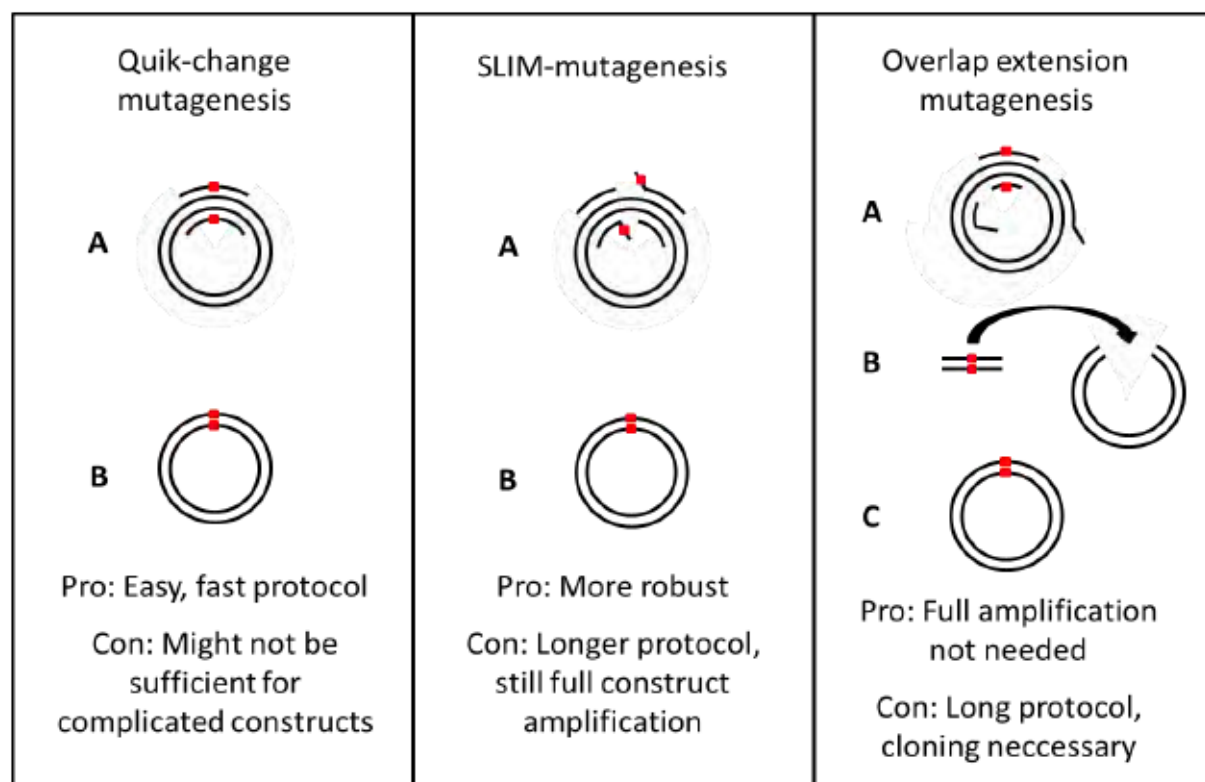
Schematic working principle of mutagenesis tested for generation of KDM6A substitution variants. Quik-change mutagenesis worked best after establishing PrimeSTAR GXL® polymerase and optimization of primer. SLIM- and OE-mutagenesis are further futurre options for more complex constructs.

ELISA-based H3K27me3 demethylation assay

The antibody-dependent assay can be used to test the activity of UTX, JMJD3 and UTY on a H3K27me3 peptide.
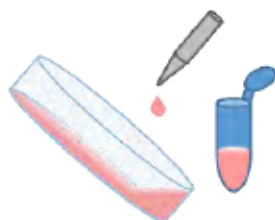
1. Harvest transfected cell lines

- Scratch cells down in medium and centrifuge at 1000 xg, 4 °C, 5 min, discard medium, keep pellet on ice.
- Resuspend and wash pellet in cold PBS, centrifuge at 1000 xg, 4 °C, 5 min, discard PBS, keep pellet on ice.

2. Lysis of cell pellets

- Resuspend pellet from 1. in 300 µl cold CO-IP buffer with freshly added protease inhibitor, shock-freeze vial in liquid nitrogen.
- Thaw on ice for 30 min, pipette up and down twice for 20 s after thawing.
- Centrifuge at 16000 xg, 4 °C, 10 min, discard pellet, keep supernatant on ice.

3. Measurement of eGFP-signal

- Take 150 µl of the supernatant from 2., measure fluorescence at 488 nm.
- After measurement, the sample will be put back into the respective supernatant.

4. GFP-Trap® MA beads binding

- Wash 10 µl slurried GFP-Trap® MA beads twice in CO-IP buffer, use magnet to separate beads from buffer.
- Pipette supernatant from 2. onto the beads.
- Rotate vial for 60 min at 4 °C, then use magnet to separate beads from buffer.
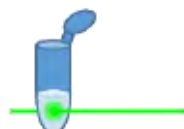
5. Measurement of eGFP-signal

- Take 150 µl of the supernatant from 4., measure fluorescence at 488 nm.
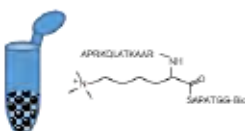

6. Incubation of beads with peptide

- Wash beads from 4. with 200 µl AA-buffer twice. Make sure no residual CO-IP buffer is left, then use magnet to separate beads from buffer.
- Resuspend beads in 150 µl AA buffer with 600 ng H3K27me3 peptide per vial.
- Shake in thermomixer for 2-4 h, 30 °C, 1000 rpm.
- Additionally, place a control-vial with 150 µl AA-buffer containing 714 ng H3K27me2 peptide in the thermomixer.
- Make sure beads stay dispersed throughout the incubation.
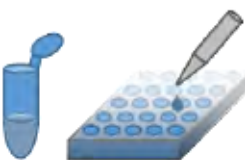
7. Binding peptide in 96-well plate

- Pipet 3x 50 µl (triplicates) of every vial from 6. into a well each.
- Cover with parafilm, gently rock for 60 min at RT.
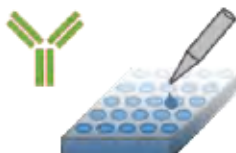
8. Primary antibody incubation

- Remove the solutions from every well.
- Add 50 µl primary antibody (1:500 in TBS-T) into each well, cover with parafilm, gently rock for 60 min at RT.
- Wash three times with TBS-T, gently rock for 5 min at RT in between while TBS-T is on sample.
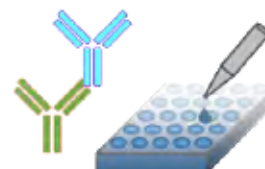
9. Secondary antibody incubation

- Remove the solutions from every well.
- Add 50 µl secondary antibody (1:1000 in TBS-T) into each well, cover with parafilm, gently rock for 30 min at RT.
- Wash four times with TBS-T, gently rock for 5 min at RT in between while TBS-T is on sample.
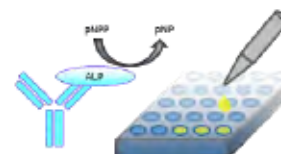
10. ALP-based development

- Remove the solutions from every well.
- Add 100 µl pNPP-solution to each well, leave for 10 min in the dark, mix by tapping every 2 min.
- End reaction by adding 100 µl 1 M NaOH to each well, mix by tapping.

11. Colorimetric readout

- A pale yellow stain will indicate for demethylated product.
- Readout plate at 405 nm with 650 nm as reference channel.

12. Data analysis

- Compare dimethylation levels.
- Calculate a correction factor for difference in protein amount by directly relating the fluorescence measurements before (3.) and after bead incubation (5.).

Chapter 4-Figure supplement 7

Plasmid map of the eGFP-TPR-SPOT construct in pEGFP-C1. TPR and extension were cloned into the leading part of the MCS, using it as an effective linker between SPOT, Poly-K (sol-tag) at the end of the MCS.

Chapter 4-Figure supplement 8
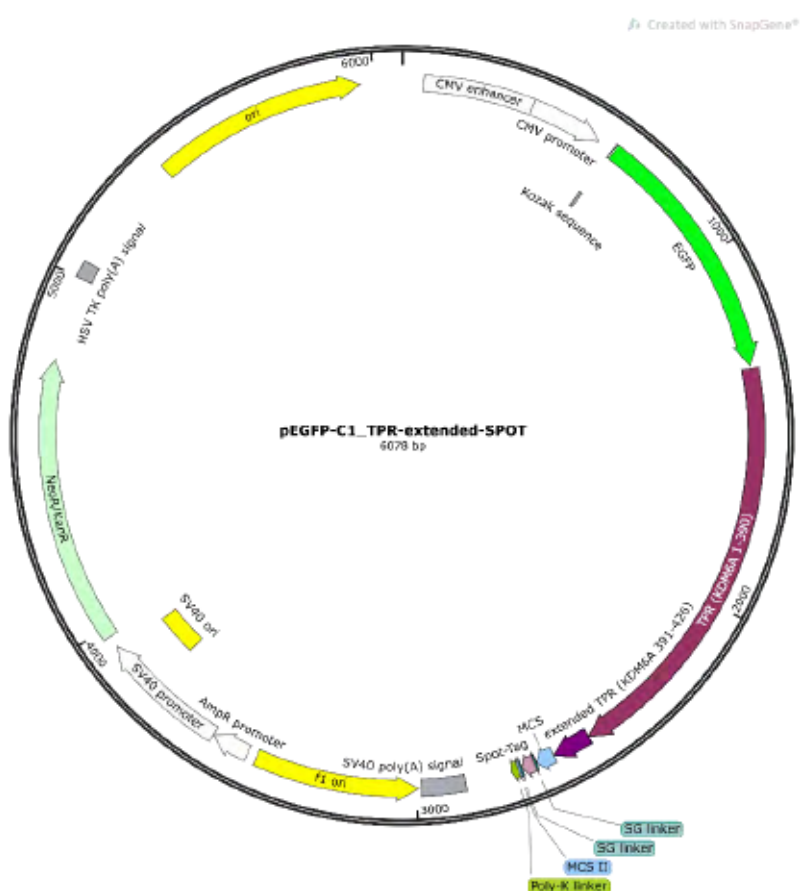
Plasmid map of the substitution mutations in in pcDNA3.1(+). Single substitutions were mutated at the respective sites. A primer network was establish to allow full sequencing of the construct.
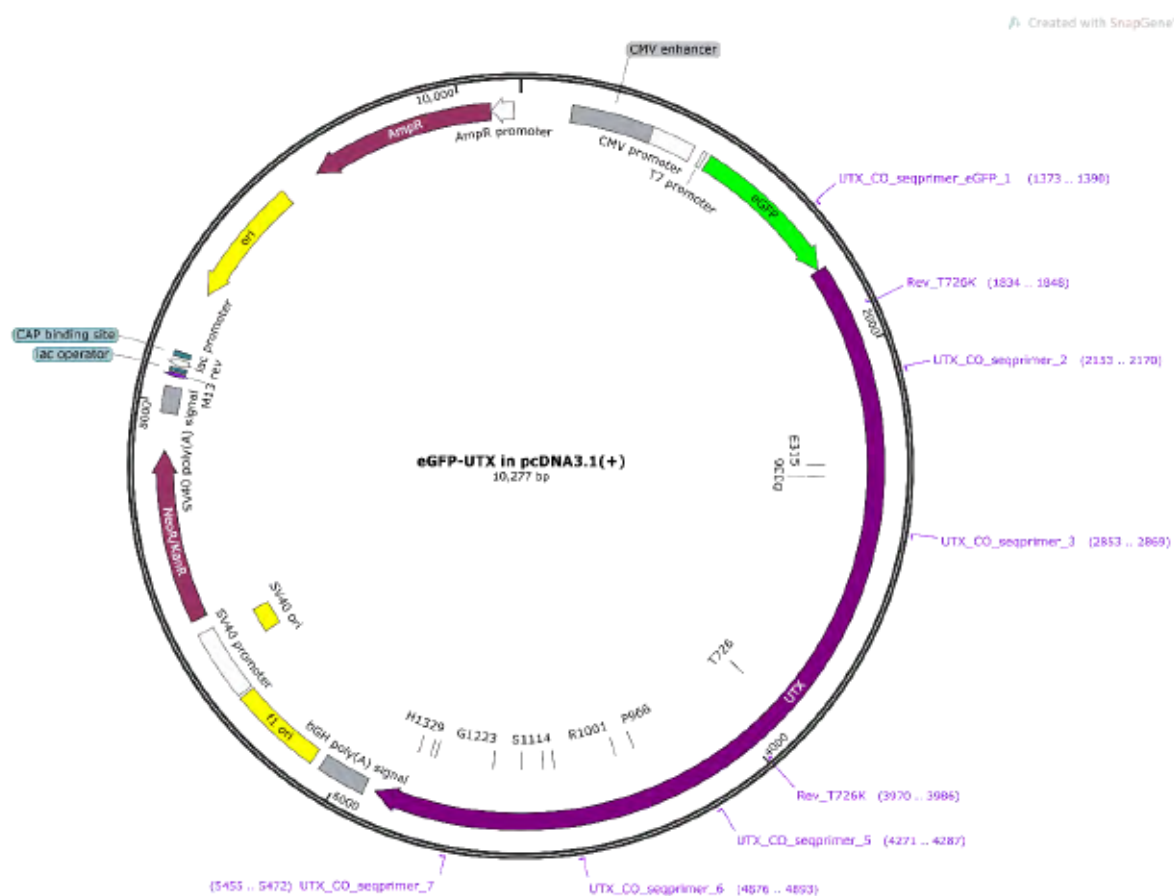
Chapter 4-Table supplement 1

Literature meta-assessment of KDM6A interaction partners. Grouped via main functionality or complex-affiliation. Interaction sources with more than two independent publications are indicated in bold font.

Compass-complex	
RBBP5	(Shpargel, Sengoku, Yokoyama, & Magnuson, 2012), (Yang et al., 2019), (Benyoucef et al., 2016), (Vandamme et al., 2012), (Shpargel, Starmer, Wang, Ge, & Magnuson, 2017), (Lang et al., 2019)
ASH2L	(Kato et al.), (Benyoucef et al., 2016), (Vandamme et al., 2012), (Rocha-Viegas et al., 2014)
KMT2D	(Vandamme et al., 2012), (Kim et al., 2014), (Rickels et al., 2020)
PTIP/PAXIP1	(Yang et al., 2019), (Benyoucef et al., 2016), (Vandamme et al., 2012)
WDR5	(Benyoucef et al., 2016), (Vandamme et al., 2012)
KMT2C	(Vandamme et al., 2012)
PA1	(Benyoucef et al., 2016)
KMT2B	(Leng et al., 2020)
Transcription, transcription factors and promotors	
CBP	(Tie, Banerjee, Conrad, Scacheri, & Harte, 2012), (Zha et al., 2015), (Zha et al., 2016), (Xie et al., 2017)
p300	(Watanabe et al., 2019), (Wang et al., 2017)
NCOA6	(Benyoucef et al., 2016), (Rocha-Viegas et al., 2014)
RNA-Pol II	(Smith et al., 2008)
TAL1	(Benyoucef et al., 2016)
GATA3	(Yu et al., 2019)
GATADB2	(Biswas et al., 2019)
HNF1A	(Kalisz et al., 2020)
B-Catenin	(Leng et al., 2020)
ORF50	(Rossetto & Pari, 2012)
JunB	(Beyaz et al., 2017)
Oct-1	(Perovanovic, Shen, Wu, & Tantin, 2020)
ZBT16	(Beyaz et al., 2017)
DNA-damage and repair	
p53	(Chen et al., 2019), (Zhang et al., 2013), (Gao et al., 2019)
p53BP1	(Yang et al., 2019)
RAD50	(Chen et al., 2019)
Chromatin remodeling	

BRM	(Tie et al., 2012)
MTA1	(Biswas et al., 2019)
MTA2	(Biswas et al., 2019)
CHD4	(Biswas et al., 2019)
RBBP4	(Biswas et al., 2019)

Receptors

RXR/RAR	(Wang et al., 2017), (Rocha-Viegas et al., 2014)
ERα	(Xie et al., 2017)

Histones

H3	(Hong et al., 2007)

Histone demethylases

KDM6A	(Shpargel et al., 2012)
KDM6C	(Shpargel et al., 2012)
KDM7A	(Xie et al., 2017)

Benyoucef, A., Palii, C. G., Wang, C. C., Porter, C. J., Chu, A., Dai, F. T., . . . Brand, M. (2016). UTX inhibition as selective epigenetic therapy against TAL1-driven T-cell acute lymphoblastic leukemia. *Genes & Development, 30*(5), 508-521. doi:10.1101/gad.276790.115

Beyaz, S., Kim, J. H., Pinello, L., Xifaras, M. E., Hu, Y., Huang, J. L., . . . Winau, F. (2017). The histone demethylase UTX regulates the lineage-specific epigenetic program of invariant natural killer T cells. *Nature Immunology, 18*(2), 184-195. doi:10.1038/ni.3644

Biswas, M., Chatterjee, S. S., Boila, L. D., Chakraborty, S., Banerjee, D., & Sengupta, A. (2019). MBD3/NuRD loss participates with KDM6A program to promote DOCK5/8 expression and Rac GTPase activation in human acute myeloid leukemia. *Faseb Journal, 33*(4), 5268-5286. doi:10.1096/fj.201801035R

Chen, H., Huang, Y. X., Zhu, X. Q., Liu, C., Yuan, Y. M., Su, H., . . . Huang, K. (2019). Histone demethylase UTX is a therapeutic target for diabetic kidney disease. *Journal of Physiology-London, 597*(6), 1643-1660. doi:10.1113/jp277367

Gao, J., Zou, J., Li, J., Zhang, Y., Qiao, L., Wang, F., . . . Liu, C. (2019). Folate deficiency facilitates coordination of KDM6A with p53 in response to DNA damage. *bioRxiv*, 591768. doi:10.1101/591768

Hong, S. H., Cho, Y. W., Yu, L. R., Yu, H., Veenstra, T. D., & Ge, K. (2007). Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. *Proceedings of the National Academy of Sciences of the United States of America, 104*(47), 18439-18444. doi:10.1073/pnas.0707292104

Kalisz, M., Bernardo, E., Beucher, A., Maestro, M. A., del Pozo, N., Millan, I., . . . Ferrer, J. (2020). HNF1A recruits KDM6A to activate differentiated acinar cell programs that suppress pancreatic cancer. *Embo Journal, 39*(9). doi:10.15252/embj.2019102808

Kato, H., Asamitsu, K., Sun, W. D., Kitajima, S., Yoshizawa-Sugata, N., Okamoto, T., . . . Poellinger, L. Cancer-derived UTX TPR mutations G137V and D336G impair interaction with MLL3/4 complexes and affect UTX subcellular localization. *Oncogene*. doi:10.1038/s41388-020-1218-3

Kim, J. H., Sharma, A., Dhar, S. S., Lee, S. H., Gu, B., Chan, C. H., . . . Lee, M. G. (2014). UTX and MLL4 Coordinately Regulate Transcriptional Programs for Cell Proliferation and Invasiveness in Breast Cancer Cells. *Cancer Research, 74*(6), 1705-1717. doi:10.1158/0008-5472.Can-13-1896

Lang, A., Yilmaz, M., Hader, C., Murday, S., Kunz, X., Wagner, N., . . . Schulz, W. A. (2019). Contingencies of UTX/KDM6A Action in Urothelial Carcinoma. *Cancers, 11*(4). doi:10.3390/cancers11040481

Leng, X., Wang, J., An, N., Wang, X., Sun, Y., & Chen, Z. (2020). Histone 3 lysine-27 demethylase KDM6A coordinates with KMT2B to play an oncogenic role in NSCLC by regulating H3K4me3. *Oncogene, 39*(41), 6468-6479. doi:10.1038/s41388-020-01449-y

Perovanovic, J., Shen, Z., Wu, Y., & Tantin, D. (2020). Oct1 recruits the histone lysine demethylase Utx to canalize lineage specification. *bioRxiv,* 2020.2012.2001.406488. doi:10.1101/2020.12.01.406488

Rickels, R., Wang, L., Iwanaszko, M., Ozark, P. A., Morgan, M. A., Piunti, A., . . . Shilatifard, A. (2020). A small UTX stabilization domain of Trr is conserved within mammalian MLL3-4/COMPASS and is sufficient to rescue loss of viability in null animals. *Genes & Development, 34*(21-22), 1493-1502. doi:10.1101/gad.339762.120

Rocha-Viegas, L., Villa, R., Gutierrez, A., Iriondo, O., Shiekhattar, R., & Di Croce, L. (2014). Role of UTX in Retinoic Acid Receptor-Mediated Gene Regulation in Leukemia. *Molecular and Cellular Biology, 34*(19), 3765-3775. doi:10.1128/mcb.00839-14

Rossetto, C. C., & Pari, G. (2012). KSHV PAN RNA Associates with Demethylases UTX and JMJD3 to Activate Lytic Replication through a Physical Interaction with the Virus Genome. *Plos Pathogens, 8*(5). doi:10.1371/journal.ppat.1002680

Shpargel, K. B., Sengoku, T., Yokoyama, S., & Magnuson, T. (2012). UTX and UTY Demonstrate Histone Demethylase-Independent Function in Mouse Embryonic Development. *Plos Genetics, 8*(9), 17. doi:10.1371/journal.pgen.1002964

Shpargel, K. B., Starmer, J., Wang, C. C., Ge, K., & Magnuson, T. (2017). UTX-guided neural crest function underlies craniofacial features of Kabuki syndrome. *Proceedings of the National Academy of Sciences of the United States of America, 114*(43), E9046-E9055. doi:10.1073/pnas.1705011114

Smith, E. R., Lee, M. G., Winter, B., Droz, N. M., Eissenberg, J. C., Shiekhattar, R., & Shilatifard, A. (2008). Drosophila UTX is a histone H3 Lys27 demethylase that colocalizes with the elongating form of RNA polymerase II. *Molecular and Cellular Biology, 28*(3), 1041-1046. doi:10.1128/mcb.01504-07

Tie, F., Banerjee, R., Conrad, P. A., Scacheri, P. C., & Harte, P. J. (2012). Histone Demethylase UTX and Chromatin Remodeler BRM Bind Directly to CBP and Modulate Acetylation of Histone H3 Lysine 27. *Molecular and Cellular Biology, 32*(12), 2323-2334. doi:10.1128/mcb.06392-11

Vandamme, J., Lettier, G., Sidoli, S., Di Schiavi, E., Jensen, O. N., & Salcini, A. E. (2012). The C. elegans H3K27 Demethylase UTX-1 Is Essential for Normal Development, Independent of Its Enzymatic Activity. *Plos Genetics, 8*(5). doi:10.1371/journal.pgen.1002647

Wang, S. P., Tang, Z. Y., Chen, C. W., Shimada, M., Koche, R. P., Wang, L. H., . . . Roeder, R. G. (2017). A UTX-MLL4-p300 Transcriptional Regulatory Network Coordinately Shapes Active Enhancer Landscapes for Eliciting Transcription. *Molecular Cell, 67*(2), 308-+. doi:10.1016/j.molcel.2017.06.028

Watanabe, S., Shimada, S., Akiyama, Y., Ishikawa, Y., Ogura, T., Ogawa, K., . . . Tanaka, S. (2019). Loss of KDM6A characterizes a poor prognostic subtype of human pancreatic cancer and potentiates HDAC inhibitor lethality. *International Journal of Cancer, 145*(1), 192-205. doi:10.1002/ijc.32072

Xie, G., Liu, X., Zhang, Y., Li, W., Liu, S., Chen, Z., . . . Liang, J. (2017). UTX promotes hormonally responsive breast carcinogenesis through feed-forward transcription regulation with estrogen receptor. *Oncogene, 36*(39), 5497-5511. doi:10.1038/onc.2017.157

Yang, X. Y., Xu, B. S., Mulvey, B., Evans, M., Jordan, S., Wang, Y. D., . . . Peng, J. C. (2019). Differentiation of human pluripotent stem cells into neurons or cortical organoids requires transcriptional co-regulation by UTX and 53BP1. *Nature Neuroscience, 22*(3), 362-+. doi:10.1038/s41593-018-0328-5

Yu, W. Q., Huang, W., Yang, Y., Qiu, R. F., Zeng, Y., Hou, Y. Q., . . . Wang, Y. (2019). GATA3 recruits UTX for gene transcriptional activation to suppress metastasis of breast cancer. *Cell Death & Disease, 10.* doi:10.1038/s41419-019-2062-7

Zha, L., Cao, Q., Cui, X., Li, F. F., Liang, H. J., Xue, B. Z., & Shi, H. (2016). Epigenetic regulation of E-cadherin expression by the histone demethylase UTX in colon cancer cells. *Medical Oncology, 33*(3), 11. doi:10.1007/s12032-016-0734-z

Zha, L., Li, F. F., Wu, R., Artinian, L., Rehder, V., Yu, L. Q., . . . Shi, H. (2015). The Histone Demethylase UTX Promotes Brown Adipocyte Thermogenic Program Via Coordinated Regulation of H3K27 Demethylation and Acetylation. *Journal of Biological Chemistry, 290*(41), 25151-25163. doi:10.1074/jbc.M115.662650

Zhang, C. W., Hong, Z. H., Ma, W. C., Ma, D., Qian, Y. C., Xie, W., . . . Fang, M. (2013). Drosophila UTX Coordinates with p53 to Regulate ku80 Expression in Response to DNA Damage. *Plos One, 8*(11). doi:10.1371/journal.pone.0078652

Chapter 4-Table supplement 2

PTM predictions for all residues mutated in our study, both in original state and after mutagenesis.

Mutation:	WT PTM (predicted):	Mutant PTM (predicted):
E315Q	E315: Proteolytic cleavage (low conf.)	Q315: Proteolytic cleavage (low confidence)
		Q315: Pyrrolidone Carboxylic acid (med. conf.)
D336G	D336: none	G336: none
T726K	T726: Phosphorylation (low confidence)	K726: Acetylation (low confidence)
		K726: Proteolytic cleavage (medium conf.)
	T726: Proteolytic cleavage (medium confidence)	K726: Methylation (high confidence)
		K726: PUPylation (medium confidence)
		K726: Ubiquitination (low confidence)
T726V	T726: As above	V726: none
P966R	P966: none	R966: none
Q1133A	Q1133: Amidation (medium confidence)	A1133: none
H1329A	H1329: none	A1329: none
V1338F	V1338: none	F1338: Amidation (high confidence)
C1361Y	C1361: Disulfide (medium confidence)	Y1361: none

Pejaver, V., et al., The structural and functional signatures of proteins that undergo multiple events of post-translational modification. Protein Science, 2014. 23(8): p. 1077-1093.

Chapter 4-Table supplement 3

Parameters which are fitted are given in **bold**, linked parameters in *italic*. fixed parameters are plain text.

Species	eGFP-KDM6A Donly			KDM6A-tagGFP2 Donly		
Fit-Nbr	1	2	3	1	2	3
bg	**229.632**	**162.772**	**191.763**	**287.520**	**196.581**	**56.288**
dt	0.032	0.032	0.032	0.032	0.032	0.032
g	1.000	1.000	1.000	1.000	1.000	1.000
l1	0.031	0.031	0.031	0.031	0.031	0.031
l2	0.037	0.037	0.037	0.037	0.037	0.037
lb	**541.428**	**188.267**	**768.486**	**603.088**	**891.574**	**1425.497**
n0	2621153.813	2716139.613	1868967.524	2988346.958	1853770.021	641789.232
r0	0.380	0.380	0.380	0.380	0.380	0.380
rep	40.000	40.000	40.000	40.000	40.000	40.000
sc	0.000	*0.000*	*0.000*	0.000	*0.000*	*0.000*
start	0.000	0.000	0.000	0.000	0.000	0.000
stop	32.768	32.768	32.768	32.768	32.768	32.768
tBg	1.000	1.000	1.000	1.000	1.000	1.000
tDead	85.000	85.000	85.000	85.000	85.000	85.000
tL1	**3.096**	**2.711**	**2.970**	**3.084**	**3.060**	**3.905**
tL2	**2.036**	**1.668**	**1.973**	**1.784**	**1.808**	**2.422**
tMeas	1.000	1.000	1.000	1.000	1.000	1.000
ts	**-0.533**	**-0.452**	**-0.625**	**-0.198**	**-0.359**	**-0.131**
win-size	17.000	17.000	17.000	17.000	17.000	17.000
xL1	**0.379**	**0.705**	**0.442**	**0.666**	**0.700**	**0.215**
xL2	**0.621**	**0.295**	**0.558**	**0.334**	**0.300**	**0.785**
Chi2r	1.0212	1.0025	1.0697	1.0689	0.9642	1.0613

Chapter 4-Table supplement 4

Parameters which are fitted are given in **bold**, linked parameters in *italic*. fixed parameters are plain text.

Species	eGFP-KDM6A/mCh-RBBP5					
Fit-Nbr	1	2	3	4	5	6
bg	**66.439**	**241.821**	**258.470**	**240.139**	**359.484**	**342.871**
dt	0.032	0.032	0.032	0.032	0.032	0.032
g	1.000	1.000	1.000	1.000	1.000	1.000
l1	0.031	0.031	0.031	0.031	0.031	0.031
l2	0.037	0.037	0.037	0.037	0.037	0.037
lb	**178.464**	**1523.192**	**1008.270**	**1306.896**	**269.472**	**216.563**
n0	1409508.580	2768683.738	2886104.133	2515776.437	3673933.814	3814994.318
r0	0.380	0.380	0.380	0.380	0.380	0.380
rep	40.000	40.000	40.000	40.000	40.000	40.000
sc	0.000	0.000	*0.000*	0.000	*0.000*	*0.000*
start	0.000	0.000	0.000	0.000	0.000	0.000
stop	32.768	32.768	32.768	32.768	32.768	32.768
tBg	1.000	1.000	1.000	1.000	1.000	1.000
tDead	85.000	85.000	85.000	85.000	85.000	85.000
tL1	**2.652**	**2.948**	**2.766**	**3.272**	**2.705**	**2.650**
tL2	**1.472**	**1.947**	**1.684**	**2.171**	**1.543**	**1.513**
tMeas	1.000	1.000	1.000	1.000	1.000	1.000
ts	**0.395**	**0.173**	**-0.065**	**0.093**	**0.065**	**-0.047**
win-size	17.000	17.000	17.000	17.000	17.000	17.000
xL1	**0.800**	**0.470**	**0.677**	**0.265**	**0.716**	**0.776**
xL2	**0.200**	**0.530**	**0.323**	**0.735**	**0.284**	**0.224**
Chi2r	1.0312	1.1216	1.0306	1.0362	1.0807	0.9753

Chapter 4-Table supplement 5

Parameters which are fitted are given in **bold**, linked parameters in *italic*. fixed parameters are plain text.

Species	eGFP-KDM6A/RBBP5-mCh					
Fit-Nbr	1	2	3	4	5	6
bg	**9.834**	**29.338**	**79.156**	**87.411**	**33.441**	**229.471**
dt	0.032	0.032	0.032	0.032	0.032	0.032
g	1.000	1.000	1.000	1.000	1.000	1.000
l1	0.031	0.031	0.031	0.031	0.031	0.031
l2	0.037	0.037	0.037	0.037	0.037	0.037
lb	**1020.779**	**288.175**	**2316.360**	**949.582**	**175.765**	**1159.677**
n0	93263.183	270559.314	799554.955	912893.086	571664.745	1933407.844
r0	0.380	0.380	0.380	0.380	0.380	0.380
rep	40.000	40.000	40.000	40.000	40.000	40.000
sc	0.000	*0.000*	*0.000*	0.000	*0.000*	*0.000*
start	0.000	0.000	0.000	0.000	0.000	0.000
stop	32.768	32.768	32.768	32.768	32.768	32.768
tBg	1.000	1.000	1.000	1.000	1.000	1.000
tDead	85.000	85.000	85.000	85.000	85.000	85.000
tL1	**2.637**	**2.554**	**2.852**	**2.896**	**2.661**	**2.783**
tL2	**1.154**	**1.093**	**1.843**	**2.007**	**1.421**	**1.927**
tMeas	1.000	1.000	1.000	1.000	1.000	1.000
ts	**0.221**	**0.010**	**-0.188**	**-0.238**	**-0.294**	**-0.295**
win-size	17.000	17.000	17.000	17.000	17.000	17.000
xL1	**0.886**	**0.902**	**0.570**	**0.485**	**0.809**	**0.621**
xL2	**0.114**	**0.098**	**0.430**	**0.515**	**0.191**	**0.379**
Chi2r	1.0407	1.0441	0.9906	1.0136	0.9841	1.0655

Chapter 4-Table supplement 6

Parameters which are fitted are given in **bold**, linked parameters in *italic*. fixed parameters are plain text.

Species	KDM6A-tagGFP2/mCh-RBBP5				
Fit-Nbr	1	2	3	4	5
bg	**49.175**	**105.140**	**182.533**	**52.266**	**55.283**
dt	0.032	0.032	0.032	0.032	0.032
g	1.000	1.000	1.000	1.000	1.000
l1	0.031	0.031	0.031	0.031	0.031
l2	0.037	0.037	0.037	0.037	0.037
lb	**2696.035**	**1144.323**	**830.971**	**6761.800**	**179.365**
n0	584556.444	1080445.790	1964020.068	1732919.929	1515426.628
r0	0.380	0.380	0.380	0.380	0.380
rep	40.000	40.000	40.000	40.000	40.000
sc	0.000	*0.000*	*0.000*	*0.000*	*0.000*
start	0.000	0.000	0.000	0.000	0.000
stop	32.768	32.768	32.768	32.768	32.768
tBg	1.000	1.000	1.000	1.000	1.000
tDead	85.000	85.000	85.000	85.000	85.000
tL1	**3.460**	**3.184**	**2.021**	**0.000**	**3.009**
tL2	**2.295**	**2.070**	**3.203**	**2.755**	**1.680**
tMeas	1.000	1.000	1.000	1.000	1.000
ts	**0.670**	**0.565**	**0.410**	**-1.415**	**0.538**
win-size	17.000	17.000	17.000	17.000	17.000
xL1	**0.408**	**0.565**	**0.416**	**0.809**	**0.731**
xL2	**0.592**	**0.435**	**0.584**	**0.191**	**0.269**
Chi2r	1.1216	1.1085	1.0522	1.1094	0.9487

Chapter 4-Table supplement 7

Parameters which are fitted are given in **bold**, linked parameters in *italic*. fixed parameters are plain text.

Species	KDM6A-tagGFP2/RBBP5-mCh				
Fit-Nbr	1	2	3	4	5
bg	**34.870**	**35.315**	**94.490**	**11.709**	**114.190**
dt	0.032	0.032	0.032	0.032	0.032
g	1.000	1.000	1.000	1.000	1.000
l1	0.031	0.031	0.031	0.031	0.031
l2	0.037	0.037	0.037	0.037	0.037
lb	**256.525**	**2597.394**	**541.273**	**3571.406**	**1139.532**
n0	276046.357	546936.667	827422.277	261950.851	973398.667
r0	0.380	0.380	0.380	0.380	0.380
rep	40.000	40.000	40.000	40.000	40.000
sc	0.000	*0.000*	*0.000*	*0.000*	*0.000*
start	0.000	0.000	0.000	0.000	0.000
stop	32.768	32.768	32.768	32.768	32.768
tBg	1.000	1.000	1.000	1.000	1.000
tDead	85.000	85.000	85.000	85.000	85.000
tL1	**2.886**	**5.205**	**1.603**	**2.554**	**3.165**
tL2	**1.133**	**2.622**	**2.977**	**5.494**	**1.979**
tMeas	1.000	1.000	1.000	1.000	1.000
ts	**0.233**	**0.634**	**0.095**	**0.480**	**1.075**
win-size	17.000	17.000	17.000	17.000	17.000
xL1	**0.894**	**0.069**	**0.194**	**0.918**	**0.383**
xL2	**0.106**	**0.931**	**0.806**	**0.082**	**0.617**
Chi2r	0.8310	0.9288	0.8848	1.0688	1.0573

Chapter 4-Table supplement 8

Scoring results from Fig. 15 (Adapted from Koch et al., 2021)

T-24						
Untransfected (n = 815)	[binucl.] %normal	[binucl.] %dmg	[binucl.] %MN	[mononucl.] %normal	[mononucl.] %dmg	[mononucl.] %MN	%mitotic
Weighted mean %:	5.767	0.736	0.245	86.258	1.718	2.822	2.454
Weighted STDDEV:	4.679	0.203	0.152	3.339	1.927	2.218	1.519
eGFP (n = 374)	[binucl.] %normal	[binucl.] %dmg	[binucl.] %MN	[mononucl.] %normal	[mononucl.] %dmg	[mononucl.] %MN	%mitotic
Weighted mean %:	4.813	1.070	1.070	74.064	11.230	5.080	2.674
Weighted STDDEV:	0.083	0.764	1.102	5.544	6.428	5.077	1.688
eGFP-KDM6A WT (n = 383)	[binucl.] %normal	[binucl.] %dmg	[binucl.] %MN	[mononucl.] %normal	[mononucl.] %dmg	[mononucl.] %MN	%mitotic
Weighted mean %:	6.527	1.305	1.044	71.802	13.838	2.350	3.133
Weighted STDDEV:	1.100	1.216	0.914	5.262	9.130	0.942	1.407
eGFP-KDM6A TPR (n = 360)	[binucl.] %normal	[binucl.] %dmg	[binucl.] %MN	[mononucl.] %normal	[mononucl.] %dmg	[mononucl.] %MN	%mitotic
Weighted mean %:	2.500	7.778	0.278	41.111	45.278	1.944	1.111
Weighted STDDEV:	0.896	1.285	0.458	6.365	6.916	1.806	0.779
eGFP-KDM6A JmjC (n = 476)	[binucl.] %normal	[binucl.] %dmg	[binucl.] %MN	[mononucl.] %normal	[mononucl.] %dmg	[mononucl.] %MN	%mitotic
Weighted mean %:	3.151	6.513	0.840	36.975	49.370	1.681	1.471
Weighted STDDEV:	2.735	2.541	0.434	3.805	8.376	0.872	0.950
SW-1710						
Untransfected (n = 429)	[binucl.] %normal	[binucl.] %dmg	[binucl.] %MN	[mononucl.] %normal	[mononucl.] %dmg	[mononucl.] %MN	%mitotic
Weighted mean %:	3.497	0.233	0.233	90.443	1.399	0.699	3.497
Weighted STDDEV:	2.185	0.278	0.264	2.546	1.306	0.403	1.216
eGFP (n = 303)	[binucl.] %normal	[binucl.] %dmg	[binucl.] %MN	[mononucl.] %normal	[mononucl.] %dmg	[mononucl.] %MN	%mitotic
Weighted mean %:	6.601	1.980	0.660	78.548	6.271	3.630	1.320
Weighted STDDEV:	4.325	1.219	0.440	4.769	3.637	3.093	0.919
eGFP-KDM6A WT (n = 287)	[binucl.] %normal	[binucl.] %dmg	[binucl.] %MN	[mononucl.] %normal	[mononucl.] %dmg	[mononucl.] %MN	%mitotic
Weighted mean %:	7.317	3.833	0.348	68.990	14.634	3.136	1.742
Weighted STDDEV:	2.734	0.175	0.543	5.270	1.870	1.255	1.208
eGFP-KDM6A TPR (n = 223)	[binucl.] %normal	[binucl.] %dmg	[binucl.] %MN	[mononucl.] %normal	[mononucl.] %dmg	[mononucl.] %MN	%mitotic
Weighted mean %:	4.484	9.417	1.345	34.081	43.049	3.587	4.036
Weighted STDDEV:	1.498	1.235	0.764	3.234	5.586	1.297	3.089
eGFP-KDM6A JmjC (n = 445)	[binucl.] %normal	[binucl.] %dmg	[binucl.] %MN	[mononucl.] %normal	[mononucl.] %dmg	[mononucl.] %MN	%mitotic
Weighted mean %:	3.820	10.562	0.449	48.989	28.539	2.697	4.944
Weighted STDDEV:	1.281	3.971	0.429	4.923	4.218	1.168	2.932

Chapter 4-Table supplement 9

P-Values from Fig. 15 (Adapted from Koch et al., 2021)

Hypothesis tested:	P-value:	Sign.:	P-value:	Sign.:
	T-24		SW-1710	
Different percentage of total (mono/binuclear, mitosis) normal cells in WT vs. eGFP	0.9883	N.S.	0.0144	*
Different percentage of total (mono/binuclear, mitosis) normal cells in TPR vs. eGFP	0.0026	**	0.0001	***
Different percentage of total (mono/binuclear, mitosis) normal cells in JmjC vs. eGFP	0.0013	**	0.0047	**
Different percentage of binuclear cells in WT vs. eGFP	0.2795	N.S.	0.5566	N.S.
Different percentage of binuclear cells in TPR vs. eGFP	0.0164	*	0.1608	N.S.
Different percentage of binuclear cells in JmjC vs. eGFP	0.3400	N.S.	0.2134	N.S.
Different percentage of cells with micronuclei (MN) in WT vs. eGFP	0.4494	N.S.	0.7903	N.S.
Different percentage of cells with micronuclei (MN) in TPR vs. eGFP	0.3132	N.S.	0.7956	N.S.
Different percentage of cells with micronuclei (MN) in JmjC vs. eGFP	0.3192	N.S.	0.6114	N.S.
Different percentage of mitotic cells in WT vs. eGFP	0.7356	N.S.	0.6552	N.S.
Different percentage of mitotic cells in TPR vs. eGFP	0.2191	N.S.	0.2182	N.S.
Different percentage of mitotic cells in JmjC vs. eGFP	0.3425	N.S.	0.1106	N.S.
Different percentage of mononuclear normal cells in WT vs. eGFP	0.6352	N.S.	0.0803	N.S.
Different percentage of mononuclear normal cells in TPR vs. eGFP	0.0025	**	0.0002	***
Different percentage of mononuclear normal cells in JmjC vs. eGFP	0.0007	***	0.0017	**
Different percentage of binuclear damaged cells in WT vs. eGFP	0.7901	N.S.	0.0597	N.S.
Different percentage of binuclear damaged cells in TPR vs. eGFP	0.0015	**	0.0018	**
Different percentage of binuclear damaged cells in JmjC vs. eGFP	0.0237	*	0.0232	*

Chapter 4-Table supplement 10

P-Values from Fig. 16 (Adapted from Koch et al., 2021)

Hypothesis tested:	P-value:	Sign.:	P-value:	Sign.:
	T-24		SW-1710	
Different percentage of viable cells in WT vs. eGFP	0.0655	N.S.	0.0758	N.S.
Different percentage of viable cells in TPR vs. eGFP	0.0002	***	0.0001	***
Different percentage of viable cells in JmjC vs. eGFP	0.0009	***	0.0213	*

Chapter 4-Table supplement 11

Constructs & variants

Construct	Protein	Vector
pcDNA 3.1 (+) eGFP-UTX	eGFP-UTX	pcDNA 3.1 (+)
pcDNA 3.1 (+) UTX-eGFP	UTX-eGFP	pcDNA 3.1 (+)
pcDNA 3.1 (+) mCh-UTX	mCh-UTX	pcDNA 3.1 (+)
pcDNA 3.1 (+) UTX-mCh	UTX-mCh	pcDNA 3.1 (+)
pEGFP-C1 eGFP-WDR5	eGFP-WDR5	pEGFP-C1
pEGFP-N1 WDR5-eGFP	WDR5-eGFP	pEGFP-N1
pmCherry-C1 mCh-RBBP5	mCh-RBPP5	pmCherry-C1
pmCherry-N1 RBBP5-mCh	RBPP5-mCh	pmCherry-N1
pmCherry-C1 mCherry-FLAG-RBBP5	mCh-FLAG-RBBP5	pmCherry-C1
puB-eGFP	free eGFP	pCAGEN
pcDNA5 mCherry	free mCherry	pcDNA 5
puc2CL12 UTX-tagGFP2	UTX-tagGFP2	pLenti-C-mGFP
puc2CL12 tagGFP2	tagFGP2	pLenti-C-mGFP
Variant	**Protein**	**Vector**
E315Q	eGFP-UTX	pCDNA 3.1 (+)
D336G	eGFP-UTX	pCDNA 3.1 (+)
P966R	eGFP-UTX	pCDNA 3.1 (+)
Q1133A	eGFP-UTX	pCDNA 3.1 (+)
H1329A	eGFP-UTX	pCDNA 3.1 (+)
V1338F	eGFP-UTX	pCDNA 3.1 (+)
C1361Y	eGFP-UTX	pCDNA 3.1 (+)
T726V	eGFP-UTX	pCDNA 3.1 (+)
T726K	eGFP-UTX	pCDNA 3.1 (+)
RE linker*	eGFP-UTX	pEGFP-C1
Δ-TPR	eGFP-UTX	pEGFP-C1
Δ-IDR	eGFP-UTX	pEGFP-C1
Δ-JmjC	eGFP-UTX	pEGFP-C1
TPR	eGFP-UTX	pEGFP-C1
JmjC	eGFP-UTX	pEGFP-C1
TPR-SPOT	eGFP-UTX	pEGFP-C1

Chapter 4-Table supplement 12

Primer

Variant/direction	Sequence 5'-3'
E315Q Fwd	ATCGACAAGAGCCAGGCCTC
E315Q Rev	AGGCCTGGCTCTTGTCGATG
D336G Fwd	ATGGGCGCCCTGCAGGCCTACAT
D336G Rev	ATGTAGGCCTGCAGGGCGCCCAT
T726K Fwd	CAACATCCTGAAGGTGCCTGAGAC
T726K Rev	GTCTCAGGCACCTTCAGGATGTTG
T726V Fwd	CAACATCCTGGTCGTGCCTGAGAC
T726V Rev	GTCTCAGGCACGACCAGGATGTTG
P966R Fwd	CAAATCCCAATAATCGTGTGACCGTGATC
P966R Rev	GATCACGGTCACACGATTATTGGGATTTG
Q1133A Fwd	CATGAATACAGTGGCGCTGTACATGAAG
Q1133A Rev	CTTCATGTACAGCGCCACTGTATTCATG
H1329A Fwd	GAGGAGCCTGCCGCCTACTGTAGCATCTG
H1329A Rev	CAGATGCTACAGTAGGCGGCAGGCTCCTC
V1338F Fwd	AGGTGGAGTTCTTCGACCTGCTGTT
V1338F Rev	AACAGCAGGTCGAAGAACTCCACCT
C1361Y Fwd	ACTGCCAGGACTACGCCAGAAAGA
C1361Y Rev	TCTTTCTGGCGTAGTCCTGGCAGT

Domain/direction	RE	Sequence 5'-3'
TPR front Fwd	BspEI	ATATTCCGGAATGAAGTCCTGCGGCGTGT
TPR front Rev	HindIII	TATAAGCTTCAGGGCGGATGTATTGCTGCA
TPR front stop Rev	HindIII	TATAAGCTTTCACAGGGCGGATGTATTGCTGCA
IDR front Fwd	BspEI	ATATTCCGGAGCCGCCAGGATCAAGTACCT
IDR front Rev	HindIII	TATAAGCTTCATGCTGGGGATGATCTGAGGG
IDR back Fwd	HindIII	TATAAGCTTGCCGCCAGGATCAAGTACCT
IDR back Rev	EcoRI	ATATGAATTCCATGCTGGGGATGATCTGAGGG
IDR back stop Rev	EcoRI	ATATGAATTCTCACATGCTGGGGATGATCTGAGGG
JmjC front Fwd	BspEI	ATATTCCGGATCCGTGTCCATCTACCCCTCCTCC
JmjC front Rev	HindIII	TATAAGCTTGCTGGAGGCGCTAGGCAGA
JmjC front stop Rev	HindIII	TATAAGCTTTCAGCTGGAGGCGCTAGGCA
JmjC back Fwd	HindIII	TATAAGCTTTCCGTGTCCATCTACCCCTCCT
JmjC back rev	EcoRI	ATATGAATTCGCTGGAGGCGCTAGGCAGA
JmjC back stop Rev	EcoRI	ATATGAATTCTCAGCTGGAGGCGCTAGGCA
JmjC third Fwd	EcoRI	ATATGAATTCTCCGTGTCCATCTACCCCTCCT
JmjC third Rev	KpnI	TATGGTACCGCTGGAGGCGCTAGGCAGA

JmjC third stop rev	KpnI	TATGGTACCTCAGCTGGAGGCGCTAGGCA
RBBP5 C1 fwd	BspEI	GATTCCGGAATGAACCTCGAGTTGCTG
RBBP5 C1 rev	ApaI	AATAGGGCCCTCTAGATG
RBBP5 N1 fwd	HindIII	TCATTAAGCTTATGAACCTCGAGTTGCTGG
RBBP5 N1 rev	AgeI	GATACCGGTAGTAACAGTTCTGAGATTGCTCCT
WDR5 C1 fwd	BspEI	ATATCCGGAATGGCGACGGAGGAGAAGAAG
WDR5 C1 rev	ApaI	ATAGGGCCCACTAGCAGTCACTCTTCCACAGTT
WDR5 N1 fwd	HindIII	ATTAAGCTTCGGATGGCGAC
WDR5 N1 rev	AgeI	GATACCGGTAGGCAGTCACTCTTCCACAGTT

Chapter 4-Table supplement 13

(Adapted from Koch et al., 2021)

Primary antibodies				
Antibody	Source/Isotype	Clone	Company	Cat.-Nr.
α-KDM6A	Rabbit IgG	D3Q1I	Cell Signaling Technology (CST)	33510
α-KDM6A	Mouse IgG	E-8	Santa Cruz Biotechnology (SC)	Sc-514859
α-RBBP5	Rabbit IgG	D3I6P	Cell Signaling Technology	13171
α-GFP	Mouse IgG2a	4B10	Cell Signaling Technology	2955
α-p-γH2AX	Rabbit IgG	20E3	Cell Signaling Technology	9718
α-NPM1	Mouse IgG1	E-3	Santa Cruz Biotechnology (SC)	sc-271737
α-α-Tubulin	Mouse IgG1	B-5-1-2	Sigma-Aldrich	T6074
α-H3K27me2	Mouse IgG2a	0324	Active Motif (AM)	61435
Secondary antibodies				
Conjugation	Antigen	Host	Company	Cat.-Nr.
HRP	Rabbit IgG	Goat	Cell Signaling Technology	7076
HRP	Mouse IgG	Horse	Cell Signaling Technology	7074
Star Orange	Mouse IgG	Goat	Abberior	STORANGE-1001
Star Red	Rabbit IgG	Goat	Abberior	STRED-1002
ALP	Mouse IgG2a	Rabbit	Sigma-Aldrich	SAB3701179

Chapter 4-Table supplement 14

(Adapted from Koch et al., 2021)

Peptide	Res.	Sequence	Modification	MW	Company	Cat. Nr.
H3K27 me3	15-34	Ac-APRKQATKAAR K(me3)SAPATGG-PEG-Bio	N-Acetyl C-PEGylate/Biotin	2477.9 g/mol	Proteo-Genix	Custom order
H3K27 me2	21-44	NH2-ATKAARK(me2)SAPAT GGVKKPHRYRPG-PEG-Bio	C-PEGylate/ Biotin	2945.5 g/mol	Epi-Gentek	R-1033

Conclusion

Throughout this work, we were able to demonstrate the power of fluorescence spectroscopy to resolve protein dynamics and structure, even in a cellular context. In addition, we used fluorescence imaging to achieve readout and statistics on cellular localization and phenotypes, namely large protein particles the severely damaged phenotype of cells. We also showed the importance of biochemical methods both as complementing and as pre-requisite for the fluorescence spectroscopy measurements. In project **T4L (Chapter 1)**, we were able to correctly assess the number of photophysical processes in the system and prepare further experimental workflows. In project **hGBP1 (Chapter 2)**, we were able to follow the domain movement of hGBP1 upon GTPγS-induced dimerization. We were able to complement existing data on the protein start-point structure and end-point structure after dimerization with the domain movement in between on a low millisecond time-resolution. We were able to recover rates from our datasets and made a comprehensive model to show the transitioning of hGBP1 through its dimer-stage from an early dimer to a somewhat elongated molecule. In project **mGBP7 (Chapter 3)**, we were able to show specific affinity and phase behavior of mGBP7 with two other mGBPs, mGBP2 and mGBP3. We identified the protein phases as dynamic and found structural basis for the observed effects. In the main project, **KDM6A (Chapter 4)**, we managed to comprehensibly characterize functional mutants and variants found in cancer in regard to their activity, localization, interaction and the cellular phenotype they generate. By using a wide variety of biochemical methods, we identified the truncation of the multi-domain protein KDM6A as a trigger for DNA-damage, DNA release into the cytoplasm and higher apoptosis-rate. In addition, the outcome of the biochemical methods enabled us to plan further highly sophisticated fluorescence spectroscopy experiments. Concluding, we strongly encourage people to do spectroscopic research in cellular high-complexity systems with attention to a couple of points: Spectroscopic experiments dealing with high levels of systemic complexity need to be planned accordingly and the desired readout should match a defined experimental question. The more complex the system is spatially; the more structural knowledge is needed to establish high-quality FRET-pairs. The experiment should be paired with a series of complementing (biochemical or biophysical) methods, which will generate additional readout but can also be used synergistic. A real potential lies within computational modelling of the system to find the right fit-model and to bolster and validate the data. Deep learning AIs for structure and soon even interface prediction can serve as a valuable tool to position the dye-pairs to the right place in unknown proteins if combined with our already developed FRET-prediction tools. If this synergy is further developed, there is a huge potential for all fluorescence spectroscopy-based methods to shine in the next decades and evolve along with the raising computational methods.

Acknowledgements

In the first place, I want to thank Prof. Claus A. M. Seidel for the opportunity to do scientific research at the top-level in his group, his continued support and for all the work and patience he invested to transform me into a grown-up scientist.

I want to thank my direct supervisor, Dr. Annemarie Greife, for the good time I had in our project, for her continuous trust, all the feedback and insights, the awesome interdisciplinary science, the laughs and for going through ups and downs together, always standing as a team.

I further want to thank Jun.-Prof Alexej Kedrov for his invaluable help as a mentor, for never turning me down when I dropped by and for always making me smile in the end.

A very special thanks to Oleg Opanasyuk, Julian Folz and Alexander Larbig, who helped and contributed tremendously to rewrite, polish and finish off Chapter 2.

I want to thank all the people who further contributed and helped me with my projects, namely Thomas-Otavio Peulen (hGBP1), Aiswaria Prakash and Laura Stephan (T4L), and Jan-Hendrik Budde (hGBP1 + mGBP7). I also want to thank Nicolaas van der Voort for all the great scientific discussion we had on many of the topics within the projects.

I further want to thank the whole urology research lab, in persona Prof. em. Wolfgang A. Schulz, Dr. Michèle J. Hoffmann, Christiane Hader and Dr. Alexander Lang for always letting me feel welcome in their lab, the interesting (medical) point of views and all the help and feedback I received.

I want to thank the whole AG Seidel for the good time together, for always being helpful and always working as a big team.

I furthermore want to thank my current PI Rune Busk Damgaard of the DTU for supporting me by giving me all the time I needed to wrap up my thesis.

Finally, I want to thank my friends & family and in particular Jacqueline for all the patience, the encouragement and the support I got over the years. It was not always easy, but you made me push through. Thank you!

Publications

1. **Koch, J.**[1], Lang, A.[2], Whongsiri, P.[2], Schulz, W. A.[2], Hoffmann, M. J.[2], Greife, A.[1,*] KDM6A mutations promote acute cytoplasmic DNA release, DNA damage response and mitosis defects, *BMC Mol and Cell Biol* 22, 54 (2021).
   Contributions: AG, **JK**, WAS conceived the project and designed the experiments. **JK** and AG conducted the experiments, collected, interpreted the data and wrote the manuscript. AL and WP helped with experiments and resources. MJH, WAS reviewed and edited the manuscript. **All authors** read and approved the final Manuscript.

2. Schulz, W. A.[2,*], Lang, A.[2], **Koch, J.**[1], Greife, A.[1,*] The Histone Demethylase UTX/KDM6A in Cancer: Progress and Puzzles. *Int J Cancer* 145, 614-620 (2019).
   Contributions: WAS, AL, **JK**, AG wrote the manuscript. WAS and AG reviewed and edited the manuscript. **All authors** read and approved the final Manuscript.

3. Lang, A.[2], Yilmaz, M.[2], Hader, C.[2], Murday, S.[2], Kunz, X.[2], Wagner, N.[2], Wiek, C.[3], Petzsch, P.[4], Köhrer, K.[4], **Koch, J.**[1], Hoffmann, M. J.[2], Greife, A.[1,*], Schulz, W. A.[2,*] Contingencies of UTX/KDM6A Action in Urothelial Carcinoma. Cancers 11, 481 (2019).
   Contributions: AL, AG, WAS conceptualized the project. Formal analysis was done by NW, PP, **JK**. Experiments were conducted by AL, MY, CH, SM, XK, PP. Resources were contributed by CW. Data curation was done by AL, PP. Writing the original draft was done by AL, WAS. Writing for reviewing and editing purposes was done by **all authors**. KK and MJH supervised the project. WAS administered the project. Funding was acquired by WAS, AG.

Legend:

*Corresponding author

Affiliations:

[1]Institute for Molecular Physical Chemistry

[2]Department of Urology, Medical Faculty

[3]Department of Otolaryngology, Medical Faculty

[4]Biological and Medical Research Center (BMFZ)

Heinrich-Heine-University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany

Manuscript in preparation

Preliminary title: Specificity of phase transition for murine guanylate binding proteins in live cells.

CLARIFICATION: As of 28.10.2021, the manuscript has not been read and approved by all authors listed below and might be subjected to further changes such as addition or removal of text or figures.

Preliminary author list: Koch, J.[1,†], Budde, J.[1,†], Kravets, E.[2], Ma, Q.[1], Steffens, N.[2], Reich, A.[2], Weidtkamp-Peters, S.[1,3], van der Voort, N.[1], Hänsch, S.[3], Kühnemuth, R.[1], Felekyan, S.[1], Degrandi, D.[2], Pfeffer, K.[2], Seidel, C. A. M.[1,*], Greife, A.[1,*]

CLARIFICATION: The author list given above is a preliminary author list, which has not been checked and approved by all authors and might be subjected to further changes such as addition or removal of authors.

Preliminary contributions: AG, CAMS, KP, DD, **JK**, JB conceptualized the project. AG, **JK**, JB, EK verified the data. AG, **JK**, JB, QM, NV, SF, RK applied formal analysis. AG, **JK**, JB, EK, QM, AR conducted investigation. CAMS, KP, DD, SH, EK provided resources. AG, **JK**, JB created the original draft. AG, CAMS, KP, DD did reviewing and editing. AG, CAMS, **JK**, JB visualized data. AG, CAMS did supervision. CAMS, KP, DD administered the project. CAMS, KP, DD acquired funding. **All authors** read and approved the final Manuscript.

CLARIFICATION: The contribution list given above is a preliminary contribution list, which has not been checked and approved by all authors and might be subjected to further changes such as addition or removal of contributions.

Legend:

*Corresponding author

†Authors contributed equally

Affiliations:

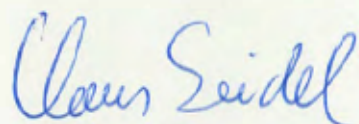[1]Institute for Molecular Physical Chemistry

[2]Institute of Medical Microbiology and Hospital Hygiene

[3]Center for Advanced Imaging

Heinrich-Heine-University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany

Credibility statement

I hereby confirm that all contributions listed under '**Publications**' and '**Manuscript in preparation**' are credible to the best of my knowledge.

Claus Seidel

Prof. Dr. Claus A. M. Seidel

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die Promotion mit dem Thema

"Multiparameter Fluorescence Spectroscopy and Microscopy of Biomolecular Systems"

am Institut für Physikalische Chemie II der Heinrich-Heine- Universität Düsseldorf unter der Leitung von Prof. Dr. Claus A.M. Seidel eigenständig und ohne unerlaubte Hilfe angefertigt und in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht habe.

Es existieren keine vorherigen Promotionsversuche.

Düsseldorf, den 17.10.2022

Julian Koch