Biophysical *Journal*
Article

Biophysical Society

# Mapping Multiple Distances in a Multidomain Protein for the Identification of Folding Intermediates

Michele Cerminara,[1,*] Antonie Schöne,[1] Ilona Ritter,[1] Matteo Gabba,[1] and Jörg Fitter[1,2,*]
[1]Forschungszentrum Jülich, Institute of Complex Systems ICS-5, Jülich, Germany and [2]RWTH Aachen University, I. Physikalisches Institut (IA), Aachen, Germany

ABSTRACT The investigation and understanding of the folding mechanism of multidomain proteins is still a challenge in structural biology. The use of single-molecule Förster resonance energy transfer offers a unique tool to map conformational changes within the protein structure. Here, we present a study following denaturant-induced unfolding transitions of yeast phosphoglycerate kinase by mapping several inter- and intradomain distances of this two-domain protein, exhibiting a quite heterogeneous behavior. On the one hand, the development of the interdomain distance during the unfolding transition suggests a classical two-state unfolding behavior. On the other hand, the behavior of some intradomain distances indicates the formation of a compact and transient molten globule intermediate state. Furthermore, different intradomain distances measured within the same domain show pronounced differences in their unfolding behavior, underlining the fact that the choice of dye attachment positions within the polypeptide chain has a substantial impact on which unfolding properties are observed by single-molecule Förster resonance energy transfer measurements. Our results suggest that, to fully characterize the complex folding and unfolding mechanism of multidomain proteins, it is necessary to monitor multiple intra- and interdomain distances because a single reporter can lead to a misleading, partial, or oversimplified interpretation.

SIGNIFICANCE The structural characterization of proteins on their way from a folded to an unfolded state and vice versa is key in elucidating the underlying principles of protein folding. Förster resonance energy transfer measurements, in particular on a single-molecule level, can give valuable structural information on the desired length scale of a few nanometers. To overcome the intrinsic limitation of monitoring only a single intramolecular distance within the protein by typical Förster resonance energy transfer measurements, we mapped five different intramolecular distances in phosphoglycerate kinase. The obtained data revealed the appearance of additional folding and unfolding intermediates and a more realistic and detailed picture about the unfolding process because we sample the energy folding funnel landscape along different directions.

## INTRODUCTION

Deciphering how the amino acid code translates into functional three-dimensional (3D) structures is the key to understanding how proteins really work. For this purpose, it is crucial to understand how the polypeptide chain can migrate from the manifold of unfolded states to a unique folded (native) state. A more detailed understanding of this process was obtained thanks to the development of the theory of funnel-shaped multidimensional folding energy landscapes (1). With the support of these theoretical predictions, experimental investigations based on single-molecule Förster resonance energy transfer (smFRET) spectroscopy and single molecule force spectroscopy contributed significantly to our current knowledge (2–7). The use of smFRET offers a unique tool to map conformational changes within the protein structure, thus allowing one to observe the distribution in populations. Although most of these studies dealt with small single-domain proteins, more recently, studies with multidomain proteins have attracted increasing interest (8–13). Multidomain proteins exhibit, typically, a much higher

*Submitted June 19, 2019, and accepted for publication December 10, 2019.*
*Correspondence: michele.cerminara@pasteur.fr or fitter@physik.rwth-aachen.de
Michele Cerminara's present address is Department of Developmental and Stem Cell Biology, UMR3738, Institut Pasteur, 75015 Paris, France.
Editor: Elizabeth Komives.
https://doi.org/10.1016/j.bpj.2019.12.006

complexity of the folding landscape than that of single-domain proteins. More than 70% of the eukaryotic proteins are composed of multiple domains, and hence, it is important to expand the knowledge acquired from model single-domain globular proteins to these more complex macromolecules. When studying multidomain proteins, the question arises whether the folding principles found for single-domain proteins also apply for multidomain proteins. Answering this question typically requires an expanded set of measured data, and smFRET offers a unique ability to measure, with a high accuracy, multiple distances between attachment positions of the two involved fluorescent dyes (14). To map all important conformational changes within multidomain proteins during folding or unfolding transitions, distances within each domain, as well as distances between the individual domains, need to be measured (8,9,15). A common limitation in studies on multidomain proteins is that they often manifest misfolding and aggregation of unfolded or not correctly folded states (16–20). This issue is intrinsically overcome by single-molecule techniques; in fact, the low sample concentrations used in these techniques minimize the chance of protein aggregation and, in addition, allow for sorting out data from aggregates.

An established model-system for studying the folding of multidomain proteins is given by phosphoglycerate kinase from yeast (yPGK), a key enzyme in glycolysis. The 45-kDa yPGK is composed of two structurally homologous domains, the N- and the C-terminal domains. Both are connected by a well-conserved hinge region, as shown in Fig. 1. Each domain consists of a core of six parallel strands of $\beta$-pleated sheets surrounded by four $\alpha$-helices. The helices are connected to the central $\beta$-strands by $\beta$-turns and unstructured loop regions. The C-domain exhibits, in addition, a pair of anti-parallel $\beta$-strands toward the C-terminal end, whereas the last 12 residues of the C-terminal domain fold back into to the N-domain (21). Interestingly, yPGK fragments, which represent only one single domain (either the N- or the C-terminal one), can fold into native-like structures, even if they exist as isolated domains (22,23). In the past, chemical- as well as heat- and cold-induced unfolding of yPGK has been studied extensively using traditional bulk methods, including circular dichroism (CD), tryptophan fluorescence, stopped-flow fluorescence, and FRET (22–26). More recently, these studies were also extended by single-molecule measurement (13,27,28). To obtain a more detailed picture about the conformational changes of yPGK, we pursued our previous smFRET studies. For this purpose, we constructed additional mutants and measured the evolution of both intradomain and interdomain distances as a function of chemical denaturant concentration. These experiments allowed us to advance toward a more detailed understanding of the complexity of the multidomain protein-unfolding transitions.

## MATERIALS AND METHODS

### Protein expression, labeling, and purification

All cysteine double mutants of the gene of phosphoglycerate kinase from *Saccharomyces cerevisiae* were generated as described in (27). Briefly, the natural cysteine C97 was replaced by a serine, and a pair of cysteine

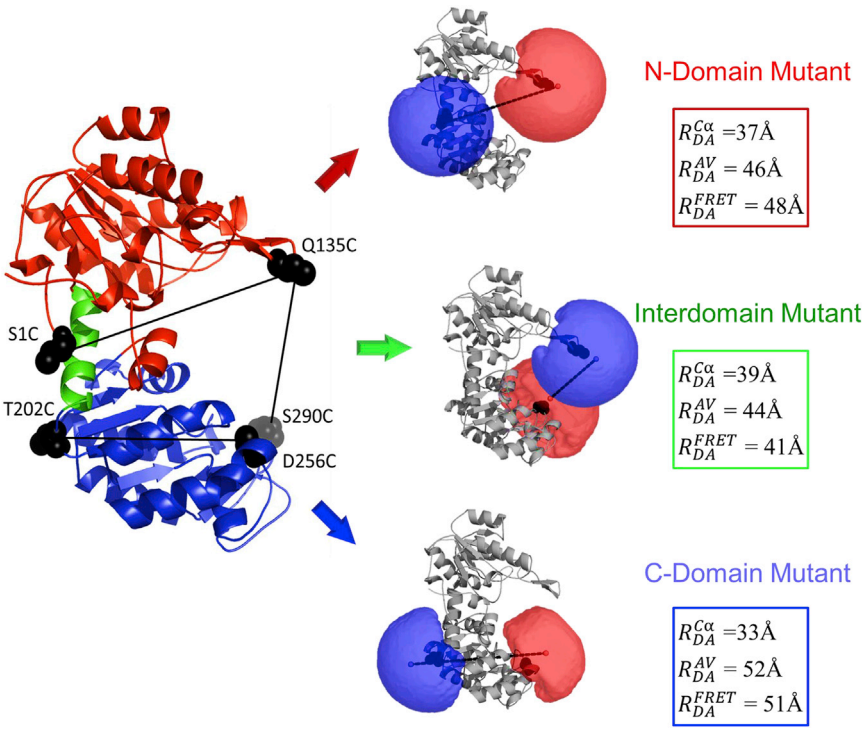


FIGURE 1 (*Left*) A 3D structure of yPGK (PDB: 1QPG) with the N-terminal domain (residues 1–185) highlighted in red, the C-terminal domain (residues 205–415) highlighted in blue, and the hinge region (residues 186–204) highlighted in green. Black spheres indicate positions at which natural amino acids are replaced by cysteines. In total, three different interdye distances are depicted: interdomain distance Q135C-S290C, intradomain distance in the N-terminal domain S1C-Q135C, and intradomain distance within the C-domain T202C-D256C. (*Right*) yPGK structures of the three mutants are shown with AVs (*red* and *blue* volumes) localized at respective dye attachment positions for the three mutants. For each mutant, the $C_\alpha$-distances of the mutated cysteines, the distance obtained from AV calculations, and the interdye distances obtained experimentally from the position of the smFRET peak are given. To see this figure in color, go online.

residues were incorporated by site-directed mutagenesis at corresponding positions; for details, see Table S1. Additionally, all corresponding single cysteine mutants were prepared for control measurements. The mutated genes were integrated into the pET15b expression system (Novagen; EMD Millipore, Merck KGaA, Darmstadt, Germany), thereby adding a hexa-histidine tag and a six-amino-acid-long thrombin cleavage site to the N-terminus of the protein sequence for purification purposes. All yPGK variants were expressed and purified from the cytoplasm of *Escherichia coli* BL21-CodonPlus(DE3)-RIL (Stratagene, La Jolla, CA) by NiNTA affinity chromatography. Far-ultraviolet CD spectroscopy and an enzymatic activity assay were performed to validate the secondary structure and the catalytic activity of the PGK variants (see Figs. S1 and S2; Table S4). After flash freezing in liquid nitrogen, proteins were stored at −80°C. The labeling strategy reported in (27,29) was followed. Briefly, a 10-$\mu$M PGK solution was incubated in a reducing environment at 4°C for 16 h with a fivefold excess of both the maleimide-functionalized dyes (C2-Alexa647 and C5-Alexa488; Sigma-Aldrich, St. Louis, MO) with a 1:1 concentration relation for both dyes. The labeled protein was separated from free dye by size exclusion chromatography using a Sephadex G25 medium (Sigma-Aldrich). After this purification step, the double-labeled protein fraction was isolated by anion exchange chromatography employing a MonoQ5/50GL (GE Healthcare Life Science, Uppsala, Sweden) on an automated high-performance liquid chromatography system (ÄKTA Explorer 10; GE Healthcare Life Science, Uppsala, Sweden). Finally, double-labeled PGK was stored at 4°C for a maximum of a few days before usage.

## CD

CD measurements were performed using a Jasco J-1100 spectropolarimeter equipped with a Peltier temperature controller (JASCO Deutschland GmbH, Pfungstadt, Germany). The spectra were acquired in the range 190–250 nm, and the background spectrum of the corresponding buffer was subtracted. Samples were prepared in 10 mM MOPS buffer at pH 7.5 with a nominal concentration of ~3 $\mu$M. All the measurements were performed at room temperature in a quartz cell having a pathlength of 1 mm.

## Tryptophan fluorescence

Fluorescence emission spectra were recorded with protein solutions (protein concentration: 0.05–0.1 mg/mL) in 5-mm path length quartz cuvettes (104F-QS; Hellma, Muehlheim, Germany) using a QuantaMaster40 spectrometer from Photon Technology International (Lawrenceville, NJ).

With excitation wavelengths of 295 nm, we obtained emission spectra that are dominated by tryptophan emission. All emission spectra (recorded between 300 and 450 nm) were corrected for background intensities as measured with pure buffer solutions. Unfolding transitions were analyzed by determining the wavelength of the emission intensity maximum ($\lambda_{max}$) as a function of GndHCl concentration (Fig. S1 *d*).

## smFRET measurements and data analysis

smFRET measurements were performed with a MicroTime 200 confocal microscope (PicoQuant, Berlin, Germany) equipped with a red (640 nm) and a blue (485 nm) diode laser (LDH-D-C 640B and LDH-D-C 485B; PicoQuant), and with a UPLSAPO 60×/1.2 NA water-immersion objective from Olympus (Hamburg, Germany). The emitted photons were collected by the microscope objective, passed through a dual-band dichroic mirror (XF2401; Omega Optical, Brattleboro, VT), focused on a 100 $\mu$m pinhole, and split into four detection channels by a further dichroic mirror (T600pxr; Chroma Technology, Bellows Falls, VT) and by two polarizing beam splitter cubes (Linos Photonics, Göttingen, Germany). The red photons were filtered by an appropriate emission filter (HQ 690/70M; Chroma Technology, Bellows Falls, VT) and focused on two SPAD detectors (SPCMCD3307M and SPCM-AGR-14; Perkin Elmer, Wood Bridge, Ontario, Canada). The blue photons were filtered by a bandpass emission filter (XF 3003 520DF40; Omega Optical, Battlesboro, VT) and detected by two silicon avalanche photodiodes ($\tau$-SPAD; PicoQuant). The lasers were operated at 20 MHz with pulsed interleaved excitation (30) by means of a computer-controlled PDL828 Sepia-II-Laser Driver (PicoQuant). The counts were processed with a HydraHarp-400 time-correlated single-photon counting acquisition unit (PicoQuant). smFRET measurements were performed with an average number of acceptor molecules within the detection spot of $<N>$ ~0.03. This condition was achieved by precisely measuring the concentration of the stock solutions by fluorescence correlation spectroscopy performed on the same setup. The yPGK buffer contained 50 mM MOPS, 50 mM NaCl, and 0.005% Tween20 (Sigma-Aldrich) at pH 7.5. For the unfolding experiments, specific amounts of GdnHCl were added. The actual concentrations of denaturant solutions were determined by refractive index measurements (31). Additionally, a photoprotection cocktail (1 mM Trolox, 10 mM cysteamine) was added to all solutions. In sample solutions with highly diluted protein molecules (a few picomolars), one can obtain energy transfer efficiencies for every single molecule by applying a burst analysis (32,33). The measured time traces give access to photon bursts originating from single molecules appearing as dips in the interphoton lag and could be selected by choosing a suitable threshold value (34). For each selected burst, donor and acceptor photon counts ($F_D$, $F_A$) after donor excitation were accumulated and corrected. For bursts with $F_D + F_A > 20$, we calculated the energy transfer efficiency

$$E = \frac{F_A}{F_A + \gamma F_D}, \tag{1}$$

for each burst, where $\gamma$ is a correction factor (for more details, see Table S3). The corrected E-values obtained for each burst were finally histogramed; depending on the number of available bursts, we chose ~20 bins for the whole range of E-values. By using the pulsed interleaved excitation scheme, it is possible to eliminate the donor-only peak, a typical artifact originated from proteins labeled only with the donor fluorophore (30). A detailed description of the data evaluation is given in (28). The obtained FRET histograms for yPGK molecules were fitted individually with one or two Gaussians, in which the obtained peak-positions of the Gaussians correspond to the mean FRET efficiencies $<E>$. The latter was used to determine the measured interdye distance (measured $R_{DA}$)

$$R_{DA} = \left(\frac{1 - <E>}{<E>}\right)^{1/6} \times R_0, \tag{2}$$

where $R_0$ represents the Förster radius. To achieve the most reliable interdye distances $R_{DA}$ from smFRET measurements, we explicitly considered the impact of high GndHCl concentrations in the employed buffer solutions on the donor and acceptor quantum yields. The latter were determined with corresponding single cysteine mutants in control measurements, according to a methodical approach described in (35) (see Table S2). Furthermore, we performed accessible volume (AV) calculations of the dyes tethered at a selected position within the yPGK structure (36). The resulting interdye distances (calculated $R_{DA}$) based on the 3D x-ray structure (Protein Data Bank, PDB: 1QPG; see (37)) and on the AV calculations were compared to those measured by smFRET with yPGK in the native state (see Table S4). All data analyses were performed with self-written MATLAB (R2015b, 64-bit; The MathWorks, Natick, MA) scripts or using OriginPro (9.0.0G, 64 bit).

# RESULTS

To map multiple inter- and intradomain distances during GndHCL-induced unfolding, we aimed to produce various

double cysteine mutants of yPGK. In total, nine double cysteine constructs were chosen for expression (see Table S1). After a scrupulous validation of the mutants in terms of conserved enzymatic activity, conserved structure, and avoidance of artifacts, we were able to perform successful smFRET measurements for five constructs, which are discussed in the following subsections in more detail (see Fig. S1; Table S1).

## Comparison of inter- and intradomain distances

We started our analysis with three double cysteine constructs that mapped one interdomain distance and two intradomain distances (Fig. 1). The global equilibrium stability of theses mutants against GndHCl was investigated by CD and fluorescence spectroscopy. The obtained data could be described by a two-state transition model, quantifying the equilibrium stability in terms of half-transition concentration $C_{1/2}$ (see Fig. S1; Table S4). The stability for all the cysteine double mutants is slightly lower as compared to that of the wild type, but the different mutants show a comparable behavior. For the investigated double cysteine mutants, transfer efficiency histograms were obtained from smFRET measurements, performed as a function of GndHCl concentrations (Fig. 2). As shown in the left panels of Fig. 2, the interdomain mutant (ID1) exhibits a classical two-state transition behavior, which we observed already in a previous smFRET study with the same mutant (27).

The native state and the unfolded state population are well separated in the histogram; upon unfolding, the native state population at $<E> \sim 0.75$ is progressively transferred to the unfolded state population (at $<E> \sim 0.2$–$0.35$). The half-transition concentration of GndHCl obtained from the single-molecule data agrees reasonably well with the $C_{1/2}$-value we obtained from ensemble measurements (see Fig. S3; Table S4).

Compared to the interdomain distance, both of the intradomain distances exhibit a clearly different behavior upon unfolding. As presented by the middle panels of Fig. 2, the histograms for intradomain distance measurements obtained from the N-domain mutant (N1), show, for native conditions, a single peak with a FRET efficiency of ~0.5, but an additional population at high FRET (representing, therefore, a compact folding intermediate) appears at concentrations of denaturant below the $C_{1/2}$ (~0.67 M). With increasing denaturant concentration, the statistical weight of the compact folding intermediate first increases, then it decreases again above 0.7 M GndHCl with a corresponding increase of a low FRET peak at $<E> \leq 0.3$, the latter corresponding to the unfolded state. Therefore, the high FRET peak appears to behave as an on-path intermediate. In comparison to the previous cases, the intradomain distance within the C-domain (C1) exhibits, at first glance, a stunningly different behavior. As shown in the corresponding histograms in Fig. 2 (*right panels*), in all the denaturing conditions, the population distribution is formed by a single peak, whose position stays at a rather constant transfer efficiency value of ~0.55 over the entire GndHCl concentration range between 0 and 1.2 M. A small, but continuous, shift of the unfolded state population toward small transfer efficiency values above 1.2 M indicates a well-known denaturant-induced coil-globule transition (28,38). Obviously, the behavior observed by monitoring these two specific intradomain distances does not show the classical two-state behavior. This deviation cannot be recognized from the global unfolding of secondary structure elements as observed by CD spectroscopy (see Fig. S1) because CD is a low-resolution technique sensible to the amount of secondary structure elements. However, a possible explanation might be given by the fact that, for this mutant, the number of amino acid residues between the two dye attachment positions is rather small (54, see Fig. 2). If we assume a random-coil structure of the amino acid chain under unfolding conditions, one can estimate the end-to-end distance of this chain by $\sqrt{<R^2>} = n \times l^2$, where l gives a persistence length (~0.65 nm) and n gives the number of amino acids (39). A <R>-value of ~48 Å obtained from this model is close to the interdye distance $R_{DA\_unf}$ of 51 Å measured for the unfolded state (see Table S4).

## Probing different intradomain distances within the C-domain

To investigate folding events occurring within individual domains in more detail, we intended to examine additional intradomain mutants. Out of six additional trials, two further mutants characterizing the C-domain were studied successfully (for details, see Fig. S1 *c*; Table S1). Compared to the C1 mutant described in the previous section, both additional intradomain distances show a clearly different behavior upon GndHCl-induced unfolding (Fig. 3). The native state of the first additional C2 mutant (see *left panels* in Fig. 3) shows a major peak around an energy transfer value of $<E> \sim 0.3$ and a pronounced shoulder toward higher E-values. With increasing denaturant concentration, the shoulder evolves into a more distinct peak at $<E> \sim 0.7$. At the same time, the major native peak population exhibits a continuing decrease in its statistical weight and a slight shift toward higher energy transfer values, a feature that is often observed because of salt-induced stabilization at low GndHCl concentrations. Above a denaturant concentration of ~0.7 M, the statistical weights of the population related to the peak at $<E> \sim 0.7$ decreases with a corresponding increase of a peak at $<E> \sim 0.3$, the latter representing the unfolded state. In this specific case, the position of the native state and that of the unfolded state are rather similar because the distance between the labeling positions is already large within the folded structure ($R_{DA}$ ~56 Å), being larger than the Förster radius for the employed dye couple ($R_0$ ~51–53 Å); therefore, a further increase in the distance due to

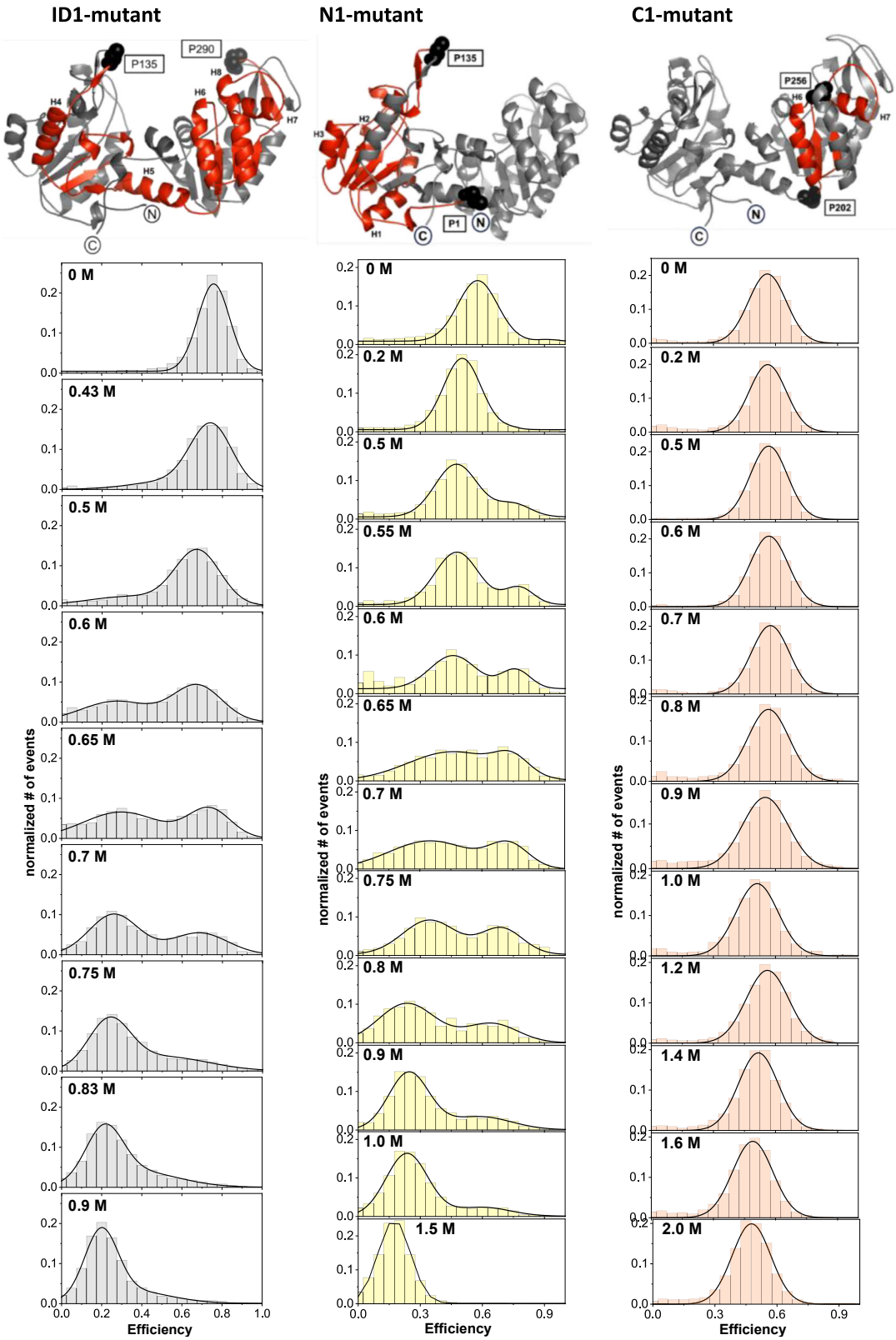


FIGURE 2 (*Upper graphs*) 3D structures of yPGK are shown with individual dye attachment positions for all three intermolecular FRET-pairs. Amino acids and related structural elements localized between the attachment positions in the amino acid sequence are highlighted in red. For the intradomain mutants, the enclosed structure elements (N1 mutant: 134 residues; C1 mutant: 54 residues) include N1 mutant with 4 $\beta$-sheets and 2 $\alpha$-helices and C1 mutant with 2 $\beta$-sheets and 2 $\alpha$-helices. (*Lower panels*) Transfer efficiency histograms for each measurement at a specific GndHCl concentration are shown. The solid lines over the histograms represent fits with one or two Gaussian peaks. To see this figure in color, go online.

unfolding results in only a small effect on the FRET efficiency (see also Table S4). The general behavior of the C2 mutant is, in part, similar to that observed for the intradomain distance of the N-domain (N1 mutant; see Fig. 2).

Finally, the other additional intradomain C-domain distance mutant (C3 mutant; see *right panels* in Fig. 3) again shows a classical two-state transition behavior, which was already observed for the interdomain mutant (ID1 mutant, Fig. 2). Here, the half-transition GndHCl concentration obtained from the smFRET data is slightly higher than that obtained for the C1 mutant from ensemble data (Table S4).

## DISCUSSION

Most of the previous smFRET studies on denaturant-induced unfolding exhibit bimodal FRET efficiency distributions. This is related to an unfolding process with two well-separated populations in the FRET efficiency histogram, the native state and the unfolded state. Such a behavior was observed mainly for single-domain proteins (see, for example, (40–43)) but also for some multidomain proteins, like for maltose-binding proteins (8) and rhodanase (44). In our study, only two mutants exhibit a clear two-state unfolding behavior. For the three intradomain distances measured within the C-domain, only the C3 mutant exhibits a two-state behavior (Fig. 3), whereas the C1 and the C2 mutants (Figs. 2 and 3) manifest a clearly different trend. We can, therefore, state that (at least for yPGK) whether a two-state behavior is observed depends not only on the investigated domain (or protein) but also on the attachment positions that are employed to follow the denaturant-induced conformational changes. In contrast to our findings, some earlier studies, in which multiple interdye distances were mapped within one and the same protein, reported that all denaturant-induced distance changes qualitatively exhibit the same classical two-state unfolding behavior (8,44,45).

To obtain a better understanding of the reason why the investigated intermolecular distances exhibit such a diverse behavior, we had a closer look at the compositions and the topology of secondary structure elements of yPGK. According to the structural classification of proteins, yPGK belongs to the class of $\alpha/\beta$-proteins (46). It is well known that many multidomain proteins belong to this class (20) (for example, the maltose-binding protein (44) and rhodanese (8), for which the obtained smFRET results differ from our findings, as already discussed above). However, a recent smFRET study on apoflavodoxin (47), a single-domain $\alpha/\beta$-protein, reports that GndHCl induced structural changes similar to those observed for the N1 mutant (Fig. 2, *middle panels*) and for the C2 mutant (Fig. 3, *left panels*). In all these cases, the corresponding FRET histograms exhibit an additional population that shows up at <E>-values larger than that of the native state. This population appears at rather low denaturant concentrations and continuously diminishes above the 0.7 M GndHCl, approximately the value of $C_{1/2}$. In the case of apoflavodoxin, the authors attribute this additional population to a compact molten globule (MG) state, which represents, for this protein, a misfolded off-pathway intermediate (47). In general, MGs are compact and exhibit some native-like secondary structure but lack a native-like tertiary structure. They can exist as stable species in substantial portions of molecules under mildly denaturing conditions, such as low GndHCl concentrations (48). Strikingly, our data and that of apoflavodoxin revealed an MG state characterized by shorter intramolecular distances and, by this, a more compact state as compared to the native state. According to the best of our knowledge, potential experimental artifacts that would lead to the rather small observed interdye distances at unfolding conditions (for example, caused by folding-dependent changes of the donor/acceptor quantum yield or by other dye-related effects) can be ruled out (see Fig. S4; Table S2). Furthermore, it is worth mentioning that a previous kinetic refolding study with yPGK reported on an MG state that was formed in the early folding phase (49).

Finally, we will focus on the question of why specific properties of the unfolding transition (e.g., number of observable states) do not show up in the same manner in all corresponding FRET histograms, for example, in those histograms that were obtained from all three C-domain intradomain distances. To elucidate this point, it is worth examining the sequence of amino acids in between both label positions. In particular, the length of this sequence and the related secondary structure elements that are formed by this sequence (*red-highlighted structure elements* in Figs. 2 and 3) may contribute, to some extent, to the folding and unfolding behavior of the corresponding mutant. Compared to all other mutants, the interdomain ID1-mutant exhibits the longest sequence and also the largest structural element sandwiched between both dyes. However, it can be assumed that, for this mutant, mainly the small hinge region (*green-colored structure element sandwiched between both domains of the yPGK structure*; see *left graph* in Fig. 1) controls the properties of the unfolding and folding transition (27,29). Because this hinge region is essentially formed only by one connecting $\alpha$-helix, the unfolding of this helix can already cause the observed transition. Out of all the mutants analyzed, the one having the smallest structural element enclosed between the two labeling positions is the C3 mutant. It includes only 34 residues and is mainly located close to the surface of the C-domain (see Fig. 3). Because this one and the interdomain mutant are the only ones that show a two-state unfolding behavior, we can assume that small (enclosed and not very heavily interacting with neighboring structure elements) portions of secondary

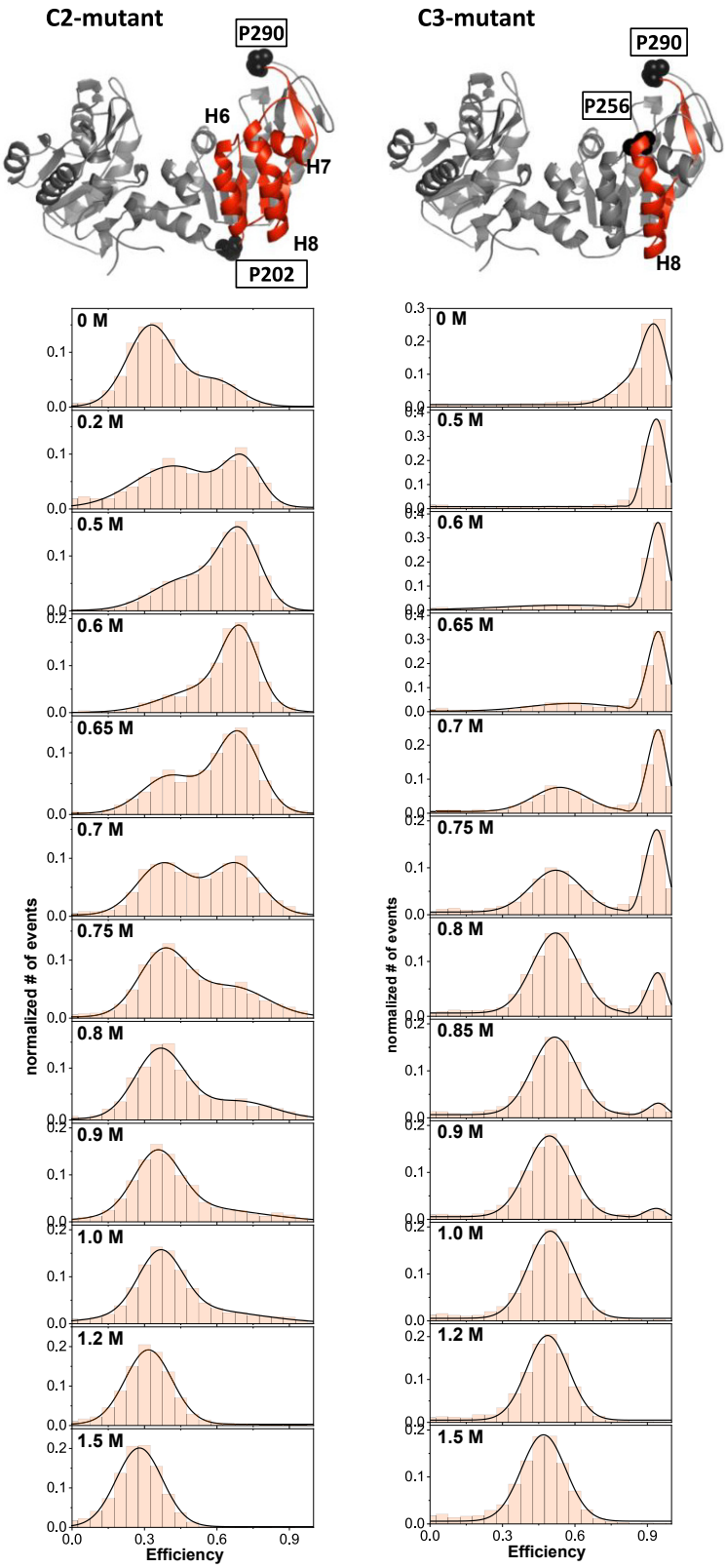


FIGURE 3 Properties of two further intradomain distances within the C-domain. The interdye distance of the C2 mutant is given by a measured $R_{DA} = 57$ Å, whereas that of the C3 mutant yields a measured $R_{DA} = 31$ Å. Structure elements enclosed between the dye attachment positions are highlighted in red (C2 mutant: 88 residues; C3 mutant: 34 residues) and includes C2 mutant with 4 $\beta$-sheets and 3 $\alpha$-helices and C3 mutant with 2 $\beta$-sheets and 1 $\alpha$-helix. To see this figure in color, go online.

structures are related to unfolding transitions characterized by a simple two-state mechanism. According to the polymer chain model, the rather short sequence of 34 amino acids between the dye attachment positions has also a small end-to-end distance of $<R> \sim 38$ Å. However, the measured interdye distance for this C3-mutant in

the unfolded state is significantly larger ($R_{DA\_unf}$ ~54 Å). In contrast, the N1 and C2 mutants are characterized by rather large enclosed structural elements with 134 and 88 residues, respectively. The respective end-to-end distances and their corresponding $R_{DA\_unf}$-values agree much better than those for the C3 mutant (see Table S4). Here, in both cases, the enclosed elements are localized in the domain center and have more pronounced interactions with neighboring structure elements, the latter of which is not part of the enclosed structure. Because in these cases, more potential intramolecular interactions are involved, the observed unfolding process appears more complex, and corresponding FRET histograms do not only show two well-separated states but also the appearance of an additional state (MG) and shifts of population peaks upon changing GndHCl concentrations. Each histogram of these mutants was also fitted with two Gaussians, but an unambiguous assignment of the two observed populations to the native, the unfolded, or the intermediate state is not straight forward (Fig. S3). The obtained fittings of the histograms indicate that one population is related to the intermediate state (see $Pop_1$ in Fig. S3), whereas the second population peak ($Pop_2$) represents different states depending on the GndHCl concentration regime: at low concentrations ($\leq$0.6–07 M), it represents the native state, whereas it represents the unfolded state at the high GndHCl concentrations ($\geq$0.6–0.7 M), in which additional pronounced gradual shifts (see *right panels* in Fig. S3) with increasing denaturant concentrations are observed. Therefore, for these two mutants, more than two states are involved, and the $Pop_2$ populations and a distinct assignment of each population to one corresponding state are not possible; in turn, a simple fit to a two-state unfolding transition is not possible. However, the curves of the relative population fractions as a function of denaturant concentration (see *left panels* in Fig. S3) qualitatively show inflection points at values close to the half-transition concentration values obtained by CD spectroscopy (0.65 and 0.6 M for the N1- and the C2-mutants, respectively), confirming that the behavior observed at the single-molecule level reproduces the global behavior. In a further case, the C1 mutant shows practically no change of the interdye distance over the whole range of GndHCl concentrations (at least up to 1.2 M GndHCl). This behavior is rather unique and was already explained in the Results. Because this mutant includes an enclosed structure element that represents a significant part of that from the C2 mutant (the residues 256–290 are missing), we would expect a similar development of FRET histograms like that of the C2 mutant. However, for the C1 mutant, a potential intermediate state population might be hidden in the measured histograms, probably because of the fact that all occurring states appear at transfer efficiency values that are very close to each other.

## CONCLUSIONS

Our approach to investigating denaturant-induced unfolded states of yPGK by mapping multiple inter- and intradomain distances highlights a rather complex unfolding behavior of this two-domain protein. In addition to the well-known two-state unfolding behavior, as detected by ensemble techniques like CD spectroscopy and also by smFRET for the interdomain mutant (27), we observed for intradomain mutants additional features differentiating these mutants from classical two-state transition. The data indicate that a compact MG intermediate state is formed transiently in each of the domains. Furthermore, our study clearly demonstrated that the choice of dye attachment positions in the polypeptide chain allows us to select which unfolding transition features are observed. In the case of complex protein folding, measuring different intermolecular distances gives us access to different unfolding and folding features because different distances allow us to slice the energy landscape along different directions and to sample the energy funnel at different depths. This may not only hold for multidomain proteins but also for single-domain proteins because pronounced differences in the FRET histograms appear already when several intradomain distances are mapped within a single domain. So, the complex folding and unfolding features within one domain may be inherent to this domain, or they may be caused by the presence of a neighboring (unfolded) domain. However, the approach of mapping multiple intermolecular distances seems to be a key to discover hidden intermediates during folding and unfolding transitions.

## SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2019.12.006.

## AUTHOR CONTRIBUTIONS

J.F., M.C., and M.G. conceived the idea and designed the experiments. A.S. and I.R. produced and characterized the samples. M.G. and M.C. developed smFRET data analysis tools. M.C. performed the smFRET measurements and evaluated the data. J.F., M.C., and M.G. wrote the manuscript with input from all authors.

## ACKNOWLEDGMENTS

The authors thank Dr. Tobias Rosenkranz, Dr. Alexander Katranidis, Henning Höfig, and Olessya Yukhnovets for support during this project.

This work was supported by the International Helmholtz Research School on Biophysics and Soft Matter (BioSoft) (to A.S.).

## REFERENCES

1. Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes. 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:545–600.

2. Zhuang, X., and M. Rief. 2003. Single-molecule folding. *Curr. Opin. Struct. Biol.* 13:88–97.
3. Nienhaus, G. U. 2006. Exploring protein structure and dynamics under denaturing conditions by single-molecule FRET analysis. *Macromol. Biosci.* 6:907–922.
4. Schuler, B., and W. A. Eaton. 2008. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* 18:16–26.
5. Borgia, A., P. M. Williams, and J. Clarke. 2008. Single-molecule studies of protein folding. *Annu. Rev. Biochem.* 77:101–125.
6. Zoldák, G., and M. Rief. 2013. Force as a single molecule probe of multidimensional protein energy landscapes. *Curr. Opin. Struct. Biol.* 23:48–57.
7. Chung, H. S., and W. A. Eaton. 2018. Protein folding transition path times from single molecule FRET. *Curr. Opin. Struct. Biol.* 48:30–39.
8. Sharma, S., K. Chakraborty, …, F. U. Hartl. 2008. Monitoring protein conformation along the pathway of chaperonin-assisted folding. *Cell.* 133:142–153.
9. Hofmann, H., F. Hillger, …, B. Schuler. 2010. Single-molecule spectroscopy of protein folding in a chaperonin cage. *Proc. Natl. Acad. Sci. USA.* 107:11793–11798.
10. Pirchi, M., G. Ziv, …, G. Haran. 2011. Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein. *Nat. Commun.* 2:493.
11. Stigler, J., F. Ziegler, …, M. Rief. 2011. The complex folding network of single calmodulin molecules. *Science.* 334:512–516.
12. Kantaev, R., I. Riven, …, G. Haran. 2018. Manipulating the folding landscape of a multidomain protein. *J. Phys. Chem. B.* 122:11030–11038.
13. Li, Q., Z. N. Scholl, and P. E. Marszalek. 2018. Unraveling the mechanical unfolding pathways of a multidomain protein: phosphoglycerate kinase. *Biophys. J.* 115:46–58.
14. Dimura, M., T. O. Peulen, …, C. A. Seidel. 2016. Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. *Curr. Opin. Struct. Biol.* 40:163–185.
15. Hellenkamp, B., P. Wortmann, …, T. Hugel. 2017. Multidomain structure and correlated dynamics determined by self-consistent FRET networks. *Nat. Methods.* 14:174–180.
16. Hartl, F. U., and M. Hayer-Hartl. 2009. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* 16:574–581.
17. Fitter, J. 2009. The perspectives of studying multi-domain protein folding. *Cell. Mol. Life Sci.* 66:1672–1681.
18. Osváth, S., M. Jäckel, …, J. Fidy. 2006. Domain interactions direct misfolding and amyloid formation of yeast phosphoglycerate kinase. *Proteins.* 62:909–917.
19. Tian, P., and R. B. Best. 2016. Structural determinants of misfolding in multidomain proteins. *PLoS Comput. Biol.* 12:e1004933.
20. Han, J. H., S. Batey, …, J. Clarke. 2007. The folding and evolution of multidomain proteins. *Nat. Rev. Mol. Cell Biol.* 8:319–330.
21. Watson, H. C., N. P. Walker, …, M. F. Tuite. 1982. Sequence and structure of yeast phosphoglycerate kinase. *EMBO J.* 1:1635–1640.
22. Osváth, S., G. Köhler, …, J. Fidy. 2005. Asymmetric effect of domain interactions on the kinetics of folding in yeast phosphoglycerate kinase. *Protein Sci.* 14:1609–1616.
23. Missiakas, D., J. M. Betton, …, J. M. Yon. 1990. Unfolding-refolding of the domains in yeast phosphoglycerate kinase: comparison with the isolated engineered domains. *Biochemistry.* 29:8683–8689.
24. Nojima, H., A. Ikai, …, H. Noda. 1977. Reversible thermal unfolding of thermostable phosphoglycerate kinase. Thermostability associated with mean zero enthalpy change. *J. Mol. Biol.* 116:429–442.
25. Lillo, M. P., J. M. Beechem, …, M. T. Mas. 1997. Design and characterization of a multisite fluorescence energy-transfer system for protein folding studies: a steady-state and time-resolved study of yeast phosphoglycerate kinase. *Biochemistry.* 36:11261–11272.
26. Szpikowska, B. K., J. M. Beechem, …, M. T. Mas. 1994. Equilibrium unfolding of yeast phosphoglycerate kinase and its mutants lacking one or both native tryptophans: a circular dichroism and steady-state and time-resolved fluorescence study. *Biochemistry.* 33:2217–2225.
27. Rosenkranz, T., R. Schlesinger, …, J. Fitter. 2011. Native and unfolded states of phosphoglycerate kinase studied by single-molecule FRET. *Chemphyschem.* 12:704–710.
28. Kempe, D., M. Cerminara, …, J. Fitter. 2017. Single-molecule FRET measurements in additive-enriched aqueous solutions. *Anal. Chem.* 89:694–702.
29. Gabba, M., S. Poblete, …, J. Fitter. 2014. Conformational state distributions and catalytically relevant dynamics of a hinge-bending enzyme studied by single-molecule FRET and a coarse-grained simulation. *Biophys. J.* 107:1913–1923.
30. Müller, B. K., E. Zaychikov, …, D. C. Lamb. 2005. Pulsed interleaved excitation. *Biophys. J.* 89:3508–3522.
31. Pace, C. N. 1986. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* 131:266–280.
32. Eggeling, C., S. Berger, …, C. A. Seidel. 2001. Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *J. Biotechnol.* 86:163–180.
33. Deniz, A. A., T. A. Laurence, …, S. Weiss. 2001. Ratiometric single-molecule studies of freely diffusing biomolecules. *Annu. Rev. Phys. Chem.* 52:233–253.
34. Fries, J. R., L. Brand, …, C. A. M. Seidel. 1998. Quantitative identification of different single molecules by selective time-resolved confocal fluorescence spectroscopy. *J. Phys. Chem. A.* 102:6601–6613.
35. Kempe, D., A. Schöne, …, M. Gabba. 2015. Accurate fluorescence quantum yield determination by fluorescence correlation spectroscopy. *J. Phys. Chem. B.* 119:4668–4672.
36. Höfig, H., M. Gabba, …, J. Fitter. 2014. Inter-dye distance distributions studied by a combination of single-molecule FRET-filtered lifetime measurements and a weighted accessible volume (wAV) algorithm. *Molecules.* 19:19269–19291.
37. McPhillips, T. M., B. T. Hsu, …, D. C. Rees. 1996. Structure of the R65Q mutant of yeast 3-phosphoglycerate kinase complexed with Mg-AMP-PNP and 3-phospho-D-glycerate. *Biochemistry.* 35:4118–4127.
38. Aznauryan, M., L. Delgado, …, B. Schuler. 2016. Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS. *Proc. Natl. Acad. Sci. USA.* 113:E5389–E5398.
39. Lapidus, L. J., P. J. Steinbach, …, J. Hofrichter. 2002. Effects of chain stiffness on the dynamics of loop formation in polypeptides. Appendix: testing a 1-dimensional diffusion model for peptide dynamics. *J. Phys. Chem. B.* 106:11628–11640.
40. Deniz, A. A., T. A. Laurence, …, S. Weiss. 2000. Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. USA.* 97:5179–5184.
41. Schuler, B., E. A. Lipman, and W. A. Eaton. 2002. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature.* 419:743–747.
42. Kuzmenkina, E. V., C. D. Heyes, and G. U. Nienhaus. 2005. Single-molecule Forster resonance energy transfer study of protein dynamics under denaturing conditions. *Proc. Natl. Acad. Sci. USA.* 102:15471–15476.
43. Haran, G. 2003. Single-molecule fluorescence spectroscopy of biomolecular folding. *J. Phys. Condens. Matter.* 15:R1291–R1317.

44. Hillger, F., D. Hänni, …, B. Schuler. 2008. Probing protein-chaperone interactions with single-molecule fluorescence spectroscopy. *Angew. Chem. Int.Engl*. 47:6184–6188.

45. Hoffmann, A., A. Kane, …, B. Schuler. 2007. Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. USA*. 104:105–110.

46. Murzin, A. G., S. E. Brenner, …, C. Chothia. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247:536–540.

47. Lindhoud, S., M. Pirchi, …, C. P. van Mierlo. 2015. Gradual folding of an off-pathway molten globule detected at the single-molecule level. *J. Mol. Biol.* 427:3148–3157.

48. Creighton, T. E. 1997. How important is the molten globule for correct protein folding? *Trends Biochem. Sci.* 22:6–10.

49. Osváth, S., L. Herényi, …, G. Köhler. 2006. Hierarchic finite level energy landscape model: to describe the refolding kinetics of phosphoglycerate kinase. *J. Biol. Chem.* 281:24375–24380.