Hypothesis: structural heterogeneity of the unfolded proteins originating from the coupling of the local clusters and the long-range distance distribution

Satoshi Takahashi^{1,2} · Aya Yoshida^{1,2} · Hiroyuki Oikawa^{1,2}

Received: 8 December 2017 / Accepted: 5 February 2018 / Published online: 14 February 2018

© International Union for Pure and Applied Biophysics (IUPAB) and Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

We propose a hypothesis that explains two apparently contradicting observations for the heterogeneity of the unfolded proteins. First, the line confocal method of the single-molecule Förster resonance energy transfer (sm-FRET) spectroscopy revealed that the unfolded proteins possess broad peaks in the FRET efficiency plot, implying the significant heterogeneity that lasts longer than milliseconds. Second, the fluorescence correlation method demonstrated that the unfolded proteins fluctuate in the time scale shorter than 100 ns. To formulate the hypothesis, we first summarize the recent consensus for the structure and dynamics of the unfolded proteins. We next discuss the conventional method of the sm-FRET spectroscopy and its limitations for the analysis of the unfolded proteins, followed by the advantages of the line confocal method that revealed the heterogeneity. Finally, we propose that the structural heterogeneity formed by the local clustering of hydrophobic residues modulates the distribution of the long-range distance between the labeled chromophores, resulting in the broadening of the peak in the FRET efficiency plot. A clustering of hydrophobic residues around the chromophore might further contribute to the broadening. The proposed clusters are important for the understanding of protein folding mechanism.

Keywords Sm-FRET spectroscopy · Protein folding · Dynamics and heterogeneity of the unfolded proteins

Introduction

Heterogeneity of unfolded proteins is a puzzling concept. By definition, the unfolded state of proteins does not possess fixed native structures, and is composed of heterogeneous mixtures of different conformations of polypeptides. Surprisingly, however, the extent of the heterogeneity of the unfolded protein can hardly be detected directly by standard experimental methods. This is because the ensemble averaging required for

conventional experimental methods usually vanishes the information of the structural heterogeneity of proteins. Even in single-molecule measurements that do not require the ensemble averaging, the very fast conformational fluctuation of the unfolded protein averages observable over the time required for the single molecule measurements used, and gives time-averaged structural parameters for the unfolded state. Accordingly, the unfolded proteins have been assumed as the state possessing heterogeneous conformations theoretically but behaving as a homogeneous ensemble practically.

Despite the difficulty explained above, the experimental characterization of the structural heterogeneity of the unfolded polypeptides and how the heterogeneity is reduced in each of the kinetic steps are the central information required to understand protein folding. Protein folding is the process by which unfolded polypeptide chains attain the ordered three-dimensional structures (Abaskharon and Gai 2016; Sosnick and Barrick 2011; Dill and MacCallum 2012; Udgaonkar 2013; Takahashi et al. 2016; Schuler et al. 2016). For many proteins smaller than 100 residues, the unfolded state can kinetically convert to the folded structures rapidly in the cooperative two-state manner. The

This article is part of a Special Issue on “Biomolecules to Bio-nanomachines-Fumio Arisaka 70th Birthday” edited by Damien Hall, Junichi Takagi and Haruki Nakamura

✉ Satoshi Takahashi
satoshi.takahashi.a6@tohoku.ac.jp

¹ Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, Katahira 2-1-1, Aoba-ku, Sendai 980-8577, Japan

² Department of Chemistry, Graduate school of Science, Tohoku University, Aramaki 6-3, Aoba-ku, Sendai 980-8578, Japan

significant heterogeneity of the unfolded state contrasts to the structural specificity of the folded state, and should oppose to the rapidity and cooperativity of the kinetic folding transitions. It is important to develop methods that can characterize the heterogeneity of proteins in the unfolded state as well as in the folding intermediate state.

To detect heterogeneity of the unfolded proteins, we recently developed a new method of single-molecule Förster resonance energy transfer (sm-FRET) spectroscopy. We were successful in increasing the time resolution of the sm-FRET measurements to less than 100 μ s and in prolonging the observation to more than 5 ms, and tried to differentiate the transient conformations in the unfolded proteins that might be time averaged in the previous measurements conducted at the time resolution of ~ 1 ms. We could detect apparent heterogeneity in the structure of the unfolded proteins that persists, to our surprise, in the time scale longer than milliseconds (Oikawa et al. 2013; Oikawa et al. 2015; Saito et al. 2016). The observations were confirmed at least for two protein systems and might be consistent with several single molecule results for other proteins (Kuzmenkina et al. 2005; Tezuka-Kawakami et al. 2006; Merchant et al. 2007). In fact, very early investigations of the unfolded proteins by the single-molecule FRET method might even be consistent with our results (Deniz et al. 2000; Schuler et al. 2002). However, the observed heterogeneity apparently contradicts to the recent consensus of the unfolded proteins, in which the fluctuations of polypeptides occur rapidly in the time domain shorter than 100 ns (Schuler et al. 2016). It can be argued that the observed heterogeneity might be an artifact caused by the labeling of two chromophores to proteins in the sm-FRET measurements; however, the data showing the very fast fluctuations of the unfolded proteins were obtained for the same double-labeled proteins. To understand the sm-FRET results for the unfolded proteins and to reveal their structural heterogeneity, it is necessary to propose a comprehensive explanation of the apparently contradicting observations.

In this short review, we will first summarize the recent consensus on the properties of the unfolded proteins including the very fast structural fluctuations. We will next explain the basic principles of the sm-FRET measurements, and our recent observations demonstrating the significant heterogeneity of the unfolded state. Finally, we will propose a hypothesis in which the coupling of the local structural heterogeneity and the long-range distance distribution of the unfolded polypeptide contribute to the observed heterogeneity in the sm-FRET results. We will further discuss implications of the hypothesis to the understanding of protein folding mechanism.

Recent consensus for properties of the unfolded proteins

The unfolded state of proteins has been extensively investigated in recent years. We summarize an emerging consensus

on the properties of the unfolded state of foldable proteins in the order of size, residual structures, heterogeneity, and dynamics. We limit our attention mainly to foldable globular proteins having less than 100 residues, which usually demonstrate the perfect two-state transition. The consensus is that the unfolded proteins are random coils, possess significant amount of residual structures, and fluctuate rapidly in the time scales of several tens of nanoseconds.

The most important observation for the size of the unfolded proteins is the scaling relationship between the size and chain length, indicating that the unfolded proteins in general can be categorized as self-avoiding random coils (Kohn et al. 2004). In the presence of high concentrations of denaturants, the radius of gyration (R_g) for the unfolded proteins determined by small angle X-ray scattering (SAXS) was demonstrated to obey the Flory scaling relationship:

$$R_g = R_0 N^\nu \quad (1)$$

where R_0 is the prefactor, N is the number of amino acid residues, and ν is the scaling factor estimated experimentally as 0.598 ± 0.028 . The scaling factor near 0.6 suggests that the chemically unfolded proteins behave as random coils that take into account of the avoidance of overlap of chains in the same space. The observation suggests the absence of long-range interactions; however, it does not exclude the presence of local structures formed by the contacts between the residues close along the amino acid sequences (Fitzkee and Rose 2004). In the absence of the denaturants, the chain might shrink as suggested in the sm-FRET measurements (Sherman and Haran 2006; Haran 2012; Hofmann et al. 2012; Schuler et al. 2016). The collapse in the absence of denaturants contradicts to the results for the time-resolved SAXS measurements, in which the expanded dimensions similar to that obtained in the presence of denaturants were detected (Plaxco et al. 1999; Konuma et al. 2011; Yoo et al. 2012). The discrepancy in the compactness detected by the SAXS and sm-FRET results has been the focus of intensive discussion (Watkins et al. 2015; Borgia et al. 2016; Aznauryan et al. 2016; Fuertes et al. 2017; Riback et al. 2017). Some investigations suggested that the deviation of polypeptides from the Gaussian chain, which is a simple random coil allowing the overlap of chains in the same space and is frequently assumed to analyze the FRET efficiency data, might cause the apparent discrepancy (Maity and Reddy 2016; Song et al. 2017). We will return to this topic below.

The secondary structure content for the unfolded proteins in the presence of denaturant is minimal; however, a small amount of residual structures are usually detected by various spectroscopic methods (Bowler 2012; Jensen et al. 2014). The self-avoiding statistical coil model of unfolded proteins, which assumes the propensities of the backbone dihedral angles of amino acids for the unfolded proteins are the same as

that sampled from the PDB data in non- α , non- β conformations, can reproduce the radius of gyration determined by SAXS measurements and the residual dipolar coupling (RDC) data obtained by NMR spectroscopy (Jha et al. 2005; Bernado et al. 2005). The statistical coil model assumes no long-range interactions in the unfolded proteins based on the scaling relationship observed by SAXS measurements (Kohn et al. 2004). The model is a good first order approximation for the unfolded proteins; however, the incorporation of the long-range contacts deduced from the paramagnetic resonance enhancement (PRE) measurements is known to increase the local residual structures. In the case for the unfolded state of ubiquitin (76 residues), the RDC measurements detected that the first β -hairpin keeps the native-like structure (Meier et al. 2007). The combined use of the RDC and PRE data suggested that ca. 20% of the first β -hairpin as well as 10–15% of helix might be formed in the unfolded state of ubiquitin (Huang and Grzesiek 2010). Importantly, the detected local structures are fluctuating and are usually averaged in the time scale of the NMR measurements ($\sim$ milliseconds). Thus, the residual local structures are significantly populated for the chemically unfolded proteins but are fluctuating rapidly.

If we reduce the concentration of the denaturant, the residual structures sometimes become extensive and involve long-range contacts. The kinetic investigation of the loop formation and breakage in the denatured state of cytochrome *c*' (125 residues) demonstrated that the collapsed form of the unfolded ensemble likely possesses the native like clusters of hydrophobic residues (Dar et al. 2011). The PRE measurements for the destabilized mutant for the N-terminal domain of ribosomal protein L9 (NTL9, 56 residues) demonstrated that the significant hydrophobic cluster is present in the absence of denaturant, whose location might be consistent with the regions of residues having higher scores of the average area buried upon folding (AABUF) (Meng et al. 2013). AABUF reflects the hydrophobicity scale along the polypeptide chain and was proposed to determine the structure of the initial collapsed intermediate of larger proteins such as apomyoglobin (153 residues) (Felitsky et al. 2008) and β -lactoglobulin (162 residues) (Sakurai et al. 2017). The thermodynamic stability of the non-local hydrophobic cluster present in the unfolded state of NTL9 was further analyzed by the careful mutant cycles (Cho et al. 2014).

The loss of conformational entropy of polypeptides upon the folding transition is related to the heterogeneity of the unfolded state ensemble. Since the quantity is difficult to evaluate by thermodynamic experiments, the estimation requires the accurate modeling of both the native and unfolded states of proteins. Based on the self-avoiding statistical coil model of the unfolded proteins (Jha et al. 2005), Sosnick and his collaborators developed a method to estimate the loss of conformational entropy upon the folding (Baxa et al. 2014). They modeled the unfolded state ensemble based on the molecular

dynamics calculations with constraints of the coil libraries. For ubiquitin, the total loss of the entropy upon the folding was estimated $T\Delta S_{\text{Total}} = 1.4 \text{ kcal}\cdot\text{mol}^{-1}$ per residue at 300 K. The backbone ($T\Delta S_{\text{BB}}$) and side chain ($T\Delta S_{\text{SC}}$) contributions of the entropy loss were 1.0–1.1 and 0.2–0.3 $\text{kcal}\cdot\text{mol}^{-1}$, respectively. The backbone contribution, $T\Delta S_{\text{BB}}$, is secondary structure dependent, and is 1.5 and 1.0 $\text{kcal}\cdot\text{mol}^{-1}$ for helical and sheet residues, respectively. The larger values for the helical residues reflect that the unfolded dihedral angles are mainly in the extended regions as assumed in the coil library.

The large-scale dynamics of the unfolded polypeptides involving the collision of two residues occurs rapidly as demonstrated by several techniques. Early investigations based on the fluorescence quenching in the unfolded state of ribonuclease A, the loop formation kinetics of the unfolded cytochrome *c*, and the quenching of tryptophan triplet state by cysteine in model peptides estimated the intraresidue diffusion coefficients (D) of around $4 \times 10^{-7} \text{ cm}^2/\text{s}$ (Buckler et al. 1995; Hagen et al. 1996; Lapidus et al. 2000). The diffusion coefficient was based on the characteristic time constant (34–40 μs) for the collision of two residues separated by 62 residues (Hagen et al. 1996). In later investigations, somewhat different values of D ranging from 3×10^{-7} to $6 \times 10^{-6} \text{ cm}^2/\text{s}$ were reported (Moglich et al. 2006; Buscaglia et al. 2006; Soranno et al. 2009). For example, the analysis of the fluorescence lifetime in model peptides gave the D values of $4.9 \pm 0.2 \times 10^{-6} \text{ cm}^2/\text{s}$ in water solution and $5.8 \pm 0.5 \times 10^{-6} \text{ cm}^2/\text{s}$ in the presence of 8 M guanidinium (Moglich et al. 2006). The molecular dynamics calculation for the unfolded state of ubiquitin in the presence of 8-M urea calibrated by the PRE data demonstrated the segmental diffusion coefficients nearly uniform along the peptide chain averaging to $D = 4.9\text{--}5.5 \times 10^{-7} \text{ cm}^2/\text{s}$ (Xue and Skrynnikov 2011). Among the range of the coefficients, the values in the larger side was reported by using proteins doubly labeled by the donor and acceptor chromophores and by examining the cross correlation of donor and acceptor fluorescence intensities in the nanosecond time domain (ns-FRET correlation method). In the case for cold shock protein (66 residues) labeled at the two termini, the time scale causing the fluorescence intensity changes (reconfiguration time) was 20 ns in the presence of 8 M guanidinium and 65 ns near the native solution, corresponding to diffusion coefficient of 5×10^{-6} and $1 \times 10^{-6} \text{ cm}^2/\text{s}$, respectively (Nettels et al. 2007). Later results by the same group were consistent with the fast dynamics of the unfolded polypeptides (Borgia et al. 2012; Sorraro et al. 2012; Schuler et al. 2016). One report suggested an extremely small diffusion coefficient at $\sim 1 \times 10^{-9} \text{ cm}^2/\text{s}$ for the denatured protein L (64 residues) in the absence of denaturant based on the quenching of tryptophan triplet state by cysteine separated by ten residues around $\sim 10 \mu\text{s}$ (Waldauer et al. 2010); however, the recent investigation for protein L based on the ns-FRET correlation suggested the reconfiguration time shorter

than 100 ns in the absence of denaturant (Sorrano et al. 2017). Thus, the current consensus for the time scale of the chain fluctuation involving the significant changes in the distance between the two residues of polypeptides occur rapidly in the time scale within 100 ns.

In summary, the extensive investigations lead to an emerging view for the properties of the unfolded proteins. The unfolded proteins largely behave as random coils that take into account of the avoidance of overlap of chains in the same space, in which no stable contacts are assumed to present. A certain amount of residual structures are identified to present in the unfolded state. In addition, there still remains controversy in the compactness for the unfolded state in the absence of denaturants. The very fast conformational fluctuations were conformed. Accordingly, the self-avoiding random coil with some rapidly fluctuating local structures is the accepted view for the unfolded proteins.

Single-molecule FRET investigations for the unfolded state of proteins

The sm-FRET spectroscopy is based on the labeling of two fluorescent chromophores to proteins, and on the detection of the fluorescent photons from the two chromophores separately at the single molecule level. The excited singlet state of the donor, formed upon the laser excitation, might emit a fluorescent photon or transfers the excitation energy to the acceptor in the FRET process, followed by the emission of an acceptor fluorescence photon. By obtaining the fluorescence intensities of the donor (F_D) and acceptor (F_A) chromophores emitted from single proteins, the FRET efficiency, E , can be determined by Eq. (2):

$$E = \frac{F_A}{F_A + \gamma F_D} \quad (2)$$

where γ is a factor determined by the ratio of the detection efficiencies of the donor and acceptor photons for the optical system utilized, and by the ratio of the fluorescence quantum efficiencies of the chromophores. If the rotational dynamics of the two dyes is faster than the time scale of the changes in the distance between the donor and acceptor chromophores R , the efficiency E becomes a function of a single variable R as described in Eq. (3):

$$E(R) = \left[1 + \left(\frac{R}{R_0} \right)^6 \right]^{-1} \quad (3)$$

where R_0 is a constant called Förster distance determined by the overlap integral of the donor fluorescence and acceptor absorbance spectra. Note that if the time scale of the rotational dynamics of the chromophores is comparable to that of the

changes in R , R_0 becomes the function of the angles describing orientations of transition dipoles of the chromophores, making E dependent on R as well as the relative orientations of the chromophores. The method has been widely utilized to investigate the structural and dynamical properties of biological macromolecules.

It is important to point out the limitations of the single-molecule FRET spectroscopy that affect the interpretation of the sm-FRET data for the unfolded proteins. (i) Due to the dependency of E on the 6th power of R/R_0 , the efficiency E changes sensitively to the changes of R near the Förster distance R_0 , typically ~ 50 Å for pairs of chromophores used for single molecule measurements. However, the method is not sensitive in the regions of R below 20 Å or longer than 80 Å. (ii) Several amino acids quench the excited state of chromophores and modulate the observed FRET efficiencies. In particular, the aromatic residues quench the excited state by electron transfers when the residues are in the vicinity to the chromophores (Chen et al. 2010). (iii) The rotational dynamics of the two dyes is assumed to be not restricted and to be much faster than the time scale of the changes in R . If the relative orientation of the two chromophores is fixed or restricted, Eq. (3) is still valid but the Förster distance R_0 should deviate significantly from the value predicted by assuming the free rotation of the chromophores. If the time scale for the rotational dynamics of the chromophores is comparable to that of the changes in R , R_0 in Eq. (3) becomes the function of the relative orientation of the chromophores, making E the function of R and the angles describing orientations of transition dipoles of the chromophores. (iv) The accuracy of the FRET efficiency determination is limited by the numbers of observed photons, and is usually low at the single molecule level. In standard experiments, we obtain at most ~ 100 photons from each of the donor and acceptor chromophores in a single event called burst that typically lasts a few milliseconds; however, the detected photon number, n , has intrinsic uncertainty of $\pm n^{1/2}$ called shot noise determined by photon statistics. Accordingly, the FRET efficiency values obtained as the time-averaged quantity over the typical duration of 1 ms still possess significant uncertainty limited by the shot noise. Experimental results displayed as the frequency histogram of the observed FRET efficiencies (the FRET efficiency plot) usually consist of broad peaks, whose width sometimes exceeds ~ 0.2 even for a single homogeneous component. In addition, it has been difficult to increase the time resolution of the sm-FRET measurements to less than 1 ms. (v) The value of γ is usually set to 1 by selecting the donor and acceptor chromophores and optical filters properly; however, the deviation of the γ value from unity affects the FRET efficiency plot significantly and may make the peak shape asymmetric, causing difficulties in the quantitative evaluation of the FRET efficiency values (Oikawa et al. 2015). In the case for $\gamma = 1$ and for the peaks near $E = 0.5$, the peak of the FRET efficiency

plot for single homogeneous component might be approximated by a Gaussian function. In other cases, the peak shapes need to be analyzed by numerical calculations based on the general formulae (Oikawa et al. 2015).

Due to the limitations explained above, the use of the single-molecule FRET spectroscopy data for the structural analysis of the unfolded proteins requires several assumptions (Hoffman 2016). First, by selecting the labeling site of chromophores carefully, the limitations (ii) and (iii) stated above are usually assumed to be absent. Second, the broadness of the peaks of the FRET efficiency plot is assumed to be caused solely by the shot noise and not by the heterogeneity of the observed species. If the sm-FRET measurements were conducted in the system having the γ value of 1, the peak corresponding to the unfolded state might be fitted by a Gaussian function to estimate the peak efficiency, $\langle E \rangle$, of the unfolded state. We stress again that no heterogeneity is assumed in this data analysis, and only the single value, $\langle E \rangle$, is obtained from the sm-FRET results. Third, the peak efficiency value $\langle E \rangle$ assigned to the unfolded state was interpreted to be the overlap of the R dependency of the FRET efficiency (Eq. (3)) and the donor-acceptor distance distribution, $P(R)$, as described in Eq. (4):

$$\langle E \rangle = \int E(R) \cdot P(R) dR \quad (4)$$

Since the accurate functional form of $P(R)$ is not known, an approximate function needs to be assumed. The simplest of the functions is the Gaussian chain model, a random coil allowing the overlap of chains in the same space, as described in Eq. (5):

$$P(R) = 4\pi R^2 \left(\frac{3}{2\pi \langle R^2 \rangle} \right)^{\frac{3}{2}} \exp \left(-\frac{3R^2}{2\langle R^2 \rangle} \right) \quad (5)$$

where $\langle R^2 \rangle$ is the mean squared distance of R and is the single fitting parameter of the model. By numerically calculating the integration of Eq. (5) to match the observed efficiency $\langle E \rangle$, the fitting parameter $\langle R^2 \rangle$ can be obtained. If the labeling sites of the donor and acceptor chromophores are both the termini, the obtained $\langle R^2 \rangle$ can be related to the radius of gyration (R_g) by $\langle R^2 \rangle = 6 R_g^2$ based on the Gaussian chain model. Thus, the compactness of the unfolded protein, first time averaged over 1 ms, and second ensemble averaged over the population of the unfolded state assuming a single homogeneous component, can be deduced by using the current method of the sm-FRET spectroscopy.

Considering the rather drastic assumptions used for the analysis of single-molecule FRET data, it might seem surprising that the analysis gave largely consistent interpretations for different unfolded proteins. Several polymer models other than the Gaussian chain model were also used and gave slightly different values of $\langle R^2 \rangle$; however, the use of other models does not alter the results significantly (Borgia et al. 2016). A

set of data obtained for ubiquitin in the presence of 8 M urea, for example, demonstrated that the unfolded state of the protein mutants labeled at the different separations of the donor and acceptor chromophores showed the dependency of $\langle R^2 \rangle$ on the numbers of the residues between the labeled positions exactly as predicted for the random coils with excluded volume effect (Eq. (1) with $\nu = 0.6$) (Aznauryan et al. 2016). Examination of the same mutants by using the ns-FRET correlation method gave the reconfiguration time shorter than 100 ns, which is consistent with the analysis of the single-molecule FRET data assuming the perfect time averaging in the millisecond time scale (Aznauryan et al. 2016). Thus, the use of Eqs. (3)–(5) for the sm-FRET data is currently an accepted strategy for the structural analysis for the unfolded proteins.

A possible failure of the above analysis of the sm-FRET data for the unfolded proteins might be reflected in the disagreement between the R_g values deduced by the sm-FRET measurements and those by the SAXS measurements. In the sm-FRET measurements, the estimated R_g values for the unfolded ensemble demonstrate significant reduction as the concentration of the denaturant was reduced. In contrast, the SAXS data for the unfolded state rather demonstrate much smaller dependency of the R_g values on the denaturant concentration. The discrepancy is currently highly debated and might require further investigations (Watkins et al. 2015; Borgia et al. 2016; Aznauryan et al. 2016; Fuertes et al. 2017; Riback et al. 2017). The use of the simple Gaussian chain model for the unfolded state for the analysis of the sm-FRET data, which was reasonable as the first order approximation given the limited information of the method, but might be too simplistic for the quantitative characterization of the unfolded proteins (Maity and Reddy 2016; Song et al. 2017).

Line confocal method of single-molecule FRET investigations

Among the limitations of the conventional single-molecule FRET method for the structural characterization of the unfolded proteins, the most critical is the small photon number available from one molecule in a single event, limiting the accuracy and the time resolution of the FRET efficiency measurements. Due to the small number, we can obtain only the time averaged FRET efficiency E from a single event typically averaged for ~ 1 ms, which still contains significant uncertainty due to the shot noise. To reduce the uncertainty, all the samples in the unfolded ensemble are assumed to possess the same FRET efficiency $\langle E \rangle$, which was estimated by fitting the peak by a Gaussian function. Accordingly, the ensemble and time averaged FRET efficiency $\langle E \rangle$ is used for the analysis of the unfolded proteins. We also explained that the very rapid conformational fluctuations elucidated for the unfolded proteins rationalizes the use of the single

and averaged value of $\langle E \rangle$, since the single molecule properties should be time averaged in the time region longer than microseconds. In this section, we will discuss our recent efforts to increase the time resolution and the observation duration in the single-molecule FRET detection, and the apparent heterogeneity of the unfolded proteins that contradicts to the accumulated observations for the unfolded proteins.

To improve the time resolution and the observation duration, we developed a new method for the detection of single-molecule fluorescence. We discovered, by chance, that it is possible to increase the number of photons available from single molecules by flowing sample molecules rapidly along the microfluidic channel and by irradiating the excitation laser in the line shape along the flow channel (Oikawa et al. 2013). We detect fluorescence photons from single molecules flowing along the channel by imaging the flow path to an imaging detector. For the conventional method of single molecule detection, the elimination of the background light is important and is achieved by using a narrow hole placed at the focal plane. Instead, we used a slit to reduce the background light and to image molecules flowing along the line shaped excitation light. In this line confocal optics, it is possible to detect more than several thousands of fluorescence photons per millisecond easily. In addition, it is possible to observe single molecules continuously for more than 5 ms. Due to the increase in the photon numbers, it becomes possible to improve the time resolution from 1 ms of the previous method to less than 100 μ s in the new method. In addition, the increased photon numbers enabled us to reduce the shot noise in the data time-averaged for one millisecond. We discussed the mechanism of the increase of the photon numbers, the optical systems used and the method of the data analysis in

our previous publications (Oikawa et al. 2013; Oikawa et al. 2015; Saito et al. 2016). In the most recent improvement of the method, we achieved the time resolution of 10 μ s and the observation duration of ~ 5 ms (Oikawa et al. 2018).

So far, we have investigated the unfolded state of two proteins by using the line-confocal method of the sm-FRET spectroscopy. The first is the B domain of protein A (BdpA), which is a single-domain three-helix bundle composed of 63 residues (Oikawa et al. 2015). To our surprise, the width of the peak ascribed to the unfolded state of BdpA did not agree with the width predicted from the shot noise (Fig. 1). The width of the unfolded state peak in the FRET efficiency plot at the fastest time resolution (120 μ s) might be consistent with that of the shot noise limited peak (Fig. 1a). However, the data averaged for 2 ms showed the peak width broader than that of the shot noise limited peak (Fig. 1b), implying that the apparent heterogeneity of the unfolded state for BdpA. The sm-FRET traces observed for the unfolded state also demonstrated the heterogeneity. The traces for different molecules possessed slightly different FRET efficiencies, but kept constant efficiency over the observation period of several milliseconds (Fig. 1c). This demonstrates that the apparent heterogeneity of the unfolded state lasts for more than several milliseconds. As the second example of the unfolded proteins, we investigated ubiquitin and again observed the broader peak width for the unfolded state (Saito et al. 2016). The FRET efficiency values for the unfolded ubiquitin differ from molecule to molecule but are constant over the observation period. We suggested that the very slow dynamics occurring in the time domain longer than several milliseconds might equilibrate the unfolded state of ubiquitin. The FRET efficiency plot for ubiquitin observed by us looked consistent with those

Fig. 1 The significant heterogeneity detected in the FRET efficiency data obtained by the line confocal method of the sm-FRET spectroscopy. The unfolded state of the B domain of protein A, labeled by Alexa488 and ATTO633 to cysteines introduced at 5th and 55th residues near the N and C termini, in the presence of 4 M guanidinium was observed. **a** presents the FRET efficiency plot for the raw data with the time resolution of 120 μ s. **b** presents the FRET efficiency plots after the moving time average of 2 ms. In **a** and **b**, pink bars represent the actual data and the

dotted lines represent the theoretical distribution caused by photon shot noise estimated based on the averages of fluorescence intensity, FRET efficiency, and background photons. **c** presents the examples of the traces observed in the presence of 4 M guanidinium. The dashed lines represent the noise width calculated based on the numbers of fluorescence and background photons for each trace. All data were taken from Oikawa et al. (2015)

reported by other groups, although the latter data were based on the conventional system and could not reveal the heterogeneity (Aznauryan et al. 2016). The results for the two proteins suggest that the heterogeneity might be a frequent observation of the unfolded proteins.

While previous sm-FRET experiments were conducted under the limited number of photons based on the conventional confocal microscope, the observed peak width for the unfolded proteins were frequently broader than the width determined by the shot noise (Denitz et al. 2000; Schuler et al. 2002; Kuzmenkina et al. 2005; Tezuka-Kawakami et al. 2006; Merchant et al. 2007). One of the first single-molecule FRET results of the unfolded state demonstrated the broadening of the peak for the unfolded cold shock protein (68 residues) beyond the shot noise limited width (Schuler et al. 2002). A similar broadening for the unfolded peak was pointed out for chymotrypsin inhibitor 2 (64 residues) (Denitz et al. 2000). The unfolded state of ribonuclease HI (155 residues) was shown to possess heterogeneous peak width, in which the jumps among different unfolded subpopulations occurred in the time scale of seconds (Kuzmenkina et al. 2005). The FRET efficiency plot for the unfolded state of protein L was reported to be broader than the peak limited by the shot noise, whose fluctuation was assumed to occur in the time scale longer than several milliseconds (Merchant et al. 2007). These results strongly suggest that the apparent heterogeneity is, at least, a frequent observation for the unfolded state.

In summary, the single-molecule investigation for BdpA and ubiquitin based on the line confocal method of the sm-FRET spectroscopy revealed significant heterogeneity for the unfolded state. The heterogeneous structures should possess lifetimes longer than a few milliseconds. The fast time resolution and the long observation time achieved by the line confocal method were the key advancements that enabled us to reveal the heterogeneity; however, the heterogeneity was actually reported repeatedly in the past investigations based on the conventional method. The results suggest a striking conclusion that the unfolded state of proteins might be significantly heterogeneous in the surprising time scales slower than a few milliseconds.

Hypothesis: the coupling of the local clusters and the long-range distance distribution

The apparent heterogeneity of the unfolded state demonstrated in the line confocal method of the sm-FRET spectroscopy implies that the structural transitions among different conformations of the unfolded state might occur in the time domain longer than several milliseconds; however, the heterogeneity clearly contradicts to the fast conformational fluctuations reported for many unfolded proteins. If the conformational dynamics of polypeptides occurs in the time domain shorter than 100 ns as demonstrated in the ns-FRET correlation method,

the donor-acceptor distance measured by the sm-FRET spectroscopy should be time averaged in the observation time of 100 μ s. In this section, we propose a hypothesis that explains the apparently contradicting observations for the unfolded proteins. The hypothesis is composed of four proposals.

First, to explain the origin of the heterogeneity, we propose that a local structural heterogeneity that lasts more than several milliseconds is present in the vicinity of the labeling site of the unfolded proteins. The local structure likely involves hydrophobic residues and differs slightly from molecule to molecule, causing the origin of the heterogeneity. The lifetime of the residual structures in the unfolded proteins is usually considered to be shorter than 1 ms. However, in one example, the RDC measurement for the unfolded state of lysozyme reported the presence of several clusters, one of which containing two consecutive tryptophan residues possesses a lifetime up to milliseconds (Klein-Seetharaman et al. 2002; Sziegat et al. 2012). Accordingly, we propose that the lifetime of the local clusters in the absence of denaturant can be longer than 1 ms. The proposed local clusters might be formed in the absence of the chromophores. However, considering a large size of the chromophores used for single molecule fluorescence studies, the clusters might also be formed around the labeled chromophores. For example, the fixation of Alexa488 to the polyproline backbone was proposed to explain the heterogeneity in the single-molecule FRET data for the polyprolines (Hoeftling et al. 2011). Accordingly, we propose that the association of the chromophore to hydrophobic side chains or to main chain for prolonged period also occurs. Formation of the relatively stable local structures that may or may not involve the chromophore-polypeptide interaction is the first important proposal of the current hypothesis.

Second, we consider the case that the local clusters are separated from the chromophores, and that the rotation of the chromophores is not restricted and is much faster than the dynamic changes in R . To link the structural heterogeneity of the local clusters and the observed broadening of the FRET peak assigned to the unfolded state, we propose that the local clusters modulate the donor-acceptor distance distribution, $P(R)$, causing slightly different forms of $P(R)$ from molecule to molecule. Considering the very fast dynamics revealed for the unfolded proteins, the local cluster formed by the association of hydrophobic residues should also collide frequently to other hydrophobic residues located separately along the sequence, and forms the transiently associated conformations. The frequency of the association and the lifetime of the associated conformation should modulate the shape of $P(R)$. If the structure of the local cluster is heterogeneous as we assumed above, the frequency of the association and the lifetime of the associated conformation should also be modulated by the local structural heterogeneity, and can cause heterogeneity in the shape of $P(R)$. In the standard analysis for the unfolded state FRET data, we assume a common polymer model for the

entire ensemble of proteins. We propose that different functional forms of $P(R)$ depending on the local structural heterogeneity explain the data of the unfolded state. We stress that the modulation of $P(R)$ will not change the very fast reconfiguration time of the unfolded polypeptides. The direct coupling of the local structural heterogeneity and the long-range distance distribution is the second important proposal of the current hypothesis.

Third, we next consider the case where the local clusters are formed around the labeled chromophores. In this case, the local dynamics of the chromophore should be restricted and become much slower, and the assumption for the use of Eq. (3), the separation of the time scales for the rotational dynamics of chromophores and the dynamics of the interchromophore distance R , might no longer be valid. It can be argued that even if the local clusters might restrict the free rotation of one or both of the donor and acceptor chromophores, the relative orientation of the two chromophores should still change very rapidly considering the structural flexibility of the unfolded polypeptides intervening the two chromophores. The argument might be correct for a fraction of molecules having longer values of R ; however, for the other fraction of molecules having shorter values of R , the rotation of the chromophores should be restricted due to the larger numbers of intra-protein contacts. In an extreme case, the relative orientation of the two chromophores might be fixed in the conformation having short R . Accordingly, the local structural heterogeneity around the chromophores can cause heterogeneity in the relative orientation of the two chromophores for the conformations having short R , and results in the broadening of the peak in the FRET efficiency plot. The enhanced broadening effect in the FRET efficiency plot due to the fixation of the rotational dynamics of chromophores is the third proposal of the hypothesis.

Fourth, to further link the local structural heterogeneity and the observed broadening of the peak in the FRET efficiency plot for the unfolded proteins, we propose that the local and heterogeneous cluster surrounding the labeled chromophores quenches the excited singlet state and causes the heterogeneity in the fluorescence quantum yield of the chromophores. The similar heterogeneity in the sm-FRET peak reported for the double-stranded DNAs labeled by donor and acceptor chromophores was explained in terms of the variance of the fluorescence quantum yields (Deniz et al. 1999; Holden et al. 2010; Kalinin et al. 2010). The aromatic side chains are known to quench the fluorescence of chromophores used for single molecule spectroscopy (Chen et al. 2010). Accordingly, the heterogeneity in the local structures surrounding the chromophores can cause the heterogeneity in their fluorescence quantum yields. The heterogeneous modulation of the photophysical properties of the labeled chromophores by the local structural heterogeneity is the fourth important proposal of this hypothesis.

In summary, to explain the apparent contradiction reported for unfolded proteins, the broad peaks in the FRET efficiency plot obtained by the sm-FRET spectroscopy and the very fast reconfiguration time observed by the ns-FRET correlation method, we first hypothesized that the local hydrophobic clusters might be present in the unfolded proteins in the absence of denaturant. The clusters are heterogeneous and last for more than a few milliseconds. Second, we proposed that the local clusters further modulate the long-range distance distribution of the donor and acceptor chromophores and cause the peak broadening of the FRET efficiency plot. The coupling of the local clusters and the long-range distance distribution is reminiscent of the proposal by Shortle (2002). Based in the correlation between the RDC data for proteins in the native and denatured conditions, he proposed that the local steric interactions might restrict the conformation of the denatured proteins to possess the native like topology (Shortle and Ackerman 2001; Ohnishi et al. 2004). Third, the fixation of the rotational dynamics of the chromophores to the local clusters further broadens the peak in the FRET efficiency plot. Fourth, the changes in the fluorescence quantum yields of the labeled chromophores due to the quenching by the surrounding chromophores might further contribute to the peak broadening of the FRET efficiency plot. The proposals two to four are compatible with the very fast reconfiguration time of the unfolded proteins. Accordingly, the presence of local structural heterogeneity and its coupling to the long-range distance distribution can explain the apparent contradiction reported for unfolded proteins.

Roles of the local heterogeneous clusters in protein folding

Based on the peak broadening observed in the FRET efficiency plots for the unfolded proteins, we proposed the presence of the local heterogeneous clusters in the unfolded state of proteins. The local clusters might be inherent to the unfolded proteins; however, the cluster might be formed artificially around the labeled chromophores. Considering the widespread use of labeling of chromophores and considering the impact of the local heterogeneous clusters in protein folding, it is critically important to distinguish the two possibilities. Several lines of indirect evidence rather suggest that the former possibility is more likely. For example, the reported stability of the labeled proteins was not significantly different from that of the non-labeled proteins (Deniz et al. 2000; Schuler et al. 2002; Saito et al. 2016). That the kinetic folding of the labeled proteins was comparable to that of the non-labeled protein was confirmed in several systems. However, these observations do not rule out the possibility that the labeling causes the heterogeneity. While we are in the opinion that the slow structural heterogeneity is present even in

proteins without the labeling of chromophores, further critical examinations of the two possibilities are important.

If slowly fluctuating elements of local structures are proven to be inherent feature of unfolded proteins, then this finding would have a significant impact for the mechanism of protein folding. Here, we limit ourselves to propose one experiment that should reveal the roles of the heterogeneous cluster in the kinetic folding. By combining the microfluidic mixing system and the line confocal method of the sm-FRET spectroscopy, we can expect to track the process occurring after the rapid dilution of the denaturant. It may be possible to distinguish if all the heterogeneous conformations of the unfolded state convert to the native state at the same rate or if some of the unfolded state conformations preferentially convert to the native state. Efforts toward this goal are currently in progress in our laboratory.

Acknowledgements We dedicate this article to Prof. Fumio Arisaka. We thank our collaborators and the former and current members of our laboratory.

Compliance with ethical standards

Funding This work was supported by JSPS KAKENHI Grant Number JP25104007 (to S.T.) and JSPS KAKENHI Grant Number JP17K17608 (to H.O.).

Conflict of interest Satoshi Takahashi declares that he has no conflict of interest. Aya Yoshida declares that she has no conflict of interest. Hiroyuki Oikawa declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Abaskharon RM, Gai F (2016) Meandering down the energy landscape of protein folding: are we there yet? *Biophys J* 110:1924–1932
- Aznauryan M, Delgado L, Soranno A, Nettels D, Huang J, Labhardt AM, Grzesiek S, Schuler B (2016) Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS. *Proc Natl Acad Sci U S A* 113:E5389–E5398
- Baxa MC, Haddadian EJ, Jumper JM, Freed KF, Sosnick TR (2014) Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations. *Proc Natl Acad Sci U S A* 111:15396–15401
- Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RW, Blackledge M (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc Natl Acad Sci U S A* 102:17002–17007
- Borgia A, Wensley BG, Soranno A, Nettels D, Borgia MB, Hoffmann A, Pfeil SH, Lipman EA, Clarke J, Schuler B (2012) Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. *Nature Commun* 3:1195 (9 pages)
- Borgia A, Zheng W, Buholzer K, Borgia MB, Schüler A, Hofmann H, Soranno A, Nettels D, Gast K, Grishaev A, Best RB, Schuler B (2016) Consistent view of polypeptide chain expansion in chemical denaturants from multiple experimental methods. *J Am Chem Soc* 138:11714–11726
- Bowler BE (2012) Residual structure in unfolded proteins. *Curr Opin Struct Biol* 22:4–13
- Buckler DR, Haas E, Scheraga HA (1995) Analysis of the structure of ribonuclease A in native and partially denatured states by time-resolved nonradiative dynamic excitation energy transfer between site-specific extrinsic probes. *Biochemistry* 34:15965–15978
- Buscaglia M, Lapidus LJ, Eaton WA, Hofrichter J (2006) Effects of denaturants on the dynamics of loop formation in polypeptides. *Biophys J* 91:276–288
- Chen H, Ahsan SS, Santiago-Berrios MB, Abruña HD, Webb WW (2010) Mechanisms of quenching of Alexa fluorophores by natural amino acids. *J Am Chem Soc* 132:7244–7245
- Cho J-H, Meng W, Sato S, Kim EY, Schindelin H, Raleigh DP (2014) Energetically significant networks of coupled interactions within an unfolded protein. *Proc Natl Acad Sci U S A* 111:12079–12084
- Dar TA, Schaeffer RD, Daggett V, Bowler BE (2011) Manifestations of native topology in the denatured state ensemble of *Rhodospseudomonas palustris* cytochrome *c*'. *Biochemistry* 50:1029–1041
- Deniz AA, Dahan M, Grunwell JR, Ha T, Faulhaber AE, Chemla DS, Weiss S, Schultz PG (1999) Single-pair fluorescence resonance energy transfer on freely diffusing molecules: observation of Förster distance dependence and subpopulations. *Proc Natl Acad Sci U S A* 96:3670–3675
- Deniz AA, Laurence TA, Beligere GS, Dahan M, Martin AB, Chemla DS, Dawson PE, Schultz PG, Weiss S (2000) Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc Natl Acad Sci U S A* 97:5179–5184
- Dill KA, MacCallum JL (2012) The protein-folding problem, 50 years on. *Science* 338:1042–1046
- Felitsky D, Lietzow MA, Dyson HJ, Wright PE (2008) Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. *Proc Natl Acad Sci U S A* 105:6278–6283
- Fitzkee NC, Rose GD (2004) Reassessing random-coil statistics in unfolded proteins. *Proc Natl Acad Sci U S A* 101:12497–12502
- Fuertes G, Banterle N, Ruff KM, Chowdhury A, Mercadante D, Koehler C, Kachala M, Estrada Girona G, Milles S, Mishra A, Onck PR, Gräter F, Esteban-Martín S, Pappu RV, Svergun DI, Lemke EA (2017) Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs FRET measurements. *Proc Natl Acad Sci USA* 114:E6342–E6351
- Hagen SJ, Hofrichter J, Szabo A, Eaton WA (1996) Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding. *Proc Natl Acad Sci U S A* 93:11615–11617
- Haran G (2012) How, when and why proteins collapse: the relation to folding. *Curr Opin Struct Biol* 22:14–20
- Hoefling M, Lima N, Haenni D, Seidel CAM, Schuler B, Grubmüller H (2011) Structural heterogeneity and quantitative FRET efficiency distributions of polyprolines through a hybrid atomistic simulation and Monte Carlo approach. *PLoS One* 6:e19791 (19 pages)
- Hoffmann H (2016) Understanding disordered and unfolded proteins using single-molecule FRET and polymer theory. *Methods Appl Fluoresc* 4:042003 (8 pages)
- Hofmann H, Soranno A, Borgia A, Gast K, Nettels D, Schuler B (2012) Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc Natl Acad Sci U S A* 109:16155–16160
- Holden SJ, Uphoff S, Hohlbein J, Yadin D, Le Reste L, Britton OJ, Kapanidis AN (2010) Defining the limits of single-molecule FRET resolution in TIRF microscopy. *Biophys J* 99:3102–3111

- Huang J-r, Grzesiek S (2010) Ensemble calculations of unstructured proteins constrained by RDC and PRE data: a case study of urea-denatured ubiquitin. *J Am Chem Soc* 132:694–705
- Jensen MR, Zweckstetter M, Huang J-r, Blackledge M (2014) Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem Rev* 114:6632–6660
- Jha AK, Colubri A, Freed KF, Sosnick TR (2005) Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc Natl Acad Sci U S A* 102:13099–13104
- Kalinin S, Sisamakos E, Magennis SW, Felekyan S, Seidel CAM (2010) On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits. *J Phys Chem B* 114:6197–6206
- Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H (2002) Long-range interactions within a nonnative protein. *Science* 295:1719–1722
- Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiyagarajan P, Sosnick TR, Hasan MZ, Pande VS, Ruczinski I, Doniach S, Plaxco KW. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. USA*. 101, 12491–12496. Erratum in: (2005) *Proc. Natl. Acad. Sci. USA*. 102, 14475
- Konuma T, Kimura T, Matsumoto S, Goto Y, Fujisawa T, Fersht AR, Takahashi S (2011) Time-resolved small-angle X-ray scattering study of the folding dynamics of barnase. *J Mol Biol* 405:1284–1294
- Kuzmenkina EV, Heyes CD, Nienhaus GU (2005) Single-molecule Förster resonance energy transfer study of protein dynamics under denaturing conditions. *Proc Natl Acad Sci U S A* 102:15471–15476
- Lapidus LJ, Eaton WA, Hofrichter J (2000) Measuring the rate of intramolecular contact formation in polypeptides. *Proc Natl Acad Sci U S A* 97:7220–7225
- Maity H, Reddy G (2016) Folding of protein L with implications for collapse in the denatured state ensemble. *J Am Chem Soc* 138:2609–2616
- Meier S, Grzesiek S, Blackledge M (2007) Direct observation of dipolar couplings and hydrogen bonds across a β -hairpin in 8 M urea. *J Am Chem Soc* 129:9799–9807
- Meng W, Luan B, Lyle N, Pappu RV, Raleigh DP (2013) The denatured state ensemble contains significant local and long-range structure under native conditions: analysis of the N-terminal domain of ribosomal protein L9. *Biochemistry* 52:2662–2671
- Merchant KA, Best RB, Louis JM, Gopich IV, Eaton WA (2007) Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc Natl Acad Sci U S A* 104:1528–1533
- Möglich A, Joder K, Kiefhaber T (2006) End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc. Natl. Acad. Sci. USA*. 103, 12394–12399. Erratum in: (2008) *Proc. Natl. Acad. Sci. USA*. 105, 6787
- Nettels D, Gopich IV, Hoffmann A, Schuler B (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc Natl Acad Sci U S A* 104:2655–2660
- Ohnishi S, Lee AL, Edgell MH, Shortle D (2004) Direct demonstration of structural similarity between native and denatured eglin C. *Biochemistry* 43:4064–4070
- Oikawa H, Suzuki Y, Saito M, Kamagata K, Arai M, Takahashi S (2013) Microsecond dynamics of an unfolded protein by a line confocal tracking of single molecule fluorescence. *Sci Rep* 3:2151
- Oikawa H, Kamagata K, Arai M, Takahashi S (2015) Complexity of the folding transition of the B domain of protein A revealed by the high-speed tracking of single-molecule fluorescence time series. *J Phys Chem B* 119:6081–6091
- Oikawa H, Takahashi T, Kamonprasertsuk S, Takahashi S (2018) Microsecond resolved single-molecule FRET time series measurements based on line confocal optical system combined with hybrid photodetectors. *Phys Chem Chem Phys* 20:3277–3285
- Plaxco KW, Millett IS, Segel DJ, Doniach S, Baker D (1999) Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nat Struct Biol* 6:554–556
- Riback JA, Bowman MA, Zmyslowski AM, Knoverek CR, Jumper JM, Hinshaw JR, Kaye EB, Freed KF, Clark PL, Sosnick TR (2017) Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science* 358:238–241
- Saito M, Kamonprasertsuk S, Suzuki S, Nanatani K, Oikawa H, Kushi K, Takai M, Chen PT, Chen EH, Chen RP, Takahashi S (2016) Significant heterogeneity and slow dynamics of the unfolded ubiquitin detected by the line confocal method of single-molecule fluorescence spectroscopy. *J Phys Chem B* 120:8818–8829
- Sakurai K, Yagi M, Konuma T, Takahashi S, Nishimura C, Goto Y (2017) Non-native α -helices in the initial folding intermediate facilitate the ordered assembly of the β -barrel in β -lactoglobulin. *Biochemistry* 56:4799–4807
- Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419:743–747
- Schuler B, Soranno A, Hofmann H, Nettels D (2016) Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. *Annu Rev Biophys* 45:207–231
- Sherman E, Haran G (2006) Coil-globule transition in the denatured state of a small protein. *Proc Natl Acad Sci U S A* 103:11539–11543
- Shortle D (2002) The expanded denatured state: an ensemble of conformations trapped in a locally encoded topological space. *Adv Protein Chem* 62:1–23
- Shortle D, Ackerman MS (2001) Persistence of native-like topology in a denatured protein in 8 M urea. *Science* 293:487–489
- Song J, Gomes G-N, Shi T, Gradinaru CC, Chan HS (2017) Conformational heterogeneity and FRET data interpretation for dimensions of unfolded proteins. *Biophys J* 113:1012–1024
- Soranno A, Longhi R, Bellini T, Buscaglia M (2009) Kinetics of contact formation and end-to-end distance distributions of swollen disordered peptides. *Biophys J* 96:1515–1528
- Soranno A, Buchli B, Nettels D, Cheng RR, Müller-Späh S, Pfeil SH, Hoffmann A, Lipman EA, Makarov DE, Schuler B (2012) Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proc Natl Acad Sci U S A* 109:17800–17806
- Soranno A, Holla A, Dingfelder F, Nettels D, Makarov DE, Schuler B (2017) Integrated view of internal friction in unfolded proteins from single-molecule FRET, contact quenching, theory, and simulations. *Proc Natl Acad Sci U S A* 114:E1833–E1839
- Sosnick TR, Barrick D (2011) The folding of single domain proteins: have we reached a consensus? *Curr Opin Struct Biol* 21:12–24
- Sziegat F, Silvers R, Hähnke M, Jensen MR, Blackledge M, Wimmer-Bartschek J, Schwalbe H (2012) Disentangling the coil: modulation of conformational and dynamic properties by site-directed mutation in the non-native state of hen egg white lysozyme. *Biochemistry* 51:3361–3372
- Takahashi S, Kamagata K, Oikawa H (2016) Where the complex things are: single molecule and ensemble spectroscopic investigations of protein folding dynamics. *Curr Opin Struct Biol* 36:1–9
- Tezuka-Kawakami T, Gell C, Brockwell DJ, Radford SE, Smith DA (2006) Urea-induced unfolding of the immunity protein Im9 monitored by spFRET. *Biophys J* 91:L42–L44
- Udgaonkar JB (2013) Polypeptide chain collapse and protein folding. *Arch Biochem Biophys* 531:24–33
- Waldauer SA, Bakajin O, Lapidus LJ (2010) Extremely slow intramolecular diffusion in unfolded protein L. *Proc Natl Acad Sci U S A* 107:13713–13717

- Watkins HM, Simon AJ, Sosnick TR, Lipman EA, Hjelm RP, Plaxco KW (2015) Random coil negative control reproduces the discrepancy between scattering and FRET measurements of denatured protein dimensions. *Proc Natl Acad Sci U S A* 112:6631–6636
- Xue Y, Skrynnikov NR (2011) Motion of a disordered polypeptide chain as studied by paramagnetic relaxation enhancements, ^{15}N relaxation, and molecular dynamics simulations: how fast is segmental diffusion in denatured ubiquitin? *J Am Chem Soc* 133:14614–14628
- Yoo TY, Meisburger SP, Hinshaw J, Pollack L, Haran G, Sosnick TR, Plaxco K (2012) Small-angle X-ray scattering and single-molecule FRET spectroscopy produce highly divergent views of the low-denaturant unfolded state. *J Mol Biol* 418:226–236