Single-Molecule Fluorescence Spectroscopy of Intrinsically Disordered Proteins

Hagen Hofmann and Wenwei Zheng

Contents

Abstract The past two decades have seen a substantial leap forward in our understanding of intrinsically disordered proteins, in terms of both thermodynamics and dynamics, but also in terms of structural ensembles. From understanding the principles and biological importance of their solvent pliability up to characterizing their dynamics including an identification of the molecular origins of internal friction, single-molecule FRET experiments have been an important driver of this progress. By now, the methods and analysis tools in single-molecule FRET have grown to an extensive toolbox that allows a straightforward comparison of experiments with analytical theories and results of molecular simulations. This chapter summarizes the technologies behind single-molecule FRET experiments and molecular simulations together with the key findings on intrinsically disordered proteins.

H. Hofmann (✉)
Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot, Israel
e-mail: hagen.hofmann@weizmann.ac.il

W. Zheng
College of Integrative Sciences and Arts, Arizona State University, Mesa, AZ, USA
e-mail: wenweizheng@asu.edu

R. Šachl, M. Amaro (eds.), *Fluorescence Spectroscopy and Microscopy in Biology*,
Springer Ser Fluoresc (2023) 20: 403–442, https://doi.org/10.1007/4243_2022_38,

Keywords Intrinsically disordered proteins · Molecular simulations · Protein dynamics · Single-molecule FRET

1 Introduction

It often takes time for discoveries to settle in. After the first findings that proteins do not necessarily fold for function, intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) changed our structure-based view on how proteins interact [1]. While initially being identified as "nasty" flexible extrusions of proteins that notoriously escaped X-ray structure determinations, we now know that IDRs and IDPs longer than 30 amino acids compose nearly 30–44% of the human proteome [1, 2] with functions as diverse as those of classical proteins. As fascinating this discovery has been, our understanding of the "why" and "how" is only slowly emerging. Why have IDPs evolved? How do they combine specificity and affinity without forming structured complexes? What evolutionary advantage do IDPs offer compared to folded proteins? In fact, prokaryotes that are nearly void of IDPs also populated our planet successfully. Answers to these questions are pending and it seems questionable whether they will ever be obtained in detail. IDPs (here synonymously also for IDRs) form molecular complexes upon binding to other proteins, DNA, RNA, or small molecules, but contrary to most folded proteins, IDPs are often multi-functional, which enables them to form hubs in protein interaction networks [1, 3, 4]. Similar to folded proteins, they are often chemically modified via post-translational modifications and their degradation rates in cells are substantially increased compared to folded proteins [5]. While some of these features, e.g., high turnover rates or efficient post-translational modifications, are explained by the lack of stable structures, their ability to interact with a multitude of different ligands [1, 3] is more difficult to understand. Although some IDPs indeed fold upon binding to ligands, thus forming specific binding interfaces in folding-coupled binding reactions, folding is not necessary for binding and specificity [6]. For instance, an extremely flexible protein complex bare of structure is formed between the two IDPs Prothymosin α and Histone H1 [7]. Other examples of high-affinity complexes that retain a large degree of disorder include the RNA-chaperone NCD and nucleic acids [8], Sic1 and Cdc4 [9], the Na^+/H^+ exchanger tail and ERK2 [10], nucleoporin tails and nuclear transport receptors [11], and the cytoplasmic tail of E-cadherin and β-catenin [12].

Clearly, structure determinations based on X-ray or cryogenic electron microscopy (cryo-EM) do not provide detailed information about this enormous flexibility such that mainly four methods have been used in the past to characterize the behavior of IDPs: NMR (nuclear magnetic resonance), SAXS (small-angle X-ray scattering), SANS (small-angle neutron scattering), and fluorescence-based methods, particularly single-molecule Förster resonance energy transfer (smFRET). In fact, the past two decades have seen an overwhelming number of studies in which smFRET approaches provided insights into the physical principles that dictate the behavior of IDPs and unfolded proteins such as their environmental sensitivity [13–32] and their dynamics [23, 33–37].

Here, we discuss some of the common smFRET methods used to characterize the size and the dynamics of IDPs. Our particular goal is to discuss the principles behind dynamic smFRET approaches and how they can be linked to analytical and numerical models of IDPs and unfolded proteins. Given the excellent introduction into the principles of FRET in Chapter "Laboratory instrumentation", we will not repeat the basic theory to avoid redundancy. We start by introducing technical aspects of smFRET experiments in Sect. 2, followed by a discussion of findings, theories, and simulation approaches to characterize IDPs using smFRET experiments (Sect. 3). In Sect. 4, we discuss smFRET methods and analytical theories used to study the dynamics of IDPs.

2 Technical Aspects of smFRET Experiments

2.1 Confocal smFRET Experiments

Before we discuss how exactly smFRET experiments help to characterize IDPs, we would like to remind the reader of a few technical aspects of single-molecule spectroscopic techniques. Based on the instrumentation, single-molecule experiments can roughly be divided into two categories: camera-based and confocal experiments. Camera-based imaging experiments offer the advantage to observe many surface-tethered molecules simultaneously. Yet, the time-resolution of such experiments is in the order of 1–10 ms, depending on the camera used (sCMOS or EM-CCD), which is insufficient to capture the very fast sub-microsecond dynamics of disordered proteins. Hence, most insights into IDPs and their behavior were obtained with confocal microscopes that offer a much higher time-resolution. With the development of detectors with high detection efficiencies (SPADs – single-photon avalanche detectors) and single-photon counting devices [38, 39], the time-resolution of confocal smFRET experiments is only limited by the counting precision of a few picoseconds. The design of a confocal microscope is rather simple (Fig. 1a). An excitation laser is focused through an objective of high numerical aperture (*NA*) to a diffraction limited spot whose diameter is given by $d = \lambda/2NA$. Since typical numerical apertures are in the order of 1.35–1.45, the confocal spot size is in the order of several hundreds of nanometers, i.e., small compared to many cellular structures, but very large compared to the size of proteins (2–5 nm) (Fig. 1b). The fluorescence of an emitter in the confocal spot is collected via the same objective, imaged on a pinhole to remove out-of-focus fluorescence, before it is directed onto a SPAD. Clearly, smFRET experiments with two colors (donor and acceptor) require two detectors and often even four detectors are being used to obtain information about the rotational degrees of freedom of protein-attached dyes in fluorescence anisotropy single-molecule measurements.

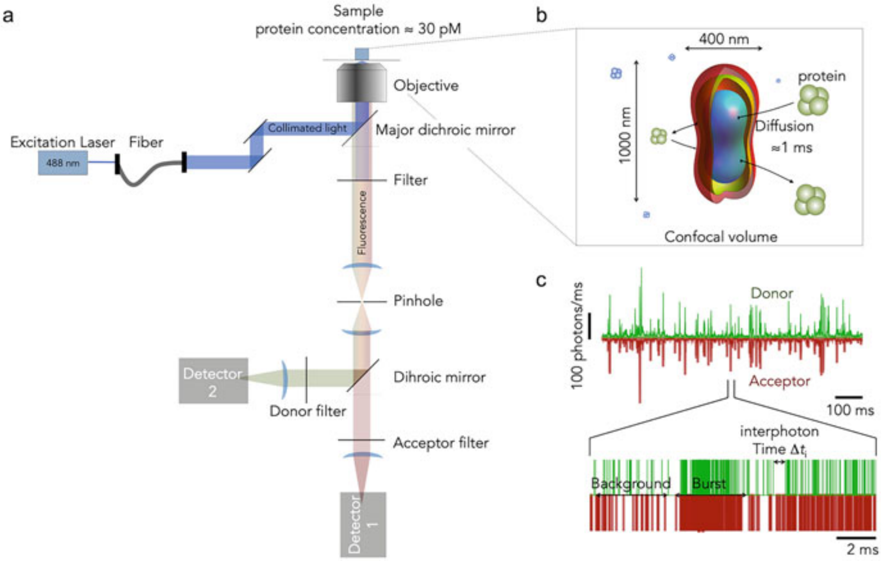

Fig. 1 The principle of confocal single-molecule FRET experiments. (**a**) Scheme of a confocal microscope. Excitation light (blue) from a laser is focused via an objective into the sample solution. Fluorescence photons from the sample are collected via the same objective, filtered, imaged on a pinhole, and finally recorded with single-photon avalanche detectors (SPADs). (**b**) The confocal volume created by the objective has a higher intensity in the center (blue) and lower intensity at the periphery [40]. The diffusion time of molecules through this volume is in the order of a millisecond. (**c**) Example of the raw data of a smFRET experiment. Molecules diffusing in and out of the confocal volume cause bursts of donor and acceptor photons in the time trace, which is shown here at a low resolution of 1 ms (upper panel). At higher time-resolution (0.1 μs), individual photons can be distinguished in the time trace (lower panel). The identification of bursts is based on a threshold of inter-photon times

In the past, most smFRET experiments on IDPs and unfolded proteins were performed with freely diffusing molecules. These types of experiments are relatively easy to realize and typically capture timescales from hundreds of picoseconds (the rotational motions of dyes attached to a protein) up to several milliseconds. The raw data of these experiments are photon traces that contain the arrival times of photons on the detectors (Fig. 1c). The construction of FRET histograms therefore requires an identification of the photons that were emitted by a molecule while it was diffusing through the confocal spot, thus resulting in a burst of photons. In addition, these raw photon counts have to be corrected for a number of experimental imperfections. Importantly, state-of-the-art burst-search algorithms already take these corrections into account [30, 41], which minimizes the risk of biasing the distribution of FRET efficiencies.

2.2 Correction Parameters in smFRET Experiments

In the following, we describe the correction of FRET values for two-color FRET experiments. However, the correction methods are general and also applicable to multi-color experiments. To determine correct FRET efficiencies (E) in a smFRET experiment with the channels 1 (acceptor) and 2 (donor), the raw photon counts (n_1 and n_2) of an individual molecule have to be corrected by the differences in brightness of the two dyes, cross-talk between the channels, background photons, and direct excitation of the acceptor at the wavelength used to excite the donor. Two main approaches are currently being used for these corrections: the determination of correction factors in separate experiments [30, 42, 43] and the determination within the same experiment [44].

Let us start with the separate-sample method. Here, the correction parameters are determined from two separate samples of the donor and acceptor dyes in which their concentrations are adjusted such that both samples have an absorbance of 0.1 at the wavelength of the donor excitation laser [42]. The fluorescence signal of both samples after donor excitation will then be measured in channel 1 and 2, and three correction factors can be determined: γ which accounts for the different detection probabilities of photons from the two dyes, β_{21} and β_{12}, the leakage of donor photons from channel 2 into the acceptor channel 1, and the leakage of acceptor photons (channel 1) into the donor channel 2, respectively. In addition, the probability α to directly excite the acceptor dye at the excitation wavelength specific for the donor can be computed from the extinction coefficients of the dyes at the donor excitation wavelength (see below). If n_1 and n_2 are the detected raw photon counts for a molecule in channel 1 and 2, respectively, and b_1 and b_2 are the background rates in both channels (in units of photon/s), the corrected photon counts for acceptor and donor (n'_D, n'_A) of this molecule are given by

$$\begin{pmatrix} n_A \\ n'_D \end{pmatrix} = \begin{pmatrix} 1 & -\beta_{21} \\ -\beta_{12} & \gamma \end{pmatrix} \begin{pmatrix} n_1 - b_1 T \\ n_2 - b_2 T \end{pmatrix} \text{ and } n'_A = n_A - \alpha\left(n'_D + n_A\right) \tag{1}$$

Here, T is the duration of a burst, which typically is in the order of a millisecond. As an example, for the frequently used FRET-pair Alexa Fluor 488 as donor and Alexa Fluor 594 as acceptor, our lab obtained $\gamma = 1.12 \pm 0.09$, $\beta_{21} = 0.050 \pm 0.003$, and $\beta_{12} = 0.0021 \pm 0.0004$ over 5 years with 21 measurements of these correction factors. As mentioned above, the probability of directly exciting the acceptor dye at the donor excitation wavelength is given by $\alpha = \epsilon_A/(\epsilon_A + \epsilon_D)$ where ϵ_A and ϵ_D are the extinction coefficients of the dyes at the donor excitation wavelength. For the dye pair mentioned above, one finds $\alpha = 0.049$. Hence, the correct FRET efficiency of an identified molecule is given by

$$E = \frac{n'_A}{n'_A + n'_D}. \tag{2}$$

Importantly, to prevent a bias in sampling molecules with different FRET efficiency, the corrections (Eq. 1) have to be taken into account during the identification of molecules (bursts) in the raw photon trace. Commonly, bursts are identified from the measured photon traces following Eggeling et al. [41] and Hoffmann et al. [30]. If $\Delta t_i = t_i - t_{i-1}$ is the inter-photon time of the i-th photon (Fig. 1c), the photon is retained if $\Delta t_i \leq \gamma_j \Delta t_{\max}$, in which $\Delta t_{\max}$ is an inter-photon time threshold set by the researcher, typically ~100 μs, and γ_j is the correction factor of the i^{th} photon detected in channel $j = [1, 2]$, i.e., $\gamma_1 = 1$ and $\gamma_2 = 1.12$ with the definition in Eq. 1. The algorithm then proceeds to the next photon $i + 1$, stops after n photons once $\Delta t_{i+n} > \gamma_j \Delta t_{\max}$, and provides the total length of the burst by $T = t_{n-1} - t_{i-1}$. The photon number in the resulting photon string is now corrected via Eq. 1 using estimated background rates b_1 and b_2. The initial guess of b_1 and b_2 is given by all detected photons in channel 1 and 2, respectively, divided by the total measurement time. A burst is finally identified if $n_A + n_D > n_T$ where again, n_T is a photon threshold that has to be set. Based on our experience, the photon threshold to identify a burst (molecule) is in the order of 30–100 photons in most smFRET experiments. In the next iteration, the photons belonging to each identified burst are removed from the photon trace and a new guess for b_1 and b_2 is computed based on the remaining photons. Subsequently, the burst search is performed again with updated background rates. This procedure converges after three iterations to the correct background rates and a constant number of identified bursts.

The disadvantage of the separate-sample method is that experiments with different samples (proteins) might require different correction factors since the quantum yield of the dyes and therefore γ also depend on the chemical environment. Similarly, experiments at different solvent conditions might change the brightness of donor and acceptor differently, thus leading to errors in the correct determination of FRET efficiency if a single set of correction factors is being used. An alternative method is based on ALEX (alternating laser excitation) [45] or PIE (pulsed-interleaved excitation) [46] excitation schemes. Here, the donor and the acceptor are excited in an alternating manner: a period of donor excitation is followed by a period of acceptor excitation with a second laser (Fig. 2a). If switching between both excitation sources is much faster than the diffusion of a molecule through the confocal spot, important additional information can be obtained. For example, by exciting the acceptor dye directly in a period, molecules that do not contain an active acceptor due to photo-bleaching or imperfect labeling can be identified and removed from the further analysis. To this end, a second parameter, the stoichiometry ratio, defined via

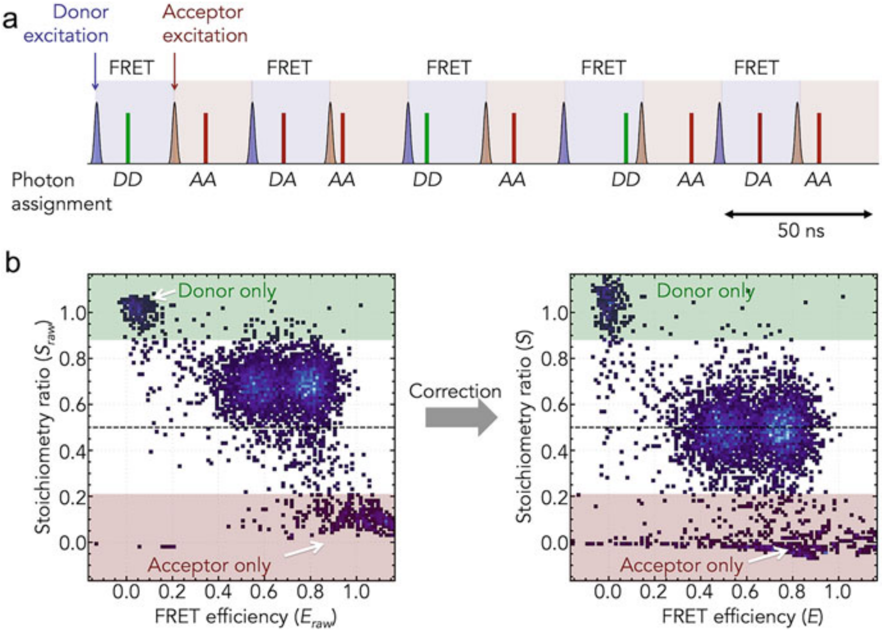

Fig. 2 Pulsed-interleaved single-molecule detection (PIE). (**a**) Excitation scheme of a PIE experiment. Excitation lasers for donor (blue peaks) and acceptor (red peaks) alternate during the experiment. Photons after donor excitation (*DD* or *DA*) are used to compute FRET efficiencies whereas photons after acceptor excitation (AA) are used to compute stoichiometry ratios. (**b**) Stoichiometry versus FRET maps for a sample containing four populations of molecules: donor-only molecules (green area), acceptor-only molecules (red area), and two FRET populations (white area). The maps are shown before (*left*) and after (*right*) correction of the photon counts, the FRET population is at $S = 0.5$

$$S = \frac{n'_{DD} + n'_{DA}}{n'_{DD} + n'_{DA} + \gamma_{PIE} n_{AA}}, \tag{3}$$

is computed for each molecule. Here, the first subscript indicates the excitation and the second subscript indicates the emission, i.e., n'_{DA} indicates the fully corrected acceptor photons after donor excitation. The parameter γ_{PIE} is a correction factor to account for the different excitation intensities for donor and acceptor. A two-dimensional plot of S versus E for all identified molecules now provides valuable information (Fig. 2b). For example, subpopulations of molecules with different FRET efficiency should show the same stoichiometry ratio if properly corrected. How is the correction performed? A detailed discussion of the method can be found in Hellenkamp et al. [44], which summarizes the work by the labs of Shimon Weiss [45, 47, 48] and Don Lamb [46]. In a first step, molecules are identified without any correction. A two-dimensional $E - S$ plot of the uncorrected FRET efficiencies and stoichiometry ratios immediately reveals the acceptor direct excitation probability and the cross-talk probability. For example, the population of molecules without an acceptor dye, often called donor-only molecules, should show a true mean FRET efficiency $\langle E \rangle = 0$. Yet, due to spectral cross-talk, some donor photons leak into the acceptor channel and the uncorrected mean FRET efficiency of the donor-only populations $\langle E_{\mathrm{raw}} \rangle$ (only corrected by background) will be located at $\langle E_{\mathrm{raw}} \rangle > 0$ (Fig. 2b, left). The cross-talk probability is then given by

$$\beta = \frac{\langle E_{\mathrm{raw}} \rangle}{1 - \langle E_{\mathrm{raw}} \rangle}. \tag{4}$$

Similarly, the effect of direct excitation of the acceptor by the donor excitation laser can be obtained. Acceptor-only molecules should have a mean stoichiometry $\langle S \rangle = 0$. Yet, due to a non-zero chance to directly excite the acceptor with the donor excitation laser, the uncorrected stoichiometry of acceptor-only molecules is at $\langle S_{\mathrm{raw}} \rangle > 0$ (Fig. 2b, left). The acceptor direct excitation coefficient is then given by

$$\mathrm{d} = \frac{\langle S_{\mathrm{raw}} \rangle}{1 - \langle S_{\mathrm{raw}} \rangle}. \tag{5}$$

Importantly, the quantity d is different from the quantity α used in the separate-sample method in Eq. 1. However, the two are directly related via

$$\mathrm{d} = \frac{\alpha}{1 - \alpha} \gamma_{PIE}. \tag{6}$$

The correction for cross-talk and acceptor direct excitation then gives a partially corrected acceptor photon number n_{DA} via

$$n_{DA} = n_{D1} - \beta n_{D2} - \mathrm{d} n_{A1}. \tag{7}$$

Again, the first subscript indicates the excitation source (D-excitation or A-excitation) and the second subscript indicates the emission, i.e., channel 1 or 2 with channel 1 being the acceptor channel. With the definitions $n_{DD} = n_{D2}$ and $n_{AA} = n_{A1}$, we can now compute stoichiometries and FRET efficiencies corrected for background, cross-talk, and acceptor direct excitation for each molecule. These partially corrected stoichiometries (S_p) and FRET efficiencies (E_p) depend on the true FRET efficiency (E) and the correction factors γ and γ_{PIE} via

$$\begin{aligned} E_p &= \frac{n_{DA}}{n_{DA} + n_{DD}} = \frac{E}{E + (1 - E)/\gamma} \text{ and } S_p = \frac{n_{DA} + n_{DD}}{n_{DA} + n_{DD} + n_{AA}} \\ &= \frac{E\gamma + (1 - E)}{E\gamma + (1 - E) + \gamma/\gamma_{PIE}}. \end{aligned} \tag{8}$$

From the resulting two-dimensional map (E_p, S_p), we obtain the two missing parameters γ and γ_{PIE} by combining Eq. 8 as

$$\langle S_p \rangle = \frac{1}{1 + \beta_{PIE}\left(\langle E_p \rangle + \gamma - \gamma \langle E_p \rangle\right)} \text{ with } \beta_{PIE} = \frac{1}{\gamma_{PIE}}. \tag{9}$$

Here, the bracket $\langle \ldots \rangle$ indicates an average over a subpopulation of molecules. Eq. 9 defines a curve in the $E_p - S_p$ map that intersects the mean position of a subpopulation of molecules. Yet, Eq. 9 has two unknown parameters (γ and γ_{PIE}), which means that their determination requires two positions (molecule populations) in the (E_p, S_p) map. The correction method is therefore only applicable for multimodal FRET histograms, i.e., histograms with at least two subpopulations of different FRET efficiencies (Fig. 2b). If this is not the case, a separate sample, e.g., a mixture of double-stranded DNA, labeled with donor–acceptor at different positions can be used as calibration sample to apply this procedure [44]. Once all correction parameters have been obtained, the fully corrected photon counts for each molecule can be obtained using Eq. 1.

3 Single-Molecule FRET: A Blurry Window into Molecular Disorder

3.1 Compaction and Expansion of Unfolded and Disordered Proteins Probed with smFRET

Ensemble fluorescence techniques, particularly when combined with fast mixing such as in a stopped-flow apparatus, provided the first hint that unfolded proteins compact after quickly diluting out chemical denaturants such as guanidinium chloride (GdmCl) or urea. Yet, at the time, it was unclear whether these so-called burst phases in protein folding experiments resembled a non-cooperative compaction of an unfolded protein or whether it was rather indicative of the fast formation of a structured, i.e., partially folded intermediate. The first clear indication that the dimension of an unfolded protein, measured in terms of the distance between chemically attached FRET donor (D) and acceptor (A) fluorophores, came in 2000 by the group of Shimon Weiss [13]. Using a confocal microscope, the authors monitored D-A labeled chymotrypsin inhibitor molecules while the molecules were randomly diffusing through the confocal volume of their microscope. The experiments revealed a clear bimodal distribution of FRET efficiencies $P(E)$ that included a peak resulting from folded molecules at high FRET and a peak from unfolded molecules at lower FRET efficiency. With the FRET efficiency being a proper molecular reaction coordinate (collective variable) in the sense that it is related to the distance between donor and acceptor dyes according to

$$E(r) = \frac{R_0^6}{R_0^6 + r^6}. \tag{10}$$

where R_0 is the Förster distance and r is the distance between D and A, the experiment provided the distribution of folded and unfolded molecules directly, without requiring fast mixing technologies. Importantly, a change in the concentration of GdmCl caused a change in the average FRET position of the unfolded peak, which indicated that the dimension of unfolded polypeptide chains responds to changing solvent conditions.

Given this result, one is tempted to equate the measured FRET histogram $P(E)$ with a Boltzmann distribution that reflects a potential of mean force $V(E)$ via $V(E) = -\ln P(E)$. Yet, nothing could be further from the truth. First, the average time it takes for a nanometer-sized molecule in water to diffuse through the confocal volume (burst time) is in the order of a millisecond. This means that conformational rearrangements faster than 1 ms will be averaged such that the resulting FRET efficiency E for a single molecule is in fact a time-averaged quantity. Second, the numbers of D- and A-photons (n'_D and n'_A) are only in the order of ~100, which causes a substantial uncertainty in the determination of FRET values of individual molecules. Hence, the width of a FRET-peak is not necessarily related to the width of the underlying distribution of DA distances. Instead, it is in large parts determined by shot noise. The expected width of the FRET histogram that results from shot-noise σ_{noise} can be estimated [49] via

$$\sigma_{\text{noise}}^2 = \left\langle E^2 \right\rangle - \langle E \rangle^2 = \langle E \rangle (1 - \langle E \rangle) \left\langle {n'}_0^{-1} \right\rangle \leq \langle E \rangle (1 - \langle E \rangle) n_T^{-1}. \tag{11}$$

Here, $n'_0 = n'_D + n'_A$ is the total number of photons obtained for a molecule, the bracket $\langle \ldots \rangle$ indicates the average over a population of molecules, and n_T is the photon threshold to identify a molecule. For example, with a typical threshold of 50 photons and a FRET-peak centered at $\langle E \rangle = 0.5$, one finds $\sigma_{\text{noise}} \leq 0.07$, which is a substantial contribution to the width of a FRET histogram. In the extreme case in which the interconversion of different conformations is much faster than the diffusion time of a molecule through the confocal spot, such as for IDPs and unfolded proteins, shot noise is often the dominating contributor to the width in the FRET histogram (Fig. 3a).

Contrary to the width, the average position of the FRET distribution $\langle E \rangle$ contains distance information, which has been used to quantify the extreme solvent-sensitivity of unfolded proteins. The first well-studied example of this sensitivity was given by the Eaton-lab, who demonstrated that the unfolded state of the cold shock protein (Csp*Tm*) compacts dramatically with decreasing concentrations of GdmCl [14, 50]. These experiments were confirmed with many other unfolded and disordered proteins in the following years [15–18, 28], indicating that an expansion or compaction of unfolded proteins with changing solvent conditions is a rather generic polymer effect. A quantification of this change in dimension in terms of average donor–acceptor distances is possible by expressing the mean FRET

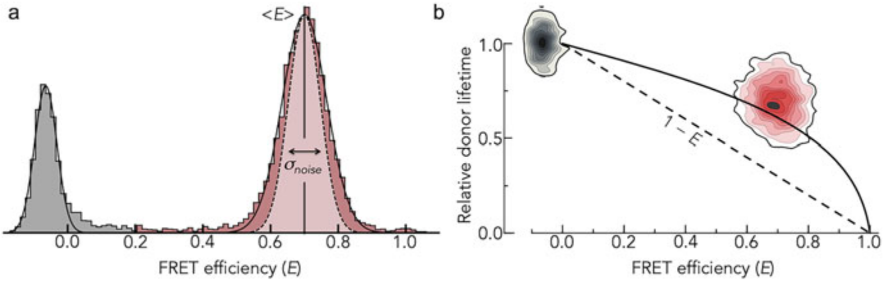

Fig. 3 Identifying conformational heterogeneity in IDPs. (**a**) SmFRET histogram of the intrinsically disordered DNA-binding domain of c-Myc. The histogram shows a single population of molecules (red) whose width is mainly determined by shot noise (dashed line). Molecules lacking an active acceptor are shown in gray. (**b**) 2D-map of the relative donor fluorescence lifetime τ_{DA}/τ_D and FRET efficiency for the data shown in (**a**). The dashed line is the prediction from Förster theory for the case in which all molecules have the same donor–acceptor distance (Eq. 18). The solid line is obtained for the case of a Gaussian distance distribution that is sampled at timescales much slower than the fluorescence lifetime (Eq. 19). Although the FRET distribution of c-Myc molecules is very narrow (**a**), conformational heterogeneity is identified in the lifetime-FRET map (**b**)

efficiency $\langle E\rangle$ as an average over the conformational distribution $P(r)$ of a polypeptide chain according to

$$\langle E\rangle = \int_a^L P(r)E(r)\mathrm{d}r. \tag{12}$$

The integration ranges from the closest possible donor–acceptor distance a to the maximum distance, which for a disordered polypeptide chain is given by its contour length $L = Nb$. Here, N is the number of peptide bonds and b is the distance between two consecutive C_α-atoms ($b = 0.38$ nm). Clearly, the distribution $P(r)$ cannot be recovered solely from the mean FRET efficiency $\langle E\rangle$ such that suitable analytical models for $P(r)$ are required to extract distance information. A variety of models have been used in the past [23], the simplest of it being the Gaussian chain model given by

$$P(r) = \frac{4}{R}\left(\frac{3}{2\pi}\right)^{1/2}\left(\frac{r}{R}\right)^2 \exp\left[-\left(\frac{r}{R}\right)^2\right]. \tag{13}$$

Here, $R = \sqrt{2/3}b_K N_K^{1/2} = \sqrt{2/3}\langle r^2\rangle^{1/2}$ is the most likely donor–acceptor distance, a parameter that can be obtained from $\langle E\rangle$ by solving Eqs. 12 and 13 numerically. The parameters b_K and N_K are the length and number of virtual Kuhn segments that are related to the real bond length b and amino acid number N via

$$b_K = 2l_p \text{ and } N_K = \frac{b}{b_K}N, \tag{14}$$

with l_p being the persistence length (stiffness) of the chain. Persistence lengths determined from smFRET experiments using Eqs. 12, 13, and 14 are ~0.4 nm at physiological condition [30], which roughly agrees with results from other methods, e.g., see [20] and references therein. Importantly, more recent experiments and simulations [22, 25, 51] found that the Gaussian chain model (Eq. 13) overestimates chain dimensions from smFRET experiments when used in Eq. 12 and a more realistic self-avoiding random walk (SAW) model

$$P(r) = A\frac{4\pi}{R}\left(\frac{r}{R}\right)^{2+g}\exp\left[-B\left(\frac{r}{R}\right)^{\delta}\right], \tag{15}$$

is better suited to extract mean donor–acceptor distances from smFRET experiments on IDPs and unfolded proteins. In Eq. 15, the parameters A and B are obtained from the conditions $1 = \int_0^\infty P(r)\mathrm{d}r$ and $R^2 = \int_0^\infty r^2 P(r)\mathrm{d}r$, and the exponents are given by $g \approx 1/6\nu$ and $\delta = \frac{1}{1-\nu}$ [51–53]. The quantity $\nu = (\ln R - \ln b_K)/\ln N_K$ is the length-scaling exponent of the chain, a parameter that we will discuss in more detail in the following section. Although Eq. 15 is strictly correct only for expanded homopolymers (good solvent) [54], a direct comparison with molecular simulations indicated that it is also a good approximation for more compact chains [51].

Although smFRET histograms do not contain direct information on the width of the distribution $P(r)$ of donor–acceptor distances r in a disordered chain, the fluorescence lifetime of the donor contains this information. By combining confocal single-molecule detection with time-correlated single-photon counting (TCSPC) using pulsed excitation sources, fluorescence lifetimes of both donor and acceptor can easily be obtained. For a distribution of distances $P(r)$, the fluorescence lifetime decay of the donor is given by

$$I(t) = I_0\int_0^\infty P(r)\, e^{-[k_D + k_T(r)]t}\mathrm{d}r, \tag{16}$$

where $\tau_D = 1/k_D$ is the fluorescence lifetime of the donor in the absence of an acceptor, $k_T(r) = k_D(R_0/r)^6$ is the rate of transferring energy to the acceptor. For the hypothetical case in which all molecules have the same donor–acceptor distance R, i.e., $P(r) = \delta(r - R)$ and $E(R) = \langle E \rangle$,[1] Eq. 16 gives

$$I(t) = I_0\, e^{-k_D\left[1+(R_0/R)^6\right]t} \tag{17}$$

and the average donor fluorescence lifetime is given by

---

[1] Here, δ indicates the Dirac delta function.

$$\langle \tau_{DA} \rangle = \int_0^\infty t\, I(t) dt \Big/ \int_0^\infty I(t) dt = k_D^{-1} \left[1 + (R_0/R)^6 \right]^{-1} = \tau_D (1 - \langle E \rangle). \tag{18}$$

Hence, the donor fluorescence lifetime depends linearly on the FRET efficiency, which is the classical prediction from Förster theory (Fig. 3b). Importantly, the same result is obtained for more complicated distance distributions, as long as the dynamics, i.e., the sampling of this distribution and therefore the fluctuations of the energy transfer rate k_T, are much faster than the intrinsic fluorescence lifetime of the donor τ_D [55]. Yet, this is not the case for IDPs and unfolded proteins that typically reconfigure at timescales of tens to hundreds of nanoseconds (Sect. 4), i.e., significantly slower than the fluorescence lifetimes of typical dyes (1–4 ns). In case of such slow sampling, we can compute the static average over the distance distribution $P(r)$ and obtain

$$\langle \tau_{DA} \rangle = \int_0^\infty t\, I(t) \mathrm{d}t \Big/ \int_0^\infty I(t) \mathrm{d}t = \int_0^\infty P(r)\, \tau_{DA}(r)^2 \mathrm{d}r \Big/ \int_0^\infty P(r) \tau_{DA}(r) \mathrm{d}r \tag{19}$$

with

$$\tau_{DA}(r) = k_D^{-1} \left[1 + (R_0/r)^6 \right]^{-1}.$$

A plot of the average donor fluorescence lifetime $\langle \tau_{DA} \rangle$ as a function of the mean FRET efficiency $\langle E \rangle$ will not follow a simple linear scaling (Fig. 3b). In fact, substituting $\tau_{DA}(r) = \tau_D[1 - E(r)]$, we can express Eq. 19 in different form

$$\langle \tau_{DA} \rangle = \tau_D (1 - \langle E \rangle) + \tau_D \frac{\sigma_E^2}{1 - \langle E \rangle}, \tag{20}$$

with $\sigma_E^2 = \langle E^2 \rangle - \langle E \rangle^2$ and $\langle E^2 \rangle = \int E(r)^2 P(r) \mathrm{d}r$, given that the distance distribution is properly normalized such that $\int P(r) \mathrm{d}r = 1$. Notably, the term σ_E^2 should not be confused with the width of the FRET histogram, which is largely affected by shot noise as explained above. It rather is the width of the "true" FRET distribution, which is typically inaccessible in smFRET experiments. Hence, a model for $P(r)$ is required to obtain information about the width of the distance distribution. Since the second term in Eq. 20 is always positive, distance fluctuations quantified by $P(r)$ will increase the average donor fluorescence lifetime compared to the prediction given by Förster theory (Eq. 18, Fig. 3b). Strong deviations from the classical Förster prediction have indeed been observed for a multitude of disordered and unfolded proteins [23, 30, 33] such that a comparison of experimental donor lifetimes with the Förster prediction in Eq. 18 can generally be used to identify distance heterogeneity in proteins [55, 56]. Importantly, these relationships only hold under the condition that the orientation of donor and acceptor dye dipoles averages sufficiently at timescales faster than the fluorescence lifetimes of the dyes. If this is not the case, the relationship between donor–acceptor distances and measured FRET efficiencies is more complicated. For example, the FRET efficiency will now depend on both, the distance and the orientation of the dyes via [57]

$$E\left(r,\kappa^2\right)=\left[1+\frac{2}{3\kappa^2}\left(\frac{r}{R_0}\right)^6\right]^{-1}\ \text{with}\ \kappa^2=\left(\cos\theta_T-3\cos\theta_D\cos\theta_A\right)^2. \tag{21}$$

Here, θ_T is the angle between donor and acceptor dipole, whereas θ_D and θ_A are the angles between these dipoles and the vector that connects both dipoles. While the value of κ^2 can range from 0 to 4, it will fluctuate in general such that several averaging regimes have to be considered. Assuming an isotropic orientation of the dye dipoles, the distribution of κ^2 is given by [57, 58]

$$p\left(\kappa^2\right)=\begin{cases}\dfrac{1}{2\sqrt{3\kappa^2}}\ln\left(2+\sqrt{3}\right) & 0\leq\kappa^2\leq1\\ \dfrac{1}{2\sqrt{3\kappa^2}}\ln\left(\dfrac{2+\sqrt{3}}{\sqrt{\kappa^2}+\sqrt{\kappa^2-1}}\right) & 1<\kappa^2\leq4\end{cases}. \tag{22}$$

In the most commonly used limit, the fluctuation of κ^2 is much faster than the fluorescence lifetime of the dyes and an average value for κ^2 obtained from Eq. 22 via $\langle\kappa^2\rangle = \int\kappa^2 p(\kappa^2)\mathrm{d}\kappa^2 = 2/3$ can be used. In this limit, the FRET efficiency (Eq. 21) simplifies to Eq. 10. Yet, exceptions have been found for cases in which the dyes interact with extended protein surfaces [59, 60], thus hampering the fast averaging of dipole orientations of the dyes. Fluorescence anisotropy experiments, ideally in a time-resolved manner, should be used to check the applicability of the $\langle\kappa^2\rangle = 2/3$ limit. For intrinsically disordered and unfolded proteins, the fast averaging limit is often fulfilled due to the absence of extended protein surfaces that would facilitate dye–protein interactions.

In summary, the combination of mean FRET efficiencies and fluorescence lifetimes does not only provide information on average chain dimensions, but it also allows an estimate of the width of distance distributions. Yet, analytical models of these distributions such as the SAW model (Eq. 15) are still required to retrieve this information from smFRET experiments. Similarly, experiments alone are insufficient to understand another important property of IDPs, their extreme sensitivity toward changes in external conditions such as denaturants [13–18, 28], temperature [21, 32], crowding [31], or ionic strength [7, 18, 19, 26]. Clearly, a change in any of these conditions will unambiguously affect the balance of attractive and repulsive interactions within the chain, thus resulting in altered chain dimensions. To obtain a more quantitative understanding of the interactions that drive such compactions or expansions, we will extend the Gaussian chain model (Eq. 13) to generate a rather general mean-field homopolymer model.

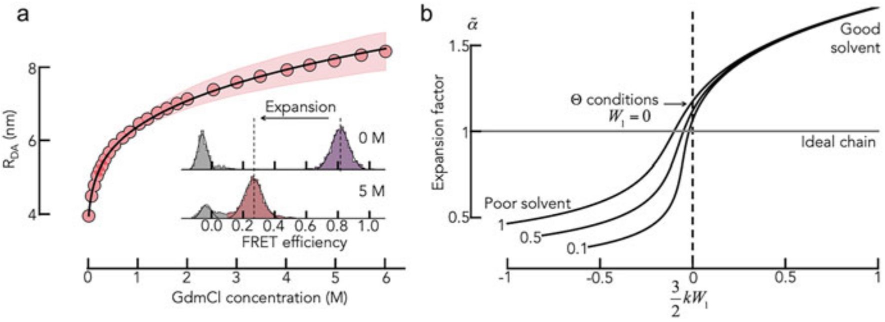

Fig. 4 Solvent-induced expansion of an IDP. (**a**) Average donor–acceptor distances (R_{DA}) of the disordered Max protein as function of the GdmCl concentration. The solid line is a fit with a polyampholyte theory [26]. *Inset:* SmFRET histograms of Max at two concentrations of the denaturant GdmCl (indicated). The FRET population shifts to lower FRET values, indicating an expansion of the chain. (**b**) Coil-to-globule transition as predicted by the mean-field theory of de Gennes (Eq. 25) for $N_K = 100$ and three different values (indicated) of the three-body interaction parameter ($3k^2W_2$)

3.2 *Mean-Field Homopolymer Theory*

The fact that polymers expand and contract in a solvent-dependent manner is not a new observation. In fact, plenty of examples have been studied in the twentieth century, the most well-known being poly-styrene in cyclohexane [61]. Although chemically tremendously more complex than homopolymers, IDPs exhibit a similar pliability to solution changes (Fig. 4a) and homopolymer theories have therefore been used in the past to model them. Clearly, the results should be interpreted with care. However, it is undeniable that homopolymer models have greatly advanced our understanding of disordered and unfolded proteins not only by providing simplified parameters to describe the molecular processes behind chain expansions or collapse but also by spotting deviations from homopolymer properties.

To model the expansion and compaction of polymers, let us start with the Gaussian chain model in Eq. 13. To simplify notation, we first express Eq. 13 in terms of a reduced distance coordinate $\alpha^2 = \left(\frac{r}{R}\right)^2$, where α is also known as the expansion factor. Keep in mind that R is the most likely donor–acceptor distance of a Gaussian chain, i.e., of an ideal mathematical chain without interactions and volume. Neglecting constant prefactors, Eq. 13 can be re-written as

$$P(\alpha) \propto \alpha^2 \exp\left(-\alpha^2\right). \tag{23}$$

As any molecular property that underlies the statistical laws of thermodynamics, the distance distribution is associated to a free energy profile ($P(\alpha) \propto e^{-F(\alpha)}$), which is given by $F(\alpha) = -\ln P(\alpha) \propto \alpha^2 - 2\ln\alpha$. Since our chain model at this point is just a mathematical chain of points connected by vectors, i.e., it lacks volume or any type of interactions, the free energy $F(\alpha)$ only describes the elastic entropy of the chain that results from the multiplicity of random conformations of this ideal chain. To introduce interactions of any type between monomers, repulsive or attractive, we will use the mean-field approximation in which the number of interactions within a chain is proportional to the density (volume fraction) $\phi = \mathrm{d}^3 N_K / \alpha^3 R^3$ of monomers in the chain. Here, d is the size (radius) of a single monomer in the chain. Thus, a reasonable extension of the free energy of an ideal chain by interactions would be given by a virial expansion of the type

$$F(\alpha) \propto \alpha^2 - 2\ln\alpha + N_K\left(W_1\phi + W_2\phi^2 + \ldots\right), \tag{24}$$

with the virial coefficients W_1 for two-body interactions and W_2 for three-body interactions. To simplify the expression, we will only consider the first two terms of the expansion, which already provide a rather general model that has first been proposed (in slightly different form)[2] by de Gennes in 1975 [62]. Finding the free energy minimum $dF/d\alpha = 0$ gives

$$\tilde{\alpha}^5 - \tilde{\alpha}^3 - \frac{3k^2 W_2}{\tilde{\alpha}^3} = \frac{3}{2} k W_1 \sqrt{N_K} \text{ with } k = (d/b_K)^3. \tag{25}$$

Here, $\tilde{\alpha}$ indicates the expansion factor at the free energy minimum. Notably, Eq. 25 can also be obtained using variational approaches [26, 63, 64]. A plot of $\tilde{\alpha}$ as function of the two-body interaction energy clearly uncovers a substantial compaction of the chain with decreasing (attractive) W_1 and positive (repulsive) W_2 (Fig. 4b). Contrary to folding/unfolding transitions, this compaction is non-cooperative [62, 65] and rather resembles a higher-order phase transition compared to the cooperative first-order like folding–unfolding transition of proteins. However, as pointed out by de Gennes [62], Eq. 25 also predicts a first-order coil-to-globule transition for homopolymers for sufficiently small values of W_2, a finding that has so far not been confirmed experimentally and that is likely an artifact of the theory. It is instructive to analyze the individual energetic contributions to the two-body interaction term W_1, often called χ, which can be expressed as a Flory-Huggins interaction parameter [66]

$$W_1 \propto w_{sp} - \left(w_{ss} + w_{pp}\right)/2, \tag{26}$$

---

[2] De Gennes derived his model for the radius of gyration (r_G) of the polymer. Since the distribution of radii of gyration for a Gaussian chain is not known in closed analytic form, he used the approximation $P(\alpha) \propto \alpha^3 \exp(-3\alpha^2/2)$ where $\alpha = r_G/R_{G,\text{ ideal}}$. Notably, compared to donor–acceptor distances measured with smFRET, the radius of gyration is by far the better quantity to construct a mean-field theory due to its direct link to the monomer-density of a chain. The long-known fact has more recently gained renewed attention in the so-called FRET-SAXS controversy [see Refs. 22 and 77].

where the energy changes w_{pp}, w_{ss}, and w_{sp} are due to polymer–polymer (pp), solvent–solvent (ss), and solvent–polymer contacts (sp), respectively. Notably, Eq. 26 had been derived for lattice-polymer models and processes such as the adsorption of denaturant molecules on the polypeptide chain might not be adequately captured by it. Hence, an awareness of these pitfalls is required when interpreting experimental results in terms of w_{pp}, w_{ss}, and w_{sp}. The exact molecular effects of GdmCl and urea on polypeptide chains, which might affect multiple interactions simultaneously, have been debated much in the past [67]. A combination of polar interactions, hydrogen bonding, and stacking to non-polar surfaces has been suggested. Yet, it is unclear whether any of these contributions really dominate the action of GdmCl and urea. Based on experimental transfer free energies (Δg_T) of amino acids from water to solutions of these denaturants, it is clear that both denaturants assist in the solvation of non-polar side chains [67, 68]. In fact, the molecular transfer model developed by Thirumalai and co-workers [69, 70] explicitly describes the change in dimension of unfolded proteins and the thermodynamic stability of folded proteins based on the experimentally determined transfer free energies of amino acids. In addition, smFRET experiments on a broad variety of IDPs and unfolded proteins suggest a clear relationship between Δg_T and W_1 [20, 21, 71, 72]. Whereas the approximate values of W_1 between high molar concentrations of denaturants and water are known to be in the order of a few k_BT [20, 26], absolute values for the three-body interaction term W_2 in Eq. 25 are more difficult to obtain. Recent results suggest values in the order of 4–30 k_BT for the term $3k^2W_2$ [26]. Yet, care has to be taken as these energies also depend on the precise numerical factors used in Eq. 25 and mean-field theories alike. In fact, a more independent parameter to quantify the compaction of disordered proteins (and polymers in general) is the length-scaling exponent ν.

For sufficiently long chains, the dimension scales with its length according to $R_{DA} \propto N_K^{\nu}$, where $R_{DA} = \tilde{\alpha} R$ is the equilibrium distance between donor and acceptor that can be determined in smFRET experiments (see Eqs. 12, 13, 14, and 15). The length-scaling exponent ν can be obtained from Eq. 25 for three limits: (1) swollen chains in good solvent, i.e., W_1 is repulsive ($W_1 > 0$) and three-body interactions can be neglected ($W_1 \ll W_2$), (2) compact chains in poor solvents, i.e., W_1 is strongly attractive ($W_1 < 0$) and W_2 is repulsive ($W_2 > 0$), and (3) conditions in which $W_1 \approx 0$ and $W_2 > 0$ such that the chain behaves like an ideal chain, a condition that is also called Θ-condition (Fig. 4b). Let us start with an expanded chain in good solvent. At such conditions, the effect of three-body interactions is substantially reduced and we can safely assume $W_2 \approx 0$. Equation 25 then simplifies to the well-known Flory result [73]

$$\tilde{\alpha}^5 - \tilde{\alpha}^3 = \frac{3}{2} k W_1 \sqrt{N_K}. \tag{27}$$

In the extreme limit of expanded chains ($\tilde{\alpha}^5 \gg \tilde{\alpha}^3$) and with the definitions for $\tilde{\alpha}$ and R, we obtain

$$R_{DA} = C_e b_K N_K^{3/5} \propto N_K^{\frac{3}{5}} \text{ with } C_e = \left(\frac{2}{3}\right)^{\frac{1}{2}} \left(\frac{3kW_1}{2}\right)^{\frac{1}{5}}. \tag{28}$$

Hence, polymers in good solvent should exhibit a length-scaling exponent of $\nu = 3/5$. In fact, the result is surprisingly close to the exact value from renormalization-group field theory of $\nu = 0.588 \pm 0.001$ [74]. Notably, this success of Flory theory is based on a fortuitous cancelation of errors: the term $W_1 N_K \phi$ in Eq. 24 (first term in brackets) overestimates the monomer interaction energy because correlations between monomers along the chain are neglected [75] and the term $\alpha^2 - 2 \ln \alpha$ (Eq. 24) of the Gaussian chain model overestimates the elastic energy of the chain.

A number of experimental studies using a diverse set of methods from SAXS [22, 27, 76, 77], smFRET [20, 22, 27], over NMR [78] to FCS (fluorescence correlation spectroscopy) [20, 22], and molecular simulations [22, 27, 51] demonstrated that IDPs and unfolded proteins at high concentration of denaturants (GdmCl and urea) indeed exhibit a length-scaling exponent of ~3/5. Hence, sequences and amino acid compositions are less relevant for chain dimensions under these conditions. This is particularly the case for proteins in GdmCl, which is not only a denaturant but also a salt, thus capable of effectively screening charge–charge interactions in unfolded proteins and IDPs as shown with smFRET on a number of systems [18, 19, 26].

The second limit that can be obtained from Eq. 25 is that of a strongly collapsed chain in poor solvent. Under these conditions ($\tilde{\alpha}^5 << \tilde{\alpha}^3 >> 1$, and $W_1 < 0$), Eq. 25 simplifies to

$$\tilde{\alpha}^3 = \frac{2kW_2}{|W_1|} N_K^{-1/2}, \tag{29}$$

which gives

$$R_{DA} = C_g b_K N_K^{1/3} \propto N_K^{\frac{1}{3}} \text{ with } C_g = \left(\frac{2}{3}\right)^{\frac{1}{2}} \left(\frac{2kW_2}{|W_1|}\right)^{\frac{1}{3}}. \tag{30}$$

Hence, if the chain forms a highly collapsed globule due to strong polymer–polymer contacts, the growth in dimension with chain length is tremendously reduced compared to the excluded volume limit (Eq. 28). It should be mentioned that although Eq. 30 follows naturally from Eq. 25, it is only valid for mean-field theories constructed in terms of radii of gyration. For a distance such as that measured with smFRET (R_{DA}), the volume fraction of a chain ϕ is ill-defined in the limit of compact chains and the chain volume cannot be reliably estimated from the end-to-end distance. In fact, the de-coupling of distances R_{DA} inferred from smFRET and radii of gyrations R_G for compact chains has been studied in detail [27] and shows that chains with similar R_G can obey largely different FRET efficiencies. Hence, scaling exponents of extremely compact chains cannot be extracted from donor–acceptor distance measurements alone but require either the additional help of molecular simulations or the direct determination of better quantities for chain dimensions such as the radius of gyration [77, 79] or the hydrodynamic radius of a chain [80, 81]. So far, only a few cases of disordered polypeptide chains with scaling exponents close to 1/3 have been published, one being poly-glutamine peptides [80]. Non-equilibrium SAXS experiments using fast continuous-flow mixing also identified a scaling exponent close to 1/3 for early folding intermediates of seven proteins [82]. It is unclear however to which extent the low scaling exponent is affected by structure formation in those intermediates after being diluted from high concentrations of denaturant.

Finally, we discuss the limit at which w_{pp}, w_{ss}, and w_{sp} balance such that $W_1 = 0$. At this Θ-condition, the chain behaves like an ideal chain and only three-body interactions (W_2) and chain elasticity, i.e., configurational entropy, contribute to the chain dimension. Importantly, since $W_2 > 0$ at Θ-conditions, the absolute donor–acceptor distance will not be identical to that of an ideal chain ($\tilde{\alpha} \neq 1$) but rather be close to it ($\tilde{\alpha} \sim 1$). To see this, we start with Eq. 25 and set $W_1 = 0$, which gives

$$\tilde{\alpha}^8 - \tilde{\alpha}^6 = 3k^2 W_2. \tag{31}$$

To obtain an approximate solution, we expand the polynomial (LHS) around $\tilde{\alpha} = 1$, which leads to the series $2(\tilde{\alpha} - 1) + 13(\tilde{\alpha} - 1)^2 + 36(\tilde{\alpha} - 1)^3 + \ldots$. Since $\tilde{\alpha} - 1 < 1$,we only keep the leading term, which admittedly is a brave assumption given the significant pre-factor of the second term, resulting in

$$R_{DA} = C_\Theta b_K N_K^{1/2} \propto N_K^{\frac{1}{2}} \text{ with } C_\Theta = \left(\frac{3}{2}\right)^{\frac{1}{2}} \left(1 + \frac{3}{2} k^2 W_2\right), \tag{32}$$

which exceeds the distance of the ideal chain that would be given by $R = \sqrt{2/3} b_K N_K^{1/2}$. Indeed, a large number of smFRET and SAXS experiments have demonstrated that the scaling exponents of IDPs and unfolded proteins under physiological conditions (absence of denaturants) cluster around $\nu = 1/2$ [20, 22, 29, 77]. These results indicated that polypeptide chains in water are surprisingly close to Θ-conditions, i.e., conditions at which attractive and repulsive interactions in the chain roughly balance. Given the rather small number of proteins tested so far, it is unclear whether this finding is pure coincidence or indeed the result of evolution. In fact, it has been speculated that the balance of attractive and repulsive interactions at Θ-conditions could be advantageous for protein folding reactions, e.g., by allowing a more efficient sampling of intra-chain contacts during folding [20, 83]. Clearly this interpretation is difficult to draw for IDPs that do not necessarily fold, yet, Θ-conditions might still be advantageous for liquid–liquid phase separation processes. Although the scaling exponents for many polypeptide chains are close to 1/2, differences in sequence composition, but even in the precise sequence such as patterning of charges [84–86] or hydrophobic and aromatic residues [87] can cause significant variations in the size of polymers [88, 89]. For example, a nearly balanced number of positively and negatively charged amino acids can strongly compact IDPs [26]. On the opposite, high net-charges cause strong electrostatic repulsions that lead to extremely expanded chains [19]. Naturally, such charge-driven interactions are sensitively affected by ionic strengths, and even small variation within the physiological regime (100–300 mM) can cause large changes in chain dimensions and scaling exponents. Many studies over the past decades explicitly incorporated the effect of charge interactions in mean-field theories [90–93] and their application in interpreting salt-induced compaction of highly charged IDPs and the expansion of nearly charge-balanced IDPs found with smFRET has been enormously successful [19, 26]. Yet, a drawback of these theories is their inability to account for specific charge patterns along the sequence. In fact, mean-field polyelectrolyte and polyampholyte theories are only applicable for well-mixed sequences in which it can be safely assumed that amino acid charges are essentially smeared out across the chain. Although attempts have been made to develop analytical heteropolymer theories for quenched sequences [94], modern approaches to account for sequence specificity in the polymer behavior of IDPs and unfolded proteins are mainly based on molecular simulations.

3.3 More Accurate Polymer Models: Combining smFRET with Molecular Simulations

Mean-field theory is a convenient first option to understand smFRET data of IDPs. However, local secondary structure preferences and long-range interactions caused by either charged amino acids or hydrophobic patches in the sequence might have significant populations and lifetimes. In such cases, homopolymer theories have limited capabilities of faithfully interpreting FRET-based distance information.

To combine smFRET data with molecular simulations, one often relies on the accuracy of the simulation model, which is described by a force field that accounts for interactions between amino acids. Since force fields were often parameterized using data from a variety of IDPs, i.e., not necessarily for the IDP studied in a particular smFRET experiment, there is no guarantee that a simulation will reproduce a FRET signal out of the box, thus complicating an interpretation of physical mechanisms with the model. To ease mechanistic interpretations of smFRET experiments therefore requires a molecular ensemble that can reproduce the smFRET data. One strategy has been to reweight conformations generated with an existing simulation model such that the calculated observables (e.g., FRET efficiency) best match the experimental values [22]. A tremendous advantage of ensemble reweighting is its low computational cost compared to the strategy of adapting parameters using iterative simulations to match experimental outcomes. However, reweighting also has its limits. For instance, if the initial ensemble obtained from a simulation deviates significantly from the experimental measurement, or in other words, if important conformations are not sampled in the simulation, reweighting will not provide a realistic description of the IDP ensemble. Moreover, the outcome from ensemble reweighting methods could be biased by the initial ensemble, which depends on the quality of the force field [95]. In such cases one has to downgrade the expectation from bottom-up simulation models with predictive power and instead either bias the simulations using experimental data [96] or adapt the simulation parameters such that the simulated trajectory can still interpret the experimental data [97]. Further cross-validation using experimental inputs other than smFRET, such as solvation free energy, SAXS, and NMR, is then necessary to verify the refined model. In the following, we will therefore mostly focus on those molecular simulation methods that can be easily integrated with smFRET experiments and do not discuss bottom-up simulation methods.

Depending on the spatial resolution of the model, all-atom implicit/explicit-solvent simulations [97, 98] and coarse-grained simulations [99] have been used in the past to describe smFRET data. Clearly, all-atom explicit-solvent models have the highest resolution but can be limited by force field accuracy and sampling efficiency. Taking advantage of the increasing amount of experimental measurements on IDPs, significant efforts have been undertaken in recent years to improve all-atom force fields [97, 100, 101] such that modern force fields will provide meaningful IDP ensembles. In terms of sampling, most software packages [102–105] have now taken advantage of the rapidly improving GPU resources such that microseconds simulations of an IDP with less than 100 amino acids can be achieved within few days in standard high-performance computing clusters. In addition, the development of specialized supercomputers for molecular dynamics such as “Anton” even enabled IDP-simulations of 100 μs in length within few days [106], which even allows investigations of IDP–ligand interactions [107].

Using all-atom simulations to interpret smFRET experiments requires modeling of the dyes. In fact, commonly used dyes such as Alexa Fluor 488 and Alexa Fluor 594 have significant sizes of a few amino acids (Fig. 5a). The simulation has to either model dyes explicitly or specific correction factors have to be included to calculate FRET efficiencies from simulations without dyes. Using a carefully parameterized force field of FRET dyes, Best et al. showed that the orientational factor $\langle \kappa^2 \rangle$ is indeed close to 2/3, as estimated in many experiments [108] and FRET efficiency calculations can be simplified using Eq. 10 (see Sect. 3.1). However, to improve the sampling efficiency, dyes are not commonly included. In such cases, only distances between the C_α-atoms of the labeled residues are accessible from the simulations. These distances should then be rescaled by a factor of $[(N + N_{\mathrm{linker}})/N]^{\nu}$, where N is the number of peptide bonds between the labeled amino acids and ν is the length-scaling exponent which can be determined by the scaling of internal chain distances with sequence separation in the simulation [109] or from experiments with multiple labeling positions [20]. Importantly, N_{linker} is a free parameter that represents the effective length of both dye linkers in terms of an equivalent number of peptide bonds. It has experimentally been estimated to be ~9 [29, 30], which agrees well with estimates obtained from all-atom simulations with and without dyes [110] (Fig. 5b).

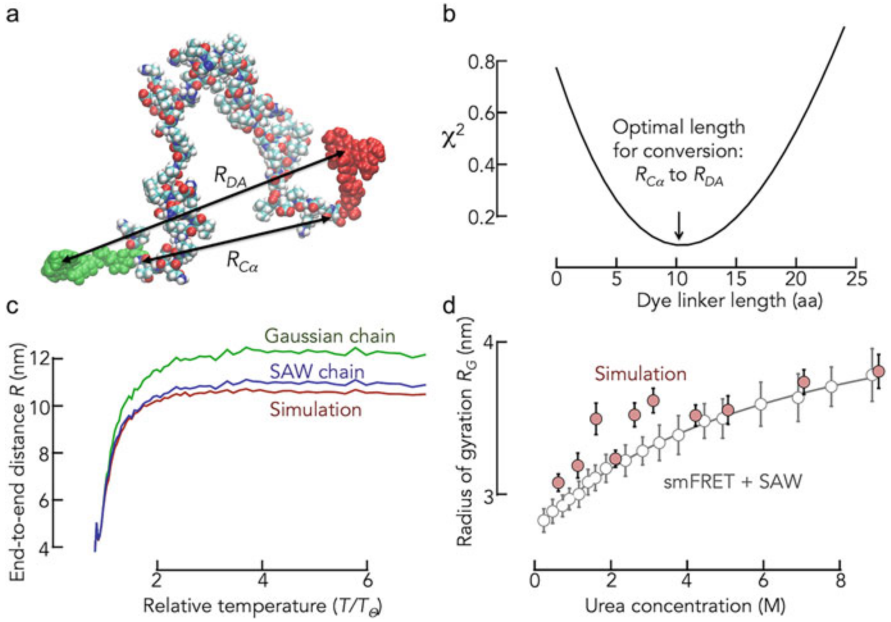

Fig. 5 Comparing smFRET experiments with molecular simulations. (**a**) One representative conformation of a simulation with the IDP ACTR, including the dyes Alexa Fluor 488 and Alexa Fluor 594, is shown. Two different ways of calculating donor–acceptor distances for comparison with smFRET experiments are depicted: R_{DA} between the donor and acceptor for simulations that explicitly include the dyes and $R_{\mathrm{C\alpha}}$ between the C_α atoms of the labeled residues for simulations without dyes. (**b**) Deviation between the FRET efficiencies calculated using R_{DA} and $R_{\mathrm{C\alpha}}$ as a function of the linker length N_{linker} (in amino acids) used for correcting FRET values from simulations without dyes [110]. (**c**) End-to-end distances from all-atom simulations (red) are compared with the distances obtained from mean FRET values that were computed from the same simulation. Two polymer models, the Gaussian chain (Eq. 13) and the SAW model (Eq. 15) were used to compute the distances from FRET using Eq. 12. (**d**) The radius of gyration R_{G} obtained from smFRET using SAW model is compared to an all-atom ensemble reweighted using both smFRET and SAXS data

Importantly, the heterogeneous ensemble of IDP conformations is better characterized by different experimental techniques rather than smFRET alone and all-atom simulations have become essential in integrating them, not only to generate realistic IDP ensembles but also to identify and reconcile discrepancies between findings with different experimental methods. For instance, while smFRET experiments showed that unfolded and disordered polypeptide chains expanded substantially upon addition of denaturants, SAXS experiments suggested that expansions were much less pronounced [111]. All-atom simulations were then used as benchmarks to test empirical ways of using polymer models for analyzing the smFRET and SAXS data (see Sect. 3.1) by comparing artificial experimental signals, e.g., FRET efficiencies and SAXS scattering curves together with the correct “answers,” i.e., donor–acceptor distance and radius of gyration (R_G) that could be directly calculated from the simulation. This comparison showed that the Gaussian chain model overestimates distances in smFRET experiments whereas Guinier analysis underestimates R_G when the chain deviates from a random coil [22]. A SAW model (Eq. 15), in which the distance distribution is adjusted according to the scaling exponent, has been found to be a better choice for interpreting smFRET and SAXS data for ν larger than 0.5 [51, 112] (Fig. 5c). In addition, simulations [51] also verified that a size-dependent factor λ [113] is necessary to convert the average donor–acceptor distance R_{DA} from smFRET into R_G

$$\lambda = \frac{R_{DA}^2}{R_g^2} = \frac{2(\gamma + 2\nu)(\gamma + 2\nu + 1)}{\gamma(\gamma + 1)}. \tag{33}$$

Here, $\gamma = 1 + g/\nu \approx 7/6$ as in Eq. 15. Equation 33 shows that the estimated conversion factor ranges between 4.1 and 6.3, depending on the chain compaction specified by the length-scaling exponent ν, which added to the deviations found between smFRET and SAXS measurements [27]. With both corrections, the radius of gyration estimated from smFRET experiments was comparable with that estimated from an ensemble using both FRET and SAXS measurement for the unfolded R17 protein at different denaturant concentrations [22, 51] (Fig. 5d). Integrated with molecular simulations, smFRET measurements have also been found to complement other experimental methods such as NMR and PRE (Paramagnetic Relaxation Enhancement) [114], thus showing that all-atom simulations provide an effective way of integrating multiple sources of experimental data.

Although all-atom simulations certainly provide the most accurate representation of molecular processes, particularly when combined with explicit water models, sampling can be time consuming and resource demanding for IDPs with more than 100 residues. It is often informative to use a coarser representation of the molecular complexity at substantially lower computational cost. An ideal “interpolation” between all-atom simulations and simple polymer models are low-resolution coarse-grained models. Coarse-grained models contain about the same number of free parameters as analytical polymer theories but additionally include specific sequence details of IDPs such as the patterning of charges and hydrophobic sequence patches that are absent in mean-field theories. The fast sampling of coarse-grained models even allows a quick adjustment to experimental data and can be used to generate multiple IDP ensembles at a variety of solvent conditions. For instance, in a recent study of the disordered cytoplasmic tail of E-cadherin [12], smFRET experiments of differently labeled variants of E-cadherin showed that a mean-field polyampholyte theory [90] failed to predict the salt-dependent conformational change of all variants. A description using a coarse-grained model instead reproduced the salt-dependence of all variants with just one global fitting parameter and identified the reason for the failure of the mean-field description. The segregation of oppositely charged residues along the sequence, i.e., charge patterning, was found to be essential for the dimension of this IDP, an effect that cannot be captured with mean-field theories. Finally, systems involving IDP assemblies such as in liquid–liquid phase separation (LLPS) processes [115] or already only those that include one IDP in complex with another macromolecular ligand [116] typically exceed the sampling-capabilities of all-atom simulations, thus making coarse-grained models unavoidable.

Clearly, the level of coarse-graining depends on the specific task and the balance between the computational efficiency and modeling accuracy. Coarse-graining levels range from representing individual residues by multiple beads over one bead per residue down to one bead for several residues. Yet, the need of calculating distances between labeled residues in describing smFRET experiments clearly favors one bead per residue models. Coarse-graining at the single-residue level is particularly important if sequence-specific effects play an important role in the system at hand. To capture electrostatic interactions at different ionic strength, which is a relevant factor considering physiological salt concentrations of 100–300 mM [117], the Debye-Hückel approximation is commonly used [118]. The computational efficiency of such a screening potential in contrast to an explicit ion model often outweighs the disadvantage of neglecting the true radial distribution of ions around the charged amino acids in studies of IDPs [119, 120]. In addition, other interactions between every pair of amino acids also add up to significant factors that can sensitively affect the dimension of an IDP. In a recently developed coarse-grained model (HPS model) [99] such pairwise interactions were introduced using the amino acid hydropathy in the Ashbaugh-Hatch functional form [121]. It was shown that a tuning of the strength of this amino acid hydropathy described the experimental results of a variety of IDPs with high accuracy [99, 122–125]. Further tuning the pairwise interaction strengths to match smFRET experiments of a specific IDP is then a straightforward way to understand the main sequence contribution to the overall dimension [12], which is information that cannot be obtained from analytical polymer theories.

The enormous flexibility of coarse-grained models can even be used to study the behavior of IDPs that interact with another folded protein. Since such complexes will also involve specific contacts between both partners, a structure of the complex, obtained experimentally or via computational methods, is a necessary starting point to identify specific contacts. Using this structural information, the stability of the folded binding partner is ensured either via strong harmonic constraints between native residue pairs or by treating the folded partner as a rigid body. In fact, the latter strategy is preferred since it saves computational resources and a number of molecular dynamics packages such as LAMMPs [126] and HOOMD-Blue [127] already provide this option. Specific inter-molecular contacts between IDPs and a folded partner can then be introduced using tunable harmonic or Lennard-Jones (LJ) potentials. Here, harmonic potentials strongly restrain the complex in its bound state such that the IDP will not dissociate from its folded partner and an LJ potential might be a more reasonable choice to sample both bound and unbound states. By tuning the interaction strengths of the potential, experimental smFRET values can be matched and a description of conformational flexibility of an IDP in complex with a folded protein can be obtained. This strategy has recently been used to quantify the conformational ensemble of disordered E-cadherin in complex with the folded protein β-catenin [12].

In summary, both all-atom explicit-solvent and coarse-grained models have their merits and limits. All-atom models provide a more accurate representation of the protein and also describe dyes explicitly, which simplifies the interpretation of smFRET experiments. However, all-atom models are limited by both sampling and force field accuracy. Even though force fields are being continuously improved with an increasing amount of experimental data, the quadratic scaling between timescales and system size will continue to limit sampling. Coarse-grained models are therefore a cost-effective solution. In fact, they can be considered as an extension of analytical polymer models. Yet, coarse-grained models average over many degrees of freedom and are less transferable among different systems. For instance, the strengths of interactions between amino acids often depend on the local sequence context. While charge–charge interactions are typically predicted well by coarse-grained models due to their substantial interaction energies, the combined effect of many weak interactions is much more difficult to catch. Hence, parameterizing (tuning) coarse-grained models with experimental data on the specific system at hand is very helpful for interpreting experimental data in terms of a molecular model. Clearly, specific interactions might be missed and the timescales of motions will be unrealistically fast such that all-atom simulation can be used as a complementation.

4 Probing and Modeling Sub-microsecond Dynamics of Disordered Proteins

4.1 Nanosecond Fluorescence Correlation Spectroscopy (nsFCS) Coupled with FRET

So far, we discussed the average dimension of disordered and unfolded proteins and its susceptibility toward changes in solution conditions. Our starting point for this discussion was the distribution of donor–acceptor distances that we expect for the extremely heterogeneous conformational ensemble of IDPs. In this section, we focus on the timescales at which such disordered ensembles are sampled, i.e., the reconfiguration time. The first attempts to identify this timescale were based on an analysis of the width in FRET histograms [14] and resulted in an upper limit of ≤25 μs for the unfolded state of Csp*Tm*. However, even extremely slow transitions at the timescale of seconds have been reported for surface-immobilized and unfolded RNase H [15]. Yet, already at that time, estimates from ensemble experiments such as contact-quenching, pioneered by Eaton and Hofrichter [128] and triplet–triplet energy transfer [129], pioneered by Kiefhaber and colleagues, suggested reconfiguration timescales <1 μs for flexible polypeptides. Indeed, the first rigorous measurement of donor–acceptor distance fluctuations in an unfolded protein identified dynamics in the order of 50 ns [34]. Technically, these smFRET measurements required a Hanbury Brown and Twiss detection scheme [130] to circumvent dead times of detectors and counting electronics[3] and the inter-photon times between photon pairs were measured with picosecond time-resolution (256–304 ps). Since then, the development of powerful counting electronics with a time-resolution down to 8 ps [38, 39] has greatly simplified smFRET-based nsFCS experiments [7, 23–25, 32, 33, 35–37, 131–135] such that they can even be used to study the dynamics of IDPs inside living cells [136, 137]. In their essentials, smFRET-based nsFCS experiments do not differ much from classical FCS experiments, which retrieve the timescale of fluorescence fluctuations, e.g., due to the diffusion of molecules in and out of the confocal volume, thus allowing the determination of translational diffusion coefficients and Stokes radii as a measure of the size of the diffusing particle. Differences are twofold: first, instead of one color, two colors are monitored in smFRET-based nsFCS experiments, and second, the timescale at which fluctuations are observed is much shorter (ns). Admittedly, the latter factor is of more of technical origin and simply requires fast counting electronics that allows the storage of photon arrival times with high, i.e., picosecond precision.

What is the principle of nsFCS? Due to the flexibility of IDPs, the distance between donor and acceptor dyes attached to them fluctuates thus giving rise to fluctuations in the rate of photons ($\tilde{n} = n/T$) from donor $\tilde{n}_D(t)$ and acceptor $\tilde{n}_A(t)$. The timescale of these fluctuations can be quantified with four correlation functions (g_{DD}, g_{AA}, g_{DA}, g_{AD}) defined by

$$g_{ij}(\tau) = \frac{\langle \tilde{n}_i(t)\tilde{n}_j(t+\tau)\rangle}{\langle \tilde{n}_i(t)\rangle\langle \tilde{n}_j(t)\rangle} \text{ with } i = \{D, A\} \text{ and } j = \{D, A\}. \tag{34}$$

For distance dynamics slower than the fluorescence lifetime of the dyes, the shape of these correlation functions differs characteristically between auto- (g_{DD}, g_{AA}) and cross-correlation functions (g_{DA}, g_{AD}) in the presence of FRET between donor and acceptor (Fig. 6a). Whereas the autocorrelation functions decay (positive amplitude), the cross-correlation functions should increase (negative amplitude) due to the anti-correlated change in photon rates from donor and acceptor. Since only distance changes coupled with FRET can cause such anti-correlated behavior, an increasing cross-correlation function unambiguously indicates the presence of distance dynamics. Conversely, an absence of this increase does not exclude distance fluctuations because other effects, e.g., static quenching of the dyes by aromatic amino acids such as tryptophan or tyrosine [12, 133], can mask the anti-correlated fluctuations of donor and acceptor emission. At very short timescales in the order of the fluorescence lifetimes of the dyes, all correlation functions increase due to photon antibunching. Organic dyes in smFRET experiments are single-photon emitters, i.e., after a photon has been emitted, the emission of a second photon takes time since the dye has to be excited again. Thinking of correlation functions as probability of observing two photons separated by the lag-time τ, this probability should ideally be

---

[3] Here, photons from each color are randomly distributed to two detectors.

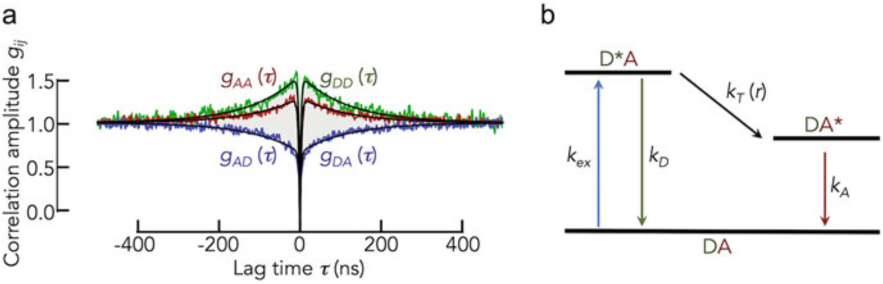

Fig. 6 Probing IDP dynamics with nsFCS. (**a**) Correlation functions from nsFCS for the intrinsically disordered DNA-binding domain of the protein c-Myc together with a global fit with multi-exponential decays as given by Eq. 35 (black lines). (**b**) Simple photo-physical scheme of a smFRET experiment (see text). The rate matrix for this scheme is given by Eq. 37

zero at $\tau = 0$, i.e., $g_{ij}(0) = 0$. Notably, this is rarely the case for cross-correlation functions because donor and acceptor can sporadically populate their excited states at the same time,[4] which can cause the simultaneous emission of a donor and an acceptor photon.

Initially, the practical goal of nsFCS experiments in the characterization of IDP dynamics was to determine the decay time of the correlation functions (Eq. 34). Importantly, the decay of the autocorrelation functions and the rise in the cross-correlation functions should provide the same relaxation time if these decays result from distance dynamics. Hence, the strategy of globally fitting the four correlation functions using exponentials such as

$$g_{ij}(\tau) = \left(1 - a_{ij}e^{-|\tau|/\tau_a^{ij}}\right)\left(1 \pm b_{ij}e^{-|\tau|/\tau_b}\right)\left(1 + c_{ij}e^{-|\tau|/\tau_c^{ij}}\right)/N_{ij} \qquad (35)$$

is frequently used [33, 34, 131]. Here, the first factor (RHS) accounts for antibunching with a correlation-function-specific amplitude (a_{ij}) and decay time (τ_a^{ij}), the second factor accounts for FRET-based distance dynamics (bunching) with a correlation-function-specific amplitude (b_{ij}) and a global decay time of distance fluctuations (τ_b), and the third factor accounts for the photo-physical triplet dynamics of the dyes that typically take place at timescales of several microseconds with a correlation-function-specific amplitude (c_{ij}) and decay time (τ_c^{ij}). The factor N_{ij}^{-1} is related to the average number of molecules in the confocal volume. The ± sign in the second factor is for autocorrelations (+) and cross-correlations (−). The global correlation time τ_b (global for all four correlation functions) then characterizes the timescale of distance fluctuation between donor and acceptor. Yet, caution has to be taken. The highly nonlinear distance dependence of FRET (Eq. 10) acts as a filter that flattens photon rate fluctuations much below and above the Förster distance R_0,

---

[4] For example, after FRET from donor to acceptor, the acceptor is in the excited state. If the donor is re-exited before the acceptor relaxes to the ground state, both dyes will be in the excited state.

which affects the correlation time. To retrieve the correct timescale of distance fluctuations, i.e., the reconfiguration time of the disordered chain, this filter has to be taken into account. Gopich et al. showed that the correlation time is given by [138]

$$\tau_b = D^{-1}\int_a^L P(r)^{-1}\left[\int_a^r \delta\tilde{n}(\rho)P(\rho)\mathrm{d}\rho\right]^2 \mathrm{d}r \Big/ \int_a^L \delta\tilde{n}(r)^2 P(r)\mathrm{d}r. \tag{36}$$

As in Eq. 12, $P(r)$ is again the donor–acceptor distance distribution and D is the intra-chain diffusion coefficient that characterizes the timescale of distance fluctuations. Knowing D and $P(r)$ then fully characterizes the dynamics of an IDP or an unfolded protein. The goal is therefore to compute D assuming a suitable model of the distance distribution. Since τ_b is identical for all correlation functions, it suffices to only consider the photon rates $\tilde{n}$ of one of the dyes, the donor in our case ($\tilde{n} = \tilde{n}_D$). The term $\delta\tilde{n}(r) = \tilde{n}(r) - \langle\tilde{n}\rangle$ can then be computed from the kinetic photo-physical scheme of excited states and ground states in a two-color FRET system (Fig. 6b). Neglecting the possibility that donor and acceptor can simultaneously populate their excited states and using the base (DA, D*A, DA*) where D and A indicate the ground state of donor and acceptor, respectively, and the asterisk indicates the excited states of the dyes, the populations of these photo-physical states expressed in the vector $\mathbf{p} = (p_{DA}\ \ p_{D^*A}\ \ p_{DA^*})$ are given by a linear and homogeneous differential equation system

$$\dot{\mathbf{p}} = \mathbf{K}\mathbf{p}$$

with the rate matrix

$$\mathbf{K} = \begin{pmatrix} -k_{ex} & k_D & k_A \\ k_{ex} & -[k_D + k_T(r)] & 0 \\ 0 & k_T(r) & -k_A \end{pmatrix}. \tag{37}$$

Here, $k_{ex} = \sigma I/h\nu$ is the excitation rate of the donor that depends on the laser intensity (I) and the absorption cross-section of the donor (σ) at the excitation wavelength λ ($\nu = c/\lambda$ where c is the speed of light), k_D and k_A are the decay rates of the excited states of donor and acceptor, respectively, and $k_T(r)$ is the distance-dependent rate of energy transfer from donor to acceptor (see Eq. 16). The photon rate of the donor as function of the distance is then given by $\tilde{n}(r) = \phi_D k_D p_{D^*A}^{ss}$ where the superscript *ss* indicates the steady-state population that is obtained from $\mathbf{0} = \mathbf{K}\mathbf{p}$ and ϕ_D is the quantum yield of the donor dye. As usual, the average donor photon rate is given by $\langle\tilde{n}\rangle = \int \tilde{n}(r)P(r)\mathrm{d}r$ such that the term $\delta\tilde{n}(r)$ is directly accessible. Typical values of these parameters for the dye pair Alexa Fluor 488 and Alexa Fluor 594 are $k_{ex} \approx 0.02\ \mathrm{ns}^{-1}$, $k_D \sim k_A \approx 0.25\ \mathrm{ns}^{-1}$, and $\phi_D > 0.9$. Thus, Eqs. 36 and 37 together with the measured correlation time and a model for the distance distribution $P(r)$ can now be used to compute the intra-chain diffusion coefficient D, i.e., the diffusion coefficient characterizing the motion of donor and acceptor dyes relative to each other.[5] Once D is known, also the reconfiguration time of the chain τ_r can be calculated using a similar approach, i.e., by replacing $\tilde{n}(r)$ with r in Eq. 36. Hence, reconfiguration times obtained using FRET-coupled nsFCS experiments typically depend on the model $P(r)$.

Importantly, Gopich and Szabo derived an elegant formalism with which auto- and cross-correlation functions can be directly computed for any kinetic model [139]. In contrast to the semi-empirical approach of determining correlation times using fits with exponential functions (Eq. 35) followed by a conversion to intra-chain diffusion coefficients and reconfiguration times using Eqs. 36 and 37, the method also effectively utilizes the amplitude information of the correlation functions. With the definitions in Eq. 37, the correlation function is given by

$$g_{ij}(\tau) = \frac{\mathbf{1}^T \mathbf{V}_j \mathrm{e}^{\mathbf{K}\tau} \mathbf{V}_i \mathbf{p}_{ss}}{\left(\mathbf{1}^T \mathbf{V}_j \mathbf{p}_{ss}\right)\left(\mathbf{1}^T \mathbf{V}_j \mathbf{p}_{ss}\right)}. \tag{38}$$

Here, $\mathbf{1}^T = (1\ 1\ldots)$ is the transposed unit vector, $\mathrm{e}^{\mathbf{K}\tau}$ is the matrix exponential of $\mathbf{K}\tau$, $\mathbf{p}_{ss}$ is the steady-state solution of Eq. 37, and $\mathbf{V}_j$ and $\mathbf{V}_i$ are detection matrices for channel i and j that indicate which transition is being monitored. For donor and acceptor detection, the matrices in our particular case are given by

$$\mathbf{V}_D = \xi_D k_D Q_D \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{and } \mathbf{V}_A = \xi_A k_A Q_A \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{39}$$

Here, Q_D and Q_A are the quantum yields of donor and acceptor, and ξ_D, ξ_A are the detection efficiencies. Also, leakage (cross-talk) from the donor to the acceptor channel can easily be implemented by correcting the acceptor detection matrix via $\mathbf{V}'_A = \mathbf{V}_A + \beta\mathbf{V}_D$. Using Eqs. 38 and 39 to directly fit the correlation functions of IDPs and unfolded proteins that sample a distance distribution $P(r)$ requires a combination of photophysics with the distance dynamics of the chain [34]. To this end, the dynamics of the chain are described as a one-dimensional diffusive process in the potential of mean force (PMF) given by the distribution $P(r)$. This approach essentially pictures the molecule as a dumbbell with donor and acceptor on opposite ends, linked by a spring whose potential is given by the PMF. The rate matrix then becomes a combination of two parts, intra-chain diffusion and photophysics

$$\mathbf{K}_{\mathrm{diff}} = D\partial/\partial r P(r)\partial/\partial r\ P(r)^{-1}\ \mathbf{I} + \mathbf{K}(r) \tag{40}$$

where $\mathbf{K}(r)$ describes the distance-dependent photophysics and the first term describes diffusion in a PMF. Here, $\mathbf{I}$ is a 3 × 3 unit matrix and D is the diffusion coefficient. To use Eq. 40 for computing correlation functions with Eq. 38, the

---

[5] Importantly, the intrinsic diffusion of donor and acceptor relative to each other should not be confused with the translational diffusion of the whole molecule.

diffusion operator (first term) has to be discretized, a procedure detailed in Nettels et al. [34]. For a practical perspective, fits of correlation functions with empirical exponentials (Eq. 35) or with an appropriate model (Eqs. 37–40) are very similar in quality. However, the number of free parameters is significantly reduced in the model-dependent fit since the excitation and emission rates of the dyes can be determined in advance, thus leaving only two free parameters, the diffusion D and the average donor–acceptor distance $\langle r^2 \rangle^{1/2} = R_{DA}$ as fitting parameters, the latter of which being additionally constrained by the average FRET efficiency $\langle E \rangle$ (Eq. 12).

4.2 Polymer Models for IDP Dynamics and Their Limitations

Which microscopic molecular processes determine the reconfiguration timescales of IDPs? Clearly, the complex arrangement of thousands of atoms in IDPs and an astronomical number of chain conformations poses challenges to analytical theories. Yet, rather simple analytically solvable bead-and-spring models, such as the Rouse model and derived models, turned out to be surprisingly successful [33, 36, 37, 140, 141]. The concept of these model goes back to P. E. Rouse who aimed at describing the viscoelastic properties of homopolymer solutions [140]. In the Rouse model [140], an IDP is composed of $n = N_K + 1$ beads that are connected by N_K segment bonds. As described in Sect. 3.1, these Kuhn segments consist of several bonds such that the chain obeys Gaussian statistics, resulting in a “phantom chain” in which beads are linked via harmonic springs [142]. The length of a segment b_K is chosen such that the mean square end-to-end distance of the chain can be written as $\langle r^2 \rangle = b_K^2 N_K$ (Eq. 14). The force-balance of a bead at position k will now be given by three terms: (1) an elastic spring force $\mathbf{f}_{ek}$, (2) a friction force due to the solvent $\mathbf{f}_{sk}$, and (3) a random force $\mathbf{f}_{bk}$ (Fig. 7a):

$$\mathbf{f}_{sk} + \mathbf{f}_{ek} = \mathbf{f}_{bk} \tag{41}$$

Assuming that solvent friction acts on individual beads, $\mathbf{f}_{sk}$ is given by Stokes’ law

$$\mathbf{f}_{sk} = \zeta_s \dot{\mathbf{r}}_k \text{ with } \zeta_s = 6\pi\eta a. \tag{42}$$

Here, ζ_s is the solvent friction coefficient for an individual bead, which is determined by its hydrodynamic radius a and by the solvent viscosity η, and $\mathbf{r}_k$ is the vector that defines the position of bead k. The second term in Eq. 41, i.e., the elastic spring force, can be derived from Hooke’s law. However, a bead in position k experiences an elastic force from two neighbors, namely at $k + 1$ and $k - 1$ such that

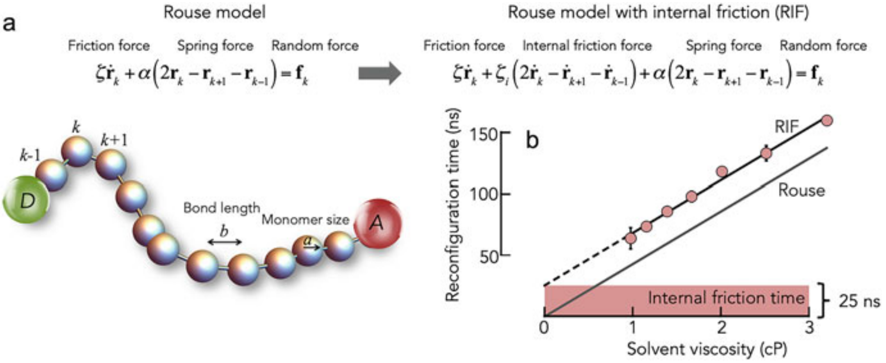

Fig. 7 Polymer dynamics models and their comparison with nsFCS. (**a**) A scheme of a coarse-grained homopolymer and a comparison of the equations of motion (Langevin equations) of the Rouse model without (*left*) and with (*right*) internal friction. (**b**) Reconfiguration times for the IDP (ACTR) obtained with nsFCS (circles) and a fit with the Rouse model with internal friction (RIF, black line). The internal friction component is indicated as red area. A Rouse model with the same monomer size (gray line) extrapolates to zero reconfiguration time at vanishing viscosity

$$\mathbf{f}_{ek} = \alpha(2\mathbf{r}_k - \mathbf{r}_{k+1} - \mathbf{r}_{k-1}) \text{ with } \alpha = 3k_BTN_K/\langle r^2\rangle \tag{43}$$

and the spring constant α describes the stiffness of the harmonic potentials between neighbored beads. With Eqs. 41–43, we can write the Langevin equation for the x-coordinate of a bead k as

$$\zeta_s\dot{x}(k, t) + \alpha[2x(k, t) - x(k+1, t) - x(k-1, t)] = X(k, t) \tag{44}$$

Here, $X(k,t)$ is the x-component of the random force that satisfies the usual fluctuation dissipation theorem of Gaussian-distributed white noise according to

$$\langle X(k, t)X(k, t')\rangle = 2k_BT\zeta_s\delta(t-t')\Delta(k-k').$$

Here, Δ indicates the Kronecker-delta. The solution has been described in the past [36, 141] and we will not re-iterate it here. It suffices to say that a spectrum of relaxation times is obtained, as expected for a coupled harmonic oscillator, and the mode-dependent relaxation time τ_R of the Rouse model is given by

$$\tau_R \approx \zeta_s/\alpha q^2 \propto \eta \tag{45}$$

with $q = \pi z/N_K$ with $z = \{1, 2, 3, \ldots\}$ being the mode number. Eq. 45 implies that the dynamics of the chain will slow down with increasing solvent viscosity and the relaxation of the chain is faster for higher modes, i.e., for larger z. Notably, Eq. 45 predicts $\tau_r \propto N_K^2$, which is different from the correct scaling $\tau_r \propto N_K^{3\nu}$ given by the Zimm model that also accounts for hydrodynamic interactions [63, 75].

Numerous nsFCS experiments on IDPs have shown that Eq. 45 does not correctly capture the dynamics of IDPs [33, 131] (Fig. 7b). Although the proportionality between reconfiguration time and solvent viscosity has indeed been found, a solvent-independent friction component of the type [143, 144]

$$\tau_r \approx \zeta_{\mathrm{eff}}/\alpha q^2 \text{ with } \zeta_{\mathrm{eff}} = \zeta_s + q^2 \zeta_i \tag{46}$$

had to be invoked to satisfactorily fit the experimental chain reconfiguration time τ_r (Fig. 7a). Here, ζ_i describes an additional friction process independent of the solvent viscosity that comes from processes internal to the protein. Eq. 46 can also be written as $\tau_r = \tau_R + \tau_i$ where τ_i is a viscosity-independent internal friction timescale (Fig. 7b). How can an additional friction process be justified? At high solvent friction such as in aqueous solutions, Kramers' reaction rate theory predicts that the reaction time τ is proportional to the friction coefficient of a particle [145]. Assuming continuum hydrodynamics, the reaction time should therefore scale with the solvent viscosity ($\tau \propto \eta$) without an offset, which has indeed been observed for slow reactions such as the barrier-dominated millisecond folding dynamics of two-state folders [146, 147]. However, substantial deviations were found for comparably fast reactions [148–150] where a linear extrapolation of the reaction time to zero viscosity also identified a limiting time scale τ_i [148]. An alternative to a viscosity dependence with an offset (as in Eq. 46) can be found if the scaling between reaction time and solvent viscosity has a weaker dependence such as $\tau \propto \eta^\beta$ with $\beta < 1$ [151, 152]. Indeed, such fractional viscosity dependencies have been identified for the diffusion of small molecular compounds [153–157] and physical interpretations range from a viscosity-dependent change of the hydrodynamic coupling between solvent and particle [158, 159], over a breakdown of the continuum hydrodynamics due to the granularity of the solvent at small length scales [153, 160], up to a breakdown of Kramers' theory due to memory effects caused by solvent relaxation [161, 162].

Yet, support for the presence of a limiting internal friction timescale (τ_i) came from recent nsFCS experiments of unfolded and intrinsically disordered proteins (IDPs) [33, 131, 163]. Though these experiments suggested that internal friction results from a rather local process that acts on individual Kuhn segments (typically around 4 peptide bonds) [33], its molecular origin remained elusive and several possibilities such as bond rotations [143, 164], mode-coupling [165], or steric interference of chain segments [166] are in discussion [36]. Identifying the molecular events that cause internal friction in unstructured polypeptides might be a crucial step toward understanding the elementary processes in the folding of proteins and IDPs. Recent simulations and experiments point toward a combination of effects, including dihedral angle flipping, intra-chain interactions, and solvent memory effects [37, 167–169]. Importantly, some of these effects have already been incorporated in theories by Bazùa and Williams [143], Allegra and Ganazzoli [164], and de Gennes [166]. Other authors even addressed more essential simplifications in Rouse-type models such as the assumption of harmonic springs in bead-and-spring models [165]. However, a direct comparison of this variety of theories with experimental results demonstrated that all these theories are capable of describing experimental results on the disordered protein ACTR [36].

It might not be surprising that analytical theories capture internal friction effects with an additional fitting parameter yet, an identification of the molecular processes that give rise to this parameter remains elusive based on experiment and theory alone. All-atom simulations with explicit solvent therefore became an ideal tool to disentangle the molecular contributions that define reconfiguration times, intra-chain diffusion coefficients, and internal friction effects. However, precise estimates from all-atom simulations require an optimized water model. In fact, viscosities of different water models vary tremendously [170], which strongly alters intra-chain dynamics. TIP4P/2005 has been found to correctly capture the viscosity of water [170, 171] and therefore force fields using this water model have been shown to provide a faithful estimate of the intra-chain dynamics measured with smFRET-nsFCS experiment in the order of hundred nanoseconds [22, 97]. Most importantly, simulations allow a direct analysis of the molecular origins of internal friction effects in IDPs. Also in simulations, the reconfiguration time of IDPs deviates from the prediction of the simple Rouse model (Eqs. 40–43). By modifying the mass of water to mimic changes in solvent viscosity, internal friction behavior was reproduced in all-atom explicit-solvent simulations down to viscosities that are typically inaccessible experimentally ($\eta \gg 1cP$) [168]. These simulations showed that internal friction effects were primarily associated with backbone dihedral transitions [167, 169] as suspected by Bazùa and Williams four decades ago [143], but also intra-chain interactions contributed to the nonlinear viscosity dependence of IDPs, thus showing that they are sequence-dependent [172].

Acknowledgments H.H. thanks the Israel Science Foundation (ISF 2253/18) and the European Research Council (ERC-CoG 864578). W.Z. acknowledges the support from the National Science Foundation (MCB-2015030) and the National Institutes of Health (R35GM146814).

References

1. van der Lee R et al (2014) Classification of intrinsically disordered regions and proteins. Chem Rev 114:6589–6631
2. Uversky VN, Dunker AK (2008) Biochemistry. Controlled chaos. Science 322(5906): 1340–1341
3. Wright PE, Dyson HJ (2014) Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol 16(1):18–29
4. Ferreon ACM, Ferreon JC, Wright PE, Deniz AA (2013) Modulation of allostery by protein intrinsic disorder. Nature 498(7454):390–394
5. Gsponer J, Futschik ME, Teichmann SA (2008) Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. Science 322:1365–1368
6. Schuler B et al (2020) Binding without folding - the biomolecular function of disordered polyelectrolyte complexes. Curr Opin Struct Biol 60:66–76

7. Borgia A et al (2018) Extreme disorder in an ultrahigh-affinity protein complex. Nature 555(7694):61–66
8. Holmstrom ED, Liu Z, Nettels D, Best RB, Schuler B (2019) Disordered RNA chaperones can enhance nucleic acid folding via local charge screening. Nat Comms 10(1):2453–2411
9. Mittag T et al (2008) Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. Proc Natl Acad Sci U S A 105(46):17772–17777
10. Hendus-Altenburger R et al (2016) The human Na + /H + exchanger 1 is a membrane scaffold protein for extracellular signal-regulated kinase 2. BMC Biol 14(1):1–17
11. Milles S et al (2015) Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors. Cell 163(3):734–745
12. Wiggers F et al (2021) Diffusion of a disordered protein on its folded ligand. Proc Natl Acad Sci U S A 118(37):e2106690118
13. Deniz AA et al (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(10):5179–5184
14. Schuler B, Lipman E, Eaton W (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. Nature 419(6908):743–747
15. Kuzmenkina E, Heyes C, Nienhaus G (2005) Single-molecule Forster resonance energy transfer study of protein dynamics under denaturing conditions. Proc Natl Acad Sci U S A 102(43):15471–15476
16. Sherman E, Haran G (2006) Coil-globule transition in the denatured state of a small protein. Proc Natl Acad Sci U S A 103(31):11539–11543
17. Merchant K, Best R, Louis J, Gopich I, Eaton W (2007) Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. Proc Natl Acad Sci U S A 104(5):1528–1533
18. Hofmann H, Golbik R, Ott M, Hübner C, Ulbrich-Hofmann R (2008) Coulomb forces control the density of the collapsed unfolded state of barstar. J Mol Biol 376(2):597–605
19. Müller-Späth S et al (2010) Charge interactions can dominate the dimensions of intrinsically disordered proteins. Proc Natl Acad Sci U S A 107:14609–14614
20. Hofmann H et al (2012) Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. Proc Natl Acad Sci U S A 109(40): 16155–16160
21. Wuttke R et al (2014) Temperature-dependent solvation modulates the dimensions of disordered proteins. Proc Natl Acad Sci U S A 111(14):5213–5218
22. Borgia A et al (2016) Consistent view of polypeptide chain expansion in chemical denaturants from multiple experimental methods. J Am Chem Soc 138:11714
23. 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
24. Grossman-Haham I, Rosenblum G, Namani T, Hofmann H (2018) Slow domain reconfiguration causes power-law kinetics in a two-state enzyme. Proc Natl Acad Sci U S A 115(3):513–518
25. Vancraenenbroeck R, Hofmann H (2018) Occupancies in the DNA-binding pathways of intrinsically disordered helix-loop-helix leucine-zipper proteins. J Phys Chem B 122:11460–11467
26. Vancraenenbroeck R, Harel YS, Zheng W, Hofmann H (2019) Polymer effects modulate binding affinities in disordered proteins. Proc Natl Acad Sci U S A 116(39):19506–19512
27. Fuertes G et al (2017) Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements. Proc Natl Acad Sci U S A 114(31):E6342–E6351
28. Mukhopadhyay S, Krishnan R, Lemke E, Lindquist S, Deniz A (2007) A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. Proc Natl Acad Sci U S A 104(8):2649–2654

29. Aznauryan M et al (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(37):E5389–E5398
30. Hoffmann A et al (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. Proc Natl Acad Sci U S A 104(1):105–110
31. Soranno A et al (2014) Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. Proc Natl Acad Sci U S A 111(13):4874–4879
32. Nettels D et al (2009) Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. Proc Natl Acad Sci U S A 106:20740–20745
33. Soranno A et al (2012) Quantifying internal friction in unfolded and intrinsically disordered proteins with single molecule spectroscopy. Proc Natl Acad Sci U S A 109:17800–17806
34. Nettels D, Gopich I, Hoffmann A, Schuler B (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. Proc Natl Acad Sci U S A 104(8):2655–2660
35. Zosel F, Haenni D, Soranno A, Nettels D, Schuler B (2017) Combining short- and long-range fluorescence reporters with simulations to explore the intramolecular dynamics of an intrinsically disordered protein. J Chem Phys 147:152708
36. Soranno A, Zosel F, Hofmann H (2018) Internal friction in an intrinsically disordered protein-comparing Rouse-like models with experiments. J Chem Phys 148(12):123326
37. Soranno A et al (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
38. Wahl M et al (2008) Scalable time-correlated photon counting system with multiple independent input channels. Rev Sci Instrum 79(12):123113
39. Wahl M, Rahn H-J, Gregor I, Erdmann R, Enderlein J (2007) Dead-time optimized time-correlated photon counting instrument with synchronized, independent timing channels. Rev Sci Instrum 78(3):033106
40. Dertinger T et al (2007) Two-focus fluorescence correlation spectroscopy: a new tool for accurate and absolute diffusion measurements. ChemPhysChem 8(3):433–443
41. Eggeling C et al (2001) Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. J Biotechnol 86(3):163–180
42. Schuler B (2007) Application of single molecule Förster resonance energy transfer to protein folding. Methods Mol Biol 350:115–138
43. Rosenblum G et al (2021) Allostery through DNA drives phenotype switching. Nat Commun 12(1):2967–2912
44. Hellenkamp B et al (2018) Precision and accuracy of single-molecule FRET measurements-a multi-laboratory benchmark study. Nat Methods 15(9):669–676
45. Kapanidis AN et al (2004) Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. Proc Natl Acad Sci U S A 101(24):8936–8941
46. Müller BK, Zaychikov E, Bräuchle C, Lamb DC (2005) Pulsed interleaved excitation. Biophys J 89(5):3508–3522
47. Kapanidis AN et al (2005) Alternating-laser excitation of single molecules. Acc Chem Res 38(7):523–533
48. Laurence T, Kong X, Jäger M, Weiss S (2005) Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. Proc Natl Acad Sci U S A 102(48): 17348–17353
49. Gopich I, Szabo A (2005) Theory of photon statistics in single-molecule Förster resonance energy transfer. J Chem Phys 122(1):14707
50. Lipman E, Schuler B, Bakajin O, Eaton W (2003) Single-molecule measurement of protein folding kinetics. Science 301(5637):1233–1235
51. Zheng W et al (2018) Inferring properties of disordered chains from FRET transfer efficiencies. J Chem Phys 148(12):123329

52. Fisher ME (1966) Shape of a self-avoiding walk or polymer chain. J Chem Phys 44(616)
53. Des Cloizeaux J (1974) Lagrangian theory for a self-avoiding random chain. Phys Rev A 10(5):1665–1669
54. O'Brien EP, Morrison G, Brooks BR, Thirumalai D (2009) How accurate are polymer models in the analysis of Förster resonance energy transfer experiments on proteins? J Chem Phys 130(12):124903
55. Gopich IV, Szabo A (2012) Theory of the energy transfer efficiency and fluorescence lifetime distribution in single-molecule FRET. Proc Natl Acad Sci U S A 109(20):7747–7752
56. Kalinin S, Valeri A, Antonik M, Felekyan S, Seidel CAM (2010) Detection of structural dynamics by FRET: a photon distribution and fluorescence lifetime analysis of systems with multiple states. J Phys Chem B 114:7983–7995
57. Dale RE, Eisinger J, Blumberg B (1979) Orientational freedom of molecular probes. Biophys J 26:161–193
58. Schuler B, Lipman EA, Steinbach PJ, Kumke M, Eaton WA (2005) Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. Proc Natl Acad Sci U S A 102(8):2754–2759
59. Hillger F et al (2008) Probing protein-chaperone interactions with single-molecule fluorescence spectroscopy. Angew Chem Int Ed Engl 47(33):6184–6188
60. Hofmann H et al (2010) Single-molecule spectroscopy of protein folding in a chaperonin cage. Proc Natl Acad Sci U S A 107(26):11793–11798
61. Sun S, Nishio I, Swislow G, Tanaka T (1980) The coil-globule transition - radius of gyration of polystyrene in cyclohexane. J Chem Phys 73(12):5971–5975
62. de Gennes P (1975) Collapse of a polymer chain in poor solvents. J Phys Lett 3:L55–L57
63. Doi M, Edwards S (1988) The theory of polymer dynamics. Oxford University Press, New York
64. Dua A, Vilgis TA (2005) Self-consistent variational theory for globules. EPL 71(1):49
65. Sanchez I (1979) Phase transition behavior of the isolated polymer chain. Macromolecules 12: 980–988
66. Hofmann H (2016) Understanding disordered and unfolded proteins using single-molecule FRET and polymer theory. Methods Appl Fluoresc 4(4):042003
67. England J, Haran G (2011) Role of solvation effects in protein denaturation: from thermodynamics to single molecules and back. Annu Rev Phys Chem 62:257–277
68. Nozaki Y, Tanford C (1970) The solubility of amino acids, diglycine, and triglycine in aqueous guanidine hydrochloride solutions. J Biol Chem 245(7):1648–1652
69. Thirumalai D, Liu Z, O'Brien EP, Reddy G (2012) Protein folding: from theory to practice. Curr Opin Struct Biol 23:1–8
70. O'Brien E, Ziv G, Haran G, Brooks B, Thirumalai D (2008) Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. Proc Natl Acad Sci U S A 105:13403–13408
71. Ziv G, Haran G (2009) Protein folding, protein collapse, and Tanford's transfer model: lessons from single-molecule FRET. J Am Chem Soc 131(8):2942–2947
72. Ziv G, Thirumalai D, Haran G (2009) Collapse transition in proteins. Phys Chem Chem Phys 11(1):83–93
73. Flory P (1949) The configuration of real polymer chains. J Chem Phys 17(3):303
74. Le Guillou JC, Zinn-Justin J (1977) Critical exponents for the n-vector model in three dimensions from field theory. Phys Rev Lett 39:95–98
75. Rubinstein M, Colby RH (2012) Polymer physics. Oxford University Press, Oxford, p 440
76. Kohn J et al (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. Proc Natl Acad Sci U S A 101(34):12491–12496
77. Riback JA et al (2017) Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. Science 358(6360):238–241
78. Wilkins D et al (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. Biochemistry 38(50):16424–16431

79. Grosberg A, Kuznetsov D (1992) Quantitative theory of the globule-to-coil transition 1. Link density distribution in a globule and its radius of gyration. Macromolecules 25(7):1970–1979
80. Crick SL, Jayaraman M, Frieden C, Wetzel R, Pappu RV (2006) Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. Proc Natl Acad Sci U S A 103(45):16764–16769
81. Grosberg A, Kuznetsov D (1992) Quantitative theory of the globule-to-coil transition. 2. Density-density correlation in a globule and the hydrodynamic radius of a macromolecule. Macromolecules 25:1980–1990
82. Uzawa T et al (2006) Time-resolved small-angle X-ray scattering investigation of the folding dynamics of heme oxygenase: implication of the scaling relationship for the submillisecond intermediates of protein folding. J Mol Biol 357(3):997–1008
83. Camacho C, Thirumalai D (1993) Kinetics and thermodynamics of folding in model proteins. Proc Natl Acad Sci U S A 90:6369–6372
84. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proc Natl Acad Sci U S A 107(18):8183–8188
85. Sherry KP, Das RK, Pappu RV, Barrick D (2017) Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor. Proc Natl Acad Sci U S A 114(44):E9243–E9252
86. Peran I et al (2019) Unfolded states under folding conditions accommodate sequence-specific conformational preferences with random coil-like dimensions. Proc Natl Acad Sci U S A 116(25):12301–12310
87. Martin EW et al (2020) Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. Science 367(6478):694–699
88. Holm C et al (2004) Polyelectrolyte theory. Springer, Berlin
89. Dobrynin AV, Rubinstein M (2005) Theory of polyelectrolytes in solutions and at surfaces. Prog Polym Sci 30(11):1049–1118
90. Higgs PG, Joanny J-F (1991) Theory of polyampholyte solutions. J Chem Phys 94(2): 1543–1554
91. Ha DT (1992) Conformations of a polyelectrolyte chain. Phys Rev A 46(6):R3012–R3015
92. Kundagrami A, Muthukumar M (2010) Effective charge and Coll-Globule transition of a polyelectrolyte chain. Macromolecules 43(5):2574–2581
93. Bhattacharjee A, Kundu P, Dua A (2011) Self-consisten theory of structures and transitions in weak polyampholytes. Macromol Theory Simul 20:75–84
94. Shakhnovich EI, Gutin AM (1989) Formation of unique structure in polypeptide chains. Theoretical investigation with the aid of a replica approach. Biophys Chem 34(3):187–199
95. Gomes G-NW, Namini A, Gradinaru CC (2022) Integrative conformational ensembles of Sic1 using different initial pools and optimization methods. Front Mol Biosci 9:910956
96. Lindorff-Larsen K, Best RB, Depristo MA, Dobson CM, Vendruscolo M (2005) Simultaneous determination of protein structure and dynamics. Nature 433:128–132
97. Best RB, Zheng W, Mittal J (2014) Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. J Chem Theory Comput 10(11): 5113–5124
98. Vitalis A, Pappu RV (2008) ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions. J Comput Chem 30:673–699
99. Dignon GL, Zheng WW, Kim YC, Best RB, Mittal J (2018) Sequence determinants of protein phase behavior from a coarse-grained model. PLoS Comput Biol 14(1):e1005941
100. Robustelli P, Piana S, Shaw DE (2018) Developing a molecular dynamics force field for both folded and disordered protein states. Proc Natl Acad Sci U S A 115(21):E4758–E4766
101. Huang J et al (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods 14(1):71–73

102. Hess B, Kutzner C, Van der Spoel D, Lindahl E (2008) GROMACS4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theor Comput 4(3): 435–447
103. Case DA et al (2005) The Amber biomolecular simulation programs. J Comput Chem 26(16): 1668–1688
104. Eastman P et al (2017) OpenMM 7: rapid development of high performance algorithms for molecular dynamics. PLoS Comput Biol 13(7):e1005659
105. Phillips JC et al (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781–1802
106. Shaw DE et al (2007) Anton, a special-purpose machine for molecular dynamics simulation. In: ISCA'07: 34th annual international symposium on computer architecture, conference proceedings, conference proceedings – annual international symposium on computer architecture. Association for Computing Machinery, New York, pp 1–12
107. Robustelli P, Piana S, Shaw DE (2020) Mechanism of coupled folding-upon-binding of an intrinsically disordered protein. J Am Chem Soc 142(25):11092–11101
108. Best RB, Hofmann H, Nettels D, Schuler B (2015) Quantitative interpretation of FRET experiments via molecular simulation: force field and validation. Biophys J 108:2721–2731
109. Das RK, Pappu RV (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc Natl Acad Sci U S A 110(33):13392–13397
110. Zheng W et al (2016) Probing the action of chemical denaturant on an intrinsically disordered protein by simulation and experiment. J Am Chem Soc 138(36):11702–11713
111. Yoo TY et al (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
112. Zheng W, Best R (2018) An extended Guinier analysis for intrinsically disordered proteins. J Mol Biol 430:2540–2553
113. Witten TA, Schäfer L (1978) Two critical ratios in polymer solutions. J Phys A 11(9): 1843–1854
114. Gomes GW et al (2020) Conformational ensembles of an intrinsically disordered protein consistent with NMR, SAXS, and single-molecule FRET. J Am Chem Soc 142(37): 15697–15710
115. Brangwynne CP et al (2009) Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science 324(5935):1729–1732
116. Dyson HJ, Wright PE (2002) Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol 12(1):54–60
117. Theillet F-X et al (2014) Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs). Chem Rev 114(13):6661–6714
118. Debye P, Hückel E (1923) Zur Theorie der Elektrolyte: I. Gefrierpunktserniedrigung und verwandte Erscheinungen. Phys Z 24(9):185–206
119. Hooper HH, Blanch HW, Prausnitz JM (1990) Configurational properties of partially ionized polyelectrolytes from Monte Carlo simulation. Macromolecules 23(22):4820–4829
120. English AE, Tanaka T, Edelman ER (1998) Polyampholytic hydrogel swelling transitions: limitations of the Debye−Hückel law. Macromolecules 31(6):1989–1995
121. Ashbaugh HS, Hatch HW (2008) Natively unfolded protein stability as a coil-to-globule transition in charge/hydropathy space. J Am Chem Soc 130(29):9536–9542
122. Latham AP, Zhang B (2019) Improving coarse-grained protein force fields with small-angle X-ray scattering data. J Phys Chem B 123(5):1026–1034
123. Dannenhoffer-Lafage T, Best RB (2021) A data-driven hydrophobicity scale for predicting liquid-liquid phase separation of proteins. J Phys Chem B 125(16):4046–4056
124. Regy RM, Thompson J, Kim YC, Mittal J (2021) Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins. Protein Sci 30(7):1371–1379

125. Tesei G, Schulze TK, Crehuet R, Lindorff-Larsen K (2021) Accurate model of liquid-liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. Proc Natl Acad Sci U S A 118(44)
126. Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. J Comput Phys 117(1):1–19
127. Anderson JA, Lorenz CD, Travesset A (2008) General purpose molecular dynamics simulations fully implemented on graphics processing units. J Comput Phys 227(10):5342–5359
128. Lapidus LJ, Eaton WA, Hofrichter J (2000) Measuring the rate of intramolecular contact formation in polypeptides. Proc Natl Acad Sci U S A 97(13):7220–7225
129. Bieri O et al (1999) The speed limit for protein folding measured by triplet-triplet energy transfer. Proc Natl Acad Sci U S A 96(17):9597–9601
130. Hanbury Brown R, Twiss RQ (1956) Correlation between photons in two coherent beams of light. Nature 177(4497):27–29
131. Borgia A et al (2012) Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. Nat Commun 3:1195
132. Schuler B, Müller-Späth S, Soranno A, Nettels D (2012) Application of confocal single-molecule FRET to intrinsically disordered proteins. Methods Mol Biol 896:21–45
133. Haenni D, Zosel F, Reymond L, Nettels D, Schuler B (2013) Intramolecular distances and dynamics from the combined photon statistics of single-molecule FRET and Photoinduced electron transfer. J Phys Chem B 117(42):13015–13028
134. Schuler B, Hofmann H (2013) Single-molecule spectroscopy of protein folding dynamics-expanding scope and timescales. Curr Opin Struct Biol 23:1–12
135. Cubuk J et al (2021) The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. Nat Commun 12(1):1936–1917
136. König I et al (2015) Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. Nat Methods 12(8):773–779
137. Schuler B, König I, Soranno A, Nettels D (2021) Impact of in-cell and in-vitro crowding on the conformations and dynamics of an intrinsically disordered protein. Angew Chem Int Ed Engl 133:2
138. Gopich IV, Nettels D, Schuler B, Szabo A (2009) Protein dynamics from single-molecule fluorescence intensity correlation functions. J Chem Phys 131(9):095102
139. Gopich IV, Szabo A (2008) Theory of photon counting in single molecule spectroscopy. World Scientific Publishing Co. Pte. Ltd., Singapore
140. Rouse PE (1953) A theory of the linear viscoelastic properties of dilute solutions of coiling polymers. J Chem Phys 21:1272–1280
141. Makarov DE (2010) Spatiotemporal correlations in denatured proteins: the dependence of fluorescence resonance energy transfer (FRET)-derived protein reconfiguration times on the location of the FRET probes. J Chem Phys 132(3):035104
142. de Gennes P-G (1979) Scaling concepts in polymer physics. Cornell University Press, Ithaca
143. Bazúa ER, Williams MC (1973) Molecular formulation of the internal viscosity in polymer dynamics, and stress symmetry. J Chem Phys 59:2858–2868
144. Khatri BS, McLeish TCB (2007) Rouse model with internal friction: a coarse grained framework for single biopolymer dynamics. Macromolecules 40(18):6770–6777
145. Kramers H (1940) Brownian motion in a field of force and the diffusion model of chemical reactions. Physica 7:284–304
146. Plaxco KW, Baker D (1998) Limited internal friction in the rate-limiting step of a two-state protein folding reaction. Proc Natl Acad Sci U S A 95(23):13591–13596
147. Jacob M, Geeves M, Holtermann G, Schmid FX (1999) Diffusional barrier crossing in a two-state protein folding reaction. Nat Struct Biol 6(10):923–926
148. Ansari A, Jones CM, Henry ER, Hofrichter J, Eaton WA (1992) The role of solvent viscosity in the dynamics of protein conformational changes. Science 256(5065):1796–1798

149. Pabit SA, Roder H, Hagen SJ (2004) Internal friction controls the speed of protein folding from a compact configuration. Biochemistry 43(39):12532–12538
150. Hagen SJ, Qiu L, Pabit SA (2005) Diffusional limits to the speed of protein folding: fact or friction? J Phys Condens Matter 17(18):S1503–S1514
151. Chung HS, Eaton WA (2013) Single-molecule fluorescence probes dynamics of barrier crossing. Nature 502(7473):685–688
152. Frauenfelder H, Fenimore PW, Chen G, McMahon BH (2006) Protein folding is slaved to solvent motions. Proc Natl Acad Sci U S A 103(42):15469–15472
153. Evans DF, Tominaga T, Davis HT (1981) Tracer diffusion in polyatomic liquids. J Chem Phys 74(2):1298
154. Evans DF, Tominaga T, Chan C (1979) Diffusion of symmetrical and spherical solutes in protic, aprotic, and hydrocarbon solvents. J Solution Chem 8:461–478
155. Pollack GL, Enyeart JJ (1985) Atomic test of the stokes-Einstein law. II. Diffusion of Xe through liquid hydrocarbons. Phys Rev A 31:980–984
156. Hiss TG, Cussler EL (1973) Diffusion in high viscosity liquids. AIChE 19:698–703
157. Ellerton HD, Mulcahy DE, Dunlop PJ, Reinfelds G (1964) Mutual frictional coefficients of several amino acids in aqueous solution at 25. J Phys Chem 68(2):403–408
158. Harris KR (2009) The fractional Stokes-Einstein equation: application to Lennard-Jones, molecular, and ionic liquids. J Chem Phys 131(5):054503
159. Zwanzig R, Harrison AK (1985) Modifications of the Stokes-Einstein formula. J Chem Phys 83(11):5861–5862
160. Bhattacharyya S, Bagchi B (1997) Anomalous diffusion of small particles in dense liquids. J Chem Phys 106(5):1757–1763
161. Grote RF, Hynes JT (1981) Reactive modes in condensed phase reactions. J Chem Phys 74: 4465–4475
162. Grote RF, Vanderzwan G, Hynes JT (1984) Frequency-dependent friction and solution reaction-rates. J Phys Chem 88(20):4676–4684
163. Neuweiler H, Johnson C, Fersht A (2009) Direct observation of ultrafast folding and denatured state dynamics in single protein molecules. Proc Natl Acad Sci U S A 106(44):18569–18574
164. Allegra G, Ganazzoli F (1981) Configurations and dynamics of real chains 2. Internal viscosity. Macromolecules 14(4):1110–1119
165. Adelman SA, Freed KF (1977) Microscopic theory of polymer internal viscosity: mode coupling approximation for the Rouse model. J Chem Phys 67(4):1380
166. de Gennes PG (1977) Origin of internal viscosities in dilute polymer solutions. J Chem Phys 66(12):5825–5826
167. Echeverria I, Makarov DE, Papoian GA (2014) Concerted dihedral rotations give rise to internal friction in unfolded proteins. J Am Chem Soc 136:8708–8713
168. Schulz JCF, Schmidt L, Best RB, Dzubiella J, Netz RR (2012) Peptide chain dynamics in light and heavy water: zooming in on internal friction. J Am Chem Soc 134(14):6273–6279
169. de Sancho D, Sirur A, Best RB (2014) Molecular origins of internal friction effects on protein-folding rates. Nat Commun 5:4307
170. Gonzalez MA, Abascal JL (2010) The shear viscosity of rigid water models. J Chem Phys 132(9):096101
171. Abascal JLF, Vega C (2005) A general purpose model for the condensed phases of water: TIP4P/2005. J Chem Phys 123:234505
172. Zheng W, Hofmann H, Schuler B, Best RB (2018) Origin of internal friction in disordered proteins depends on solvent quality. J Phys Chem B 122(49):11478–11487