Dynamical buffering of reconfiguration dynamics in intrinsically disordered proteins

Miloš T. Ivanović,[†] Andrea Holla,[†] Mark F. Nüesch,[†] Valentin von Roten,[†] Benjamin Schuler,[*,†,‡] and Robert B. Best[*,¶]

†*Department of Biochemistry, University of Zurich, Zurich, Switzerland*

‡*Department of Physics, University of Zurich, Zurich, Switzerland*

¶*Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health*

E-mail: schuler@bioc.uzh.ch; robert.best2@nih.gov

Abstract

Although they populate diverse conformational ensembles, intrinsically disordered proteins have well-defined functions determined by their equilibrium and dynamic properties. Such properties can be probed by single-molecule FRET experiments, which have revealed a strong dependence of the equilibrium degree of compaction on sequence for a set of 16 naturally occurring disordered regions of the same length from RNA-binding proteins. Remarkably, however, we find chain reconfiguration times measured for the same proteins to be almost independent of sequence, although chain dynamics would be expected to be slower for more compact proteins because of stronger intrachain interactions leading to increased internal friction. We investigate this effect with the aid of multi-microsecond, all-atom explicit-solvent simulations of all 16 disordered proteins. The simulations reproduce the experimental FRET efficiencies across the sequence set with near-quantitative accuracy. Explicitly including the FRET dyes

improves agreement while minimally perturbing the protein ensemble, with residual deviations arising mainly from slightly overstabilized salt bridges. The simulations further reproduce the experimentally observed lack of correlation between reconfiguration times and chain dimensions across sequences. We are able to rationalize this observation as arising from two compensating factors as the chains get more compact: The narrowing of end-to-end equilibrium distance distributions and a concomitant reduction of the effective end-to-end diffusion coefficients have opposite effects that result in the the small variation of reconfiguration times with protein collapse, helping to "buffer" the effect of sequence on linker dynamics.

Introduction

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) are now recognized to play a broad range of functional roles in biology, including molecular signaling, transcription regulation, and formation of biomolecular condensates.[1,2] They are also implicated in aberrant behaviour, as most amyloids are formed by disordered proteins,[3,4] and "aging" of the disordered regions within condensates has been associated with disease outcomes such as ALS.[5,6] Disordered linker sequences have also been shown to play an important role in maintaining an appropriate distance between the folded domains which they bridge.[7]

Relating the structure and dynamics of disordered proteins to their functions and role in disease is challenging because of the very broad structural ensembles they populate and their rapid dynamics. This means that apart from NMR,[2] conventional structural biology methods used for folded proteins are generally not applicable. Alternative experiments that enable the distance distributions of IDPs to be probed include small angle X-ray scattering (SAXS)[8] and Förster resonance energy transfer (FRET).[9,10] However, these methods give limited information about the rapidly interconverting ensemble of disordered structures, often making their interpretation challenging. In this context, molecular simulations with

either all-atom or coarse-grained force fields can play a critical role in filling in the structural and dynamic details. With only limited constraints from experiment, however, the results of such simulations are strongly dependent on the accuracy of the underlying simulation force field.[11]

While a decade ago, molecular simulations of IDPs resulted in ensembles that were both too collapsed and contained excessive amounts of either α-helical or β-structure, developments in the intervening years have dramatically improved their quality, and independent simulations by several groups have generally confirmed the accuracy of these new developments.[12–14] However, it is also known that the properties of IDPs are strongly sequence-dependent,[10,15–18] a feature sometimes referred to as their "molecular grammar".[19,20] Although it has been shown that force fields can reproduce the degree of compaction of selected proteins of different lengths,[12,13] a more stringent test lies in a side-by-side comparison of proteins of the same length.

Recently, single-molecule FRET measurements have been performed on a set of intrinsically disordered linker regions from RNA-binding proteins, carefully selected to have very diverse sequence properties[22] (see Table S1 for all sequences). The measured FRET efficiencies for these linker IDRs range from roughly 0.4 to 0.9 for the dye pair Cy3B/CF660R (Fig. 1A,B), indicating that these sequences indeed encode a broad range of chain expansion. Their polymer scaling exponents ν inferred from the SAW-ν model[21] range from values slightly above 3/5, close to the expectation for a self-avoiding walk (no attractive interactions), to slightly below 1/2, the value expected when attractive and repulsive interactions are balanced. However, none of the chains approach the compaction of collapsed globules with $\nu \approx 1/3$. Examples of distance distributions and configurations corresponding to the most expanded and most collapsed states are shown in Fig. 1B, illustrating the substantial range of compaction sampled.

To probe the corresponding dynamics of these 16 IDRs, we have now quantified their chain reconfiguration times, τ_r, by nanosecond fluorescence correlation spectroscopy (ns-

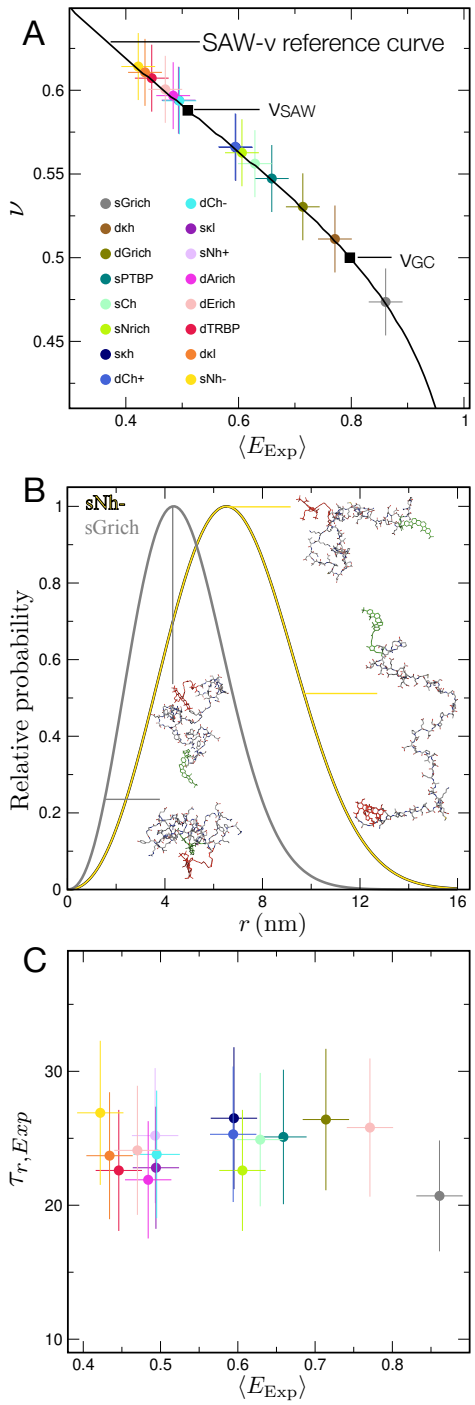

Figure 1: Chain dimensions in a set of linker IDRs with diverse sequence properties. (A) Scaling exponents, ν, inferred from experimental FRET efficiencies,$\langle E_{Exp}\rangle$, using the SAW-ν model[21] show linkers ranging from close to self-avoiding walk (SAW) chains to Gaussian chain-like. The uncertainty in the FRET efficiencies is taken as 0.03, and the uncertainty in the scaling exponents as 0.02.[22] The naming of the different sequences (see legend) reflects the sequence properties (see Table S1 and Holla et al.[22]). (B) Examples of SAW-ν end-to-end distance distributions $P(r)$ inferred from the experimental FRET efficiencies, contrasting the most expanded (sNh-) and most collapsed (sGrich) chains, with illustrative configurations from the simulations. Hydrogen atoms are omitted for clarity. (C) Lack of correlation of experimentally measured chain reconfiguration times, $\tau_{r,Exp}$, with FRET efficiencies. The uncertainty of the chain reconfiguration times is estimated to be about 20%.

FCS). This method allows us to measure the end-to-end distance dynamics by means of the fluorescence fluctuations of the donor and acceptor fluorophores[10,23] (see Methods). We observe very rapid chain dynamics, with values of τ_r between roughly 20 and 30 ns, in the range expected for IDPs of this chain length.[10] However, despite the broad range of chain dimensions of these IDRs, the reconfiguration times are remarkably similar for all sequences and surprisingly uncorrelated with the mean FRET efficiency (Fig. 1C). To investigate the relation between chain dimensions and dynamics in mechanistic detail, we have performed multi-microsecond simulations of the entire set of linker proteins, both with and without the FRET chromophores included. We used a state-of-the-art force field (Amber ff99sbws), which has been developed to reproduce both secondary structure[24] and overall degree of compaction[25] of unfolded and disordered proteins. The force field reproduces the experimental trends in FRET efficiency almost quantitatively, and the chromophores only minimally perturb the conformational ensemble; however, inclusion of the chromophores does yield FRET efficiencies in closer accord with experiment. By further analysis of the discrepancies between experiment and simulation, we identify residual deficiencies in the force field to be addressed in the future, with salt bridge interactions within the proteins being the most significant factor.

Because all-atom explicit solvent simulations with a sufficiently accurate force field can reproduce absolute time scales, unlike coarse-grained models, we can use them to investigate the origin of the experimentally observed reconfiguration times. We show that the narrow range of reconfiguration times and lack of correlation with average FRET efficiencies seen in experiment is reproduced by the simulations. We then use the equilibrium and dynamic properties accessible in the simulations to rationalize the lack of correlation between reconfiguration times and FRET efficiencies by approximating the dynamics of the interchromophore distance as one-dimensional diffusion. As we show, compensating effects arising from the variation of diffusion coefficient and of chain compactness with sequence lead to the approximately sequence-independent reconfiguration times. We further cross-validate

our explanation against the available experimental data.

Results and Discussion

All-atom molecular simulations reproduce FRET and SAXS data for a diverse set of IDRs

We performed all-atom explicit-solvent simulations of the entire set of linker IDRs described previously[22] for over six microseconds each in the Amber ff99sbws force field[25] and conditions matched to experiment. To compute FRET efficiencies from the simulations with the fewest possible assumptions, we have explicitly included the chromophores in the molecular model (for sequences and dye positions, see Table S1). Parameters for the dyes Cy3B and CF660R were derived for this work using standard procedures as described in Methods. A key adjustment for the chromophores was a scaling of the interactions between the chromophores and water, analogous to that used for the protein force field, as discussed previously.[26,27]

The mean FRET efficiencies calculated based on the distance between the chromophores are compared with the corresponding experimental efficiencies in Fig. 2A, showing them to be in good accord. The difference between simulation and experiment is within the calculated standard errors for eight of the sixteen proteins, close to statistical expectations. We discuss possible reasons for residual discrepancies below.

As a further test of the simulations, we computed SAXS form factors for the four linker IDRs for which experimental data are available[22] (Fig. 2B, Fig. S1). Besides being a complementary technique to FRET, SAXS experiments can be performed without the need for conjugating the proteins to chromophores. We also observe good agreement between the SAXS curves from experiment and from simulations, independently confirming the quality of the simulations.

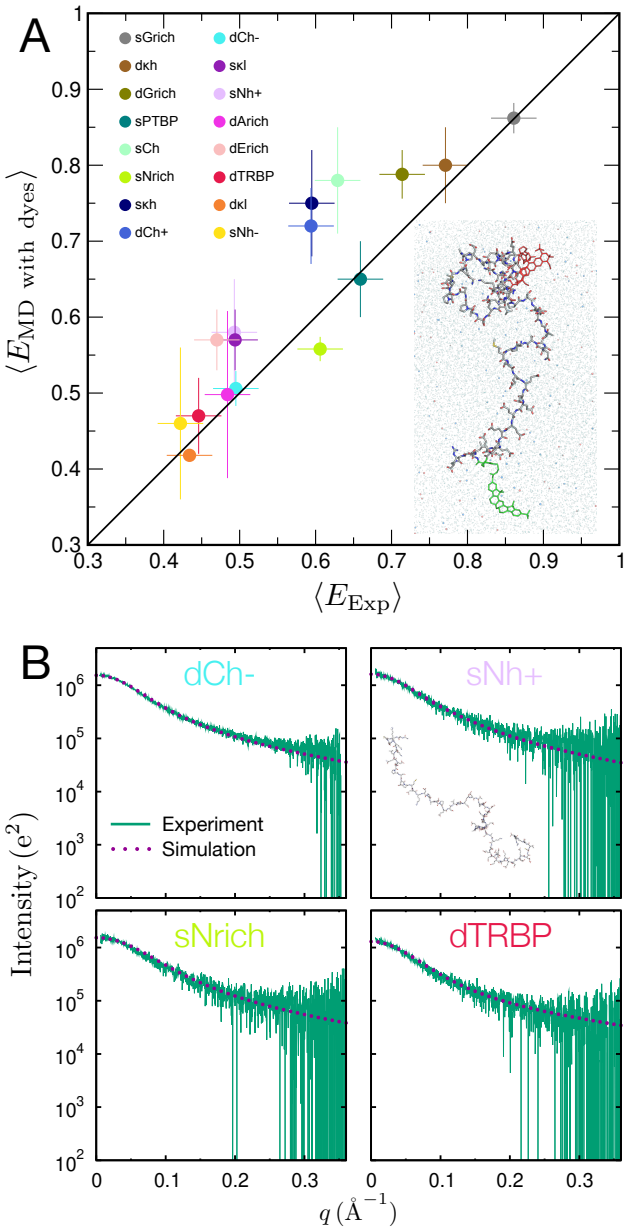

Figure 2: (A): Comparison between FRET efficiencies calculated from simulations with explicit chromophores and experimentally observed values (solid line: identity). Linear correlation coefficient and concordance correlation coefficient[28] are 0.90 and 0.84, respectively. Color code for IDRs is shown in the legend. The error bars represent one standard deviation estimated from the three independent simulation runs. Inset: A simulation snapshot showing the protein with explicit dyes, as well as the surrounding water and ions. The FRET donor and acceptor dyes are shown in green and red, respectively. K^+ ions are shown as blue spheres and Cl^- ions as red spheres. Water molecules are shown as light blue, transparent spheres. Hydrogen atoms were omitted for clarity. (B): Comparison of the experimental SAXS curves with those calculated from the simulations (both without dyes attached).

FRET chromophores minimally affect the dimensions of IDPs

A common concern in FRET experiments is the perturbation by the extrinsic chromophores introduced for the purpose of measuring distance distributions. Clearly, addition of a probe

must have some effect on the conformational distribution, and for some very hydrophobic probes, this effect has been clearly observed in experiment,[9] while in other cases the effect is more subtle.[22] In our previous work on these linker IDRs, we had used a coarse-grained force field trained against experimental data that can empirically account for the small perturbations of conformational distributions caused by commonly used chromophores, which indicated that Cy3B/CF660R, the pair of chromophores we are using here, has weaker interactions with the protein than an alternate pair of dyes, the AlexaFluors 488 and 594.[22] However, this force field optimization was based on a simplified "top-down" model driven only by experimental data, rather than by directly accounting for the detailed interactions involving the chromophores. To address the effect of the chromophores on the results, we thus performed all-atom simulations of all the linker IDRs, both with and without chromophores for a direct comparison. We estimated transfer efficiencies for the unlabelled proteins using the distance between the α carbon atoms of the labelled residues and scaling it to account for the effective increase in chain length from the cysteine side chains and dye linkers.[29]

We find that the inferred FRET efficiencies are very similar with and without the chromophores (Fig. 3A, Fig. S2), i.e., the effect of the dyes appears to be modest. However, there is an apparent shift to slightly lower transfer efficiency in simulations with explicit chromophores. This is certainly in contrast to the suggestion that chromophores cause proteins to collapse.[30,31] Note also that inclusion of the chromophores actually results in an improvement of the agreement with experiment (Fig. S2). To more directly compare the equilibrium ensembles with and without dyes, we computed the protein radii of gyration, R_g, in both cases (excluding chromophores from the calculation, Fig. 3B), and find no difference between them outside of statistical error. One can therefore infer that the slight shift in dye-dye separation to longer distances relative to the end-end distribution of the unlabelled protein primarily reflects the anisotropic distribution of configurations adopted by the chromophores in the context of the protein (see Fig. S3), as expected from their excluded volume, rather than a perturbation of the protein ensemble by attractive interactions.

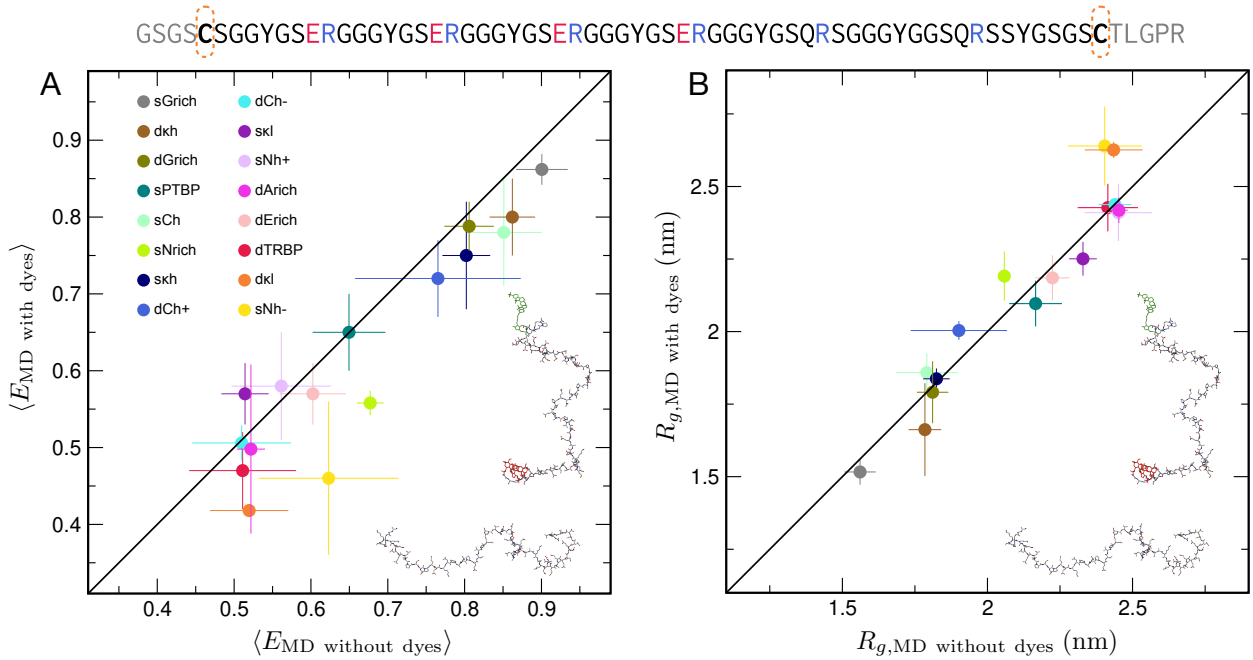

Figure 3: Effect of FRET dyes on protein configurations. (A) Comparison of average FRET efficiencies computed using the distances between the chromophore centers and the orientations between the dyes in simulations with FRET chromophores and the efficiencies estimated from simulations without dyes.[29] (B) Comparison of average protein radii of gyration (R_g) from simulations with and without chromophores (chromophores excluded from R_g calculation). Error bars represent the standard deviations estimated from the three independent simulation runs. An example of one of the IDR sequences (sGrich) with the Cys residues used for labelling indicated is given at the top (see Table S1 for all sequences), and examples of snapshots of a dye-labeled and unlabeled linker IDR are shown as insets.

Salt bridges appear to be responsible for most of the residual discrepancy from experiment

Although the overall agreement with the experimental FRET efficiencies is good, there are still some linker IDRs for which the efficiency computed from simulation is significantly different from the experimental value. To gain some further insight into the origin of these remaining discrepancies, we have examined the relation between the deviation from experiment and various properties of the sequences under consideration, as plotted in Fig. 4A and Fig. S4. The strongest correlations with the deviation from experiment are to the fraction of charged residues (FCR) and the number of possible salt bridges. Together, these observations suggest that the salt bridges in the simulations may be somewhat too strong, a deficiency which has been highlighted for several force fields.[32–37] To assess this hypothesis further,

we have tested the effect of reweighting the simulations based on the total number of salt bridges, N_{sb}, by weighting each simulation frame i as $w_i \propto \exp[-\beta\epsilon_{\mathrm{sb}}]$, where $\beta = 1/k_B T$, k_{B} is the Boltzmann constant, T is the temperature, and ϵ_{sb} is a test perturbation to salt bridge strength. As Fig. 4B shows, the optimal perturbation is $\epsilon_{\mathrm{sb}} \sim 0.75$. This small penalty for salt bridge formation suggests that salt bridges in the current force field are slightly too strong. If we instead use a correction with separate parameters for each type of salt bridge (Glu-Lys, Glu-Arg, Asp-Lys, Asp-Arg), $w_i \propto \exp[-\beta(\epsilon_{\mathrm{EK}} + \epsilon_{\mathrm{ER}} + \epsilon_{\mathrm{DK}} + \epsilon_{\mathrm{DR}})]$, we are able to improve the agreement with experiment slightly further, but still with only small perturbations to the salt bridge energies. We find $\epsilon_{\mathrm{EK}} \approx 1.29$, $\epsilon_{\mathrm{ER}} \approx 3.75$, $\epsilon_{\mathrm{DK}} \approx 0.04$ and $\epsilon_{\mathrm{DR}} \approx 0.71$, suggesting that any force field correction for salt bridges would likely need to account for the identity of the residues involved. The relatively small magnitude of the corrections required helps to explain the generally good agreement between experiments and simulations of complex coacervates of protein polyelectrolytes performed using the same force field.[14,38] However, the larger corrections involving arginine salt bridges may also help to explain why timescales for coacervates involving arginine-rich sequences are slightly too large relative to experiment in simulations with this force field.[38]

The already good agreement between the simulations and the FRET and SAXS experiments suggests that the simulation ensembles presented here are already of high accuracy. However, we can further improve the ensembles by incorporating the experimental data via reweighting to obtain an optimal combination of information from the force field and from experiment. To do so, we used a Bayesian reweighting procedure[41] which has successfully been employed in a number of recent studies,[42–44] whereby each frame of the simulation is independently reweighted so that their average matches the experimental observables while also subject to a penalty term keeping the weights as close to uniformity as possible. We employed two experimental observables for reweighting each linker domain simulation: the mean FRET efficiency and the variance of the FRET efficiency distribution obtained from fluorescence lifetime information.[43] As shown in Fig. S5 and Table S2, the reweighting is

able to reproduce these data for each sequence to be within experimental error while still keeping the weights relatively uniform across the trajectory, a testament to the quality of the force field. The resulting equilibrium distributions and mean values of the end-end distance are thus only modestly perturbed from the original simulations; the most notable changes being to linkers sCh, sκh and dCh+ (Fig. S5), as expected from the mean FRET efficiencies plotted in Fig. 2. Note that, for all 16 IDRs, the orientational factor κ^2 is very close to the expected value of 2/3 (for rapid isotropic rotational averaging of the dye transition dipoles) and remains unaffected by reweighting (Table S2), further confirming the quality of the simulations.

Simulations explain lack of correlation between reconfiguration times and transfer efficiencies

Quantifying conformational distributions is only one important component of understanding the functional behavior of IDPs. A complementary aspect is the role of chain dynamics, which can have a direct effect on binding kinetics and mechanisms.[45–47] To complement the previously measured transfer efficiencies of the linker IDRs,[22] we have thus probed the chain dynamics with nanosecond fluorescence correlation spectroscopy (nsFCS), which allows us to measure the chain reconfiguration time within the disordered ensemble.[10,23] Based on previous work on the dynamics in disordered proteins, we had expected a variation of these distance correlation times with sequence, because "internal friction" — dissipative forces within proteins that slow down their dynamics[48–50] — tends to be higher for more compact chains, slowing down their dynamics,[10,50–53] and the sequences under consideration here span a wide range of R_g. As a measure of the relation between compaction and dynamics, we plot the correlation times determined from nsFCS data versus FRET efficiencies in Fig. 5A. The correlation times lie within a band of roughly 20 to 30 ns and lack any discernible correlation with the FRET efficiencies (linear correlation coefficient -0.03), in apparent conflict with earlier findings of slower dynamics for more compact chains.[10,50]

We therefore used the simulations to understand the origin of this apparent discrepancy. To cast the simulation results in an easily interpretable form, we use a one-dimensional diffusion model to describe the dynamics of the inter-chromophore distance, and we use the dynamical data from the three separate simulation runs of each linker to optimize the parameters of the model to match the distance distributions and chain dynamics observed in the trajectories.[54] The optimal equilibrium distance distribution $P(r)$ and distance-dependent diffusion coefficients determined from the model are shown in Fig. 5B and 5C, respectively. In Fig. S6 we confirm that distance autocorrelation functions computed from this model are consistent with those estimated directly from the trajectories, with the advantage that the model correlation functions avoid the statistical noise seen in the long-time tails of those from the trajectories. As expected, the end-to-end distance distributions $P(r)$ shift to larger distances for proteins with lower FRET efficiencies. Less obviously, the diffusion coefficients tend to decrease at shorter distances, a phenomenon which is common to all the linker proteins, and which has been observed earlier in simulations of model peptides:[55] this may be considered one manifestation of internal friction effects.[10,52] We have used the diffusion model to compute inter-dye distance correlation times for each protein, finding the values to lie in a similarly narrow range as seen in experiment (Fig. 5A).

Position-dependent diffusion coefficients decreasing with r would be expected to lead to slower dynamics for more compact chains, as each protein is most sensitive to the value of $D(r)$ at the most populated regions of r, given by the shape of $P(r)$. Why is this not the case? A possible explanation can be simply described using the expression for the correlation time τ_c in a harmonic potential with constant diffusion coefficient D_{const},

$$\tau_c = \frac{\text{Var}[r]}{D_{\text{const}}} \tag{1}$$

where $\text{Var}[r]$ is the variance of the distance coordinate r. One can immediately see that changes both in the diffusion coefficient and in the variance, or width, of the distribution

of r can affect the correlation time. If the relative change in variance was similar to the change in diffusion coefficient, the effects could approximately cancel, i.e. for compact proteins, moving in a narrower potential could compensate a slower diffusion coefficient. In fact, as might be anticipated from Fig. 5B, the variance does indeed decrease in the simulations for proteins with increasing FRET efficiency (ρ=-0.86), and similarly, the SAW $-\ \nu$ model predicts smaller variance as FRET efficiency increases (Fig. 5D). A direct experimental measure of the width of the distributions is provided by the variance of the FRET efficiency, Var[E], determined from a combination of ratiometric and lifetime-based FRET measurements.[56] Indeed, the Var[E] from simulation and experiment are strongly correlated, with $\rho = 0.88$, confirming the observations from simulation. While this is a simplified picture, a similar effect is expected in the more general case of diffusion within a potential well with position-dependent diffusion coefficients. An interplay between collapse and diffusion is also expected from models of polymer dynamics, such as the Rouse and Zimm models, where chain relaxation times depend on the ratio of the chain dimensions and the diffusion coefficient determining the Brownian motion of the chain segments,[57,58] and its role has been invoked to explain the invariance of the reconfiguration time of the charged IDP prothymosin α as a function of salt concentration.[50]

While we have seen that the diffusion coefficients are not position-independent, they can be approximated as such, with the value of the effective constant diffusion coefficients being most sensitive to the diffusion coefficient near the peak in the distance distribution for each protein. If the diffusion model is refitted with the assumption of constant diffusion coefficients, we indeed find that the effective position-independent diffusion coefficient decreases with increasing FRET efficiency (ρ=-0.95), as expected (5E). The diffusion coefficient that would be inferred from the experimental τ_r and Var[r] from the SAW-ν model exhibits a very similar decrease. We can conduct a thought experiment to test the effect of these systematic changes in Var[r] and D_{const} by computing correlation times with Eq. 1 and assuming either (i) the same diffusion coefficient for all proteins (taken to be the average of those in Fig. 5E)

or (ii) identical Var[r] for all proteins (taken to be the average of those in Fig. 5D). In each of these cases, there is a strong linear correlation of the calculated τ_r with $\langle E_{\mathrm{MD}} \rangle$, with correlation coefficients -0.86 and 0.93 for cases (i) and (ii), respectively. Combining the effects of varying both Var[r] and D_{const} via Eq. 1 yields almost no correlation with FRET efficiency ($\rho = 0.05$), and in addition good agreement with τ_r computed from the full diffusion model (Fig. 5A). We can thus conclude that the small variation in τ_r, and consequent insignificant correlation with FRET efficiency, is a consequence of almost perfectly compensating effects from the narrowing distribution of end-end distances for more compact IDRs and a local reduction in $D(r)$ for shorter distances. We refer to these compensatory effects that help to maintain an almost invariant reconfiguration time as "dynamical buffering", in analogy to "conformational buffering", compensatory changes in amino acid sequence composition and sequence length that have been suggested to lead to the conservation of optimal tethering in a large family of disordered adenovirus early gene 1A protein (E1A) linkers.[7]

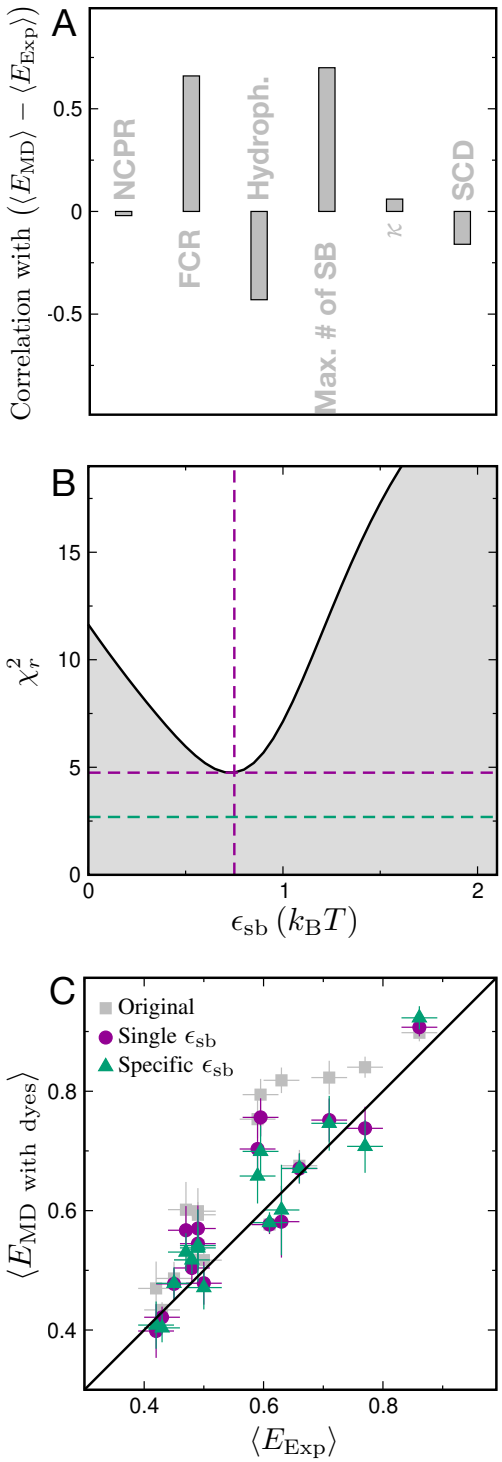

Figure 4: Identifying sources of residual discrepancies between simulated and experimentally measured FRET efficiencies. (A): The correlation between the deviation of simulated and experimental FRET efficiencies and sequence properties commonly used to characterize disordered proteins: net charge per residue (NCPR), fraction of charged residues (FCR), hydrophobicity, number of possible salt bridges, κ parameter[39] and sequence charge decoration (SCD).[40] (B) and (C): The result of the optimization of a uniform salt bridge energy correction parameter, ϵ_{sb}, used for reweighting the simulations based on the number of salt bridges. Effect of salt bridge energy corrections using a uniform parameter for all salt-bridges is shown in purple, and a correction using four individual parameters characterizing each possible type of inter-residue salt bridge is shown in green (see text for details).

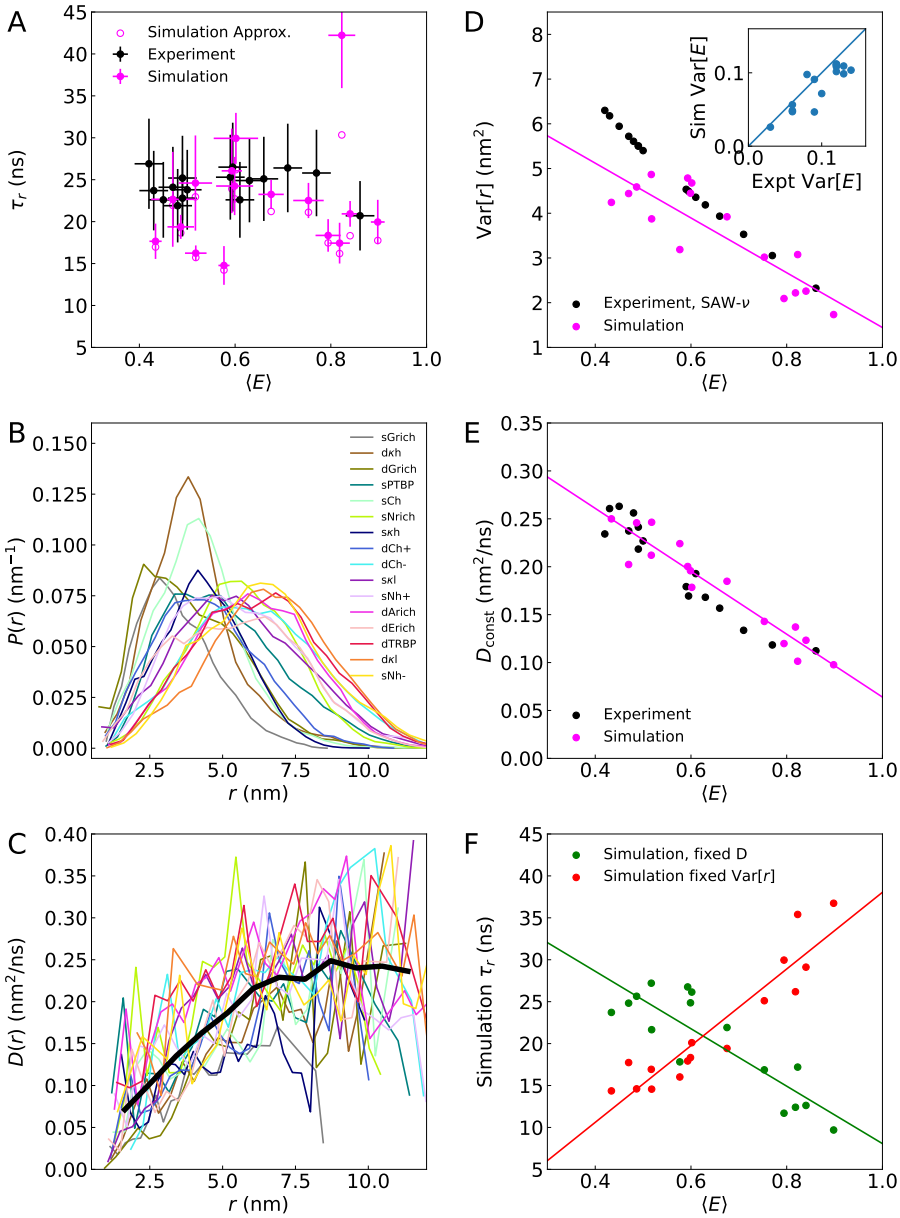

Figure 5: Rationalizing the lack of correlation between chain reconfiguration times and FRET efficiencies. (A) Weak correlation between reconfiguration times derived from nsFCS and FRET efficiencies from experiment (black, $\rho = -0.03$), and between donor-acceptor distance autocorrelation times and FRET efficiencies in simulation (pink, $\rho = 0.20$). Distributions of end-end distances (B) and position-dependent diffusion coefficients (C) inferred from a one-dimensional diffusion model applied to the simulation data. The thick black line in (C) represents the position-dependent diffusion coefficient averaged over all protein sequences. (D) Correlation between variances of end-to-end distance distributions and transfer efficiencies from simulation (pink), and between variances of distances inferred from SAW-ν model and transfer efficiencies from experiment (black). (inset) Correlation between Var[E] obtained from simulation with that from experimental lifetime measurements. (E) Correlation between position-independent diffusion coefficients and transfer efficiencies from simulation (pink) or experiment (black). Diffusion coefficients from simulation are inferred via optimization of a 1D diffusion model, while from experiment, $D_{\text{const}} = \text{Var}_{\text{SAW}-\nu}[r]/\tau_r$. (F) Correlation times estimated from simulation via $\tau_c = \text{Var}[r]/D_{\text{const}}$, showing strong correlations if either Var[r] or D_{const} is fixed to the average value over all proteins from the simulations ($\rho = 0.93$ and -0.86, respectively).

Discussion

The functional versatility of IDPs comes from both the broad equilibrium ensemble of structures which they populate as well as how rapidly they explore it to find, form, and release interactions. These properties are exquisitely sensitive to the underlying sequence as well as environment. The set of intrinsically disordered linker IDRs explored here showcases the diversity of ensembles that are populated, ranging from highly expanded sequences near the excluded volume limit to sequences that are slightly more compact than ideal chains, as characterized by single-molecule FRET spectroscopy. Since FRET measures properties averaged over a broad ensemble of structures, it is essential to complement such experiments with a theoretical or simulation model to aid in their interpretation. Analytical polymer models and coarse-grained molecular simulation models provide a starting point for analyzing structure and dynamics,[10] but all-atom simulations provide the highest spatial and temporal resolution and are able to predict both structural ensembles as well as absolute timescales of dynamics, provided the force fields used are of sufficient quality. However, it often is a challenge to sample long enough time scales for accurate comparison to experiment.

In this work, we have successfully used long-timescale all-atom simulations to characterize this set of 16 IDRs. Our simulations agree quantitatively with both single–molecule FRET efficiencies and, where available, SAXS, providing a consistent validation from complementary experiments. Further, by modeling the FRET dyes explicitly at atomistic detail, we find that the chromophores have a minimal influence on ensemble dimensions in these systems—consistent with the current view[59] and contrary to suggestions that common fluorophores collapse disordered chains.[30] In fact, including dyes in the simulations yields closer correspondence to the measured FRET efficiencies without distorting the protein ensemble, in line with broader assessments of smFRET reliability.[27] In contrast to the situation that would be expected if the dyes caused collapse, the FRET efficiencies including the explicit chromophores are systematically *lower* than without them. Taken together, this highlights that more rigorous methods of back-calculating experimental observables from simulations

do result in better agreement with experiment.[27,60]

Nanosecond fluorescence correlation spectroscopy (nsFCS) combined with single–molecule FRET directly reports on end–to–end distance relaxation, or reconfiguration times, which are commonly on tens–of–nanosecond timescales for disordered proteins.[10] By varying solution conditions and interpreting the results with polymer models, it has been possible to separate solvent friction effects from internal friction due to transient intrachain contacts. Extensive previous work has established that increased compaction tends to raise internal friction and slow intrachain motions.[50,53,61,62] Surprisingly, however, the reconfiguration times of the 16 linker IDRs reported here cluster around similar values despite their large sequence–encoded differences in equilibrium compaction. All–atom explicit–solvent simulations reproduce this small variation in time scales together with the changes in compaction. The simulations can furthermore explain the apparent paradox with a simple mechanism: while more compact ensembles indeed exhibit reduced local diffusivity, in accord with the established internal friction picture, the narrower end-to-end distribution for those proteins has the opposite effect. Because reconfiguration times can be approximated as $\tau_r \sim \frac{\mathrm{Var}[r]}{D_{\mathrm{eff}}}$, the decrease in variance offsets the decrease in effective diffusivity, preserving a nearly constant τ_r across sequences. Note that this observed compensation may not extend to scenarios of more extreme compaction approaching the globule limit, where internal friction effects are felt most strongly, and a sharper reduction in diffusion coefficient may be expected. Indeed, simulations of a coarse-grained model with explicit solvent also showed a dynamical buffering effect, with almost constant reconfiguration times over a range of solvent quality and degree of chain collapse, and a sharp slowdown only being observed once the chain approaches the most compact extreme as $\nu \to 1/3$.[52]

The quantitative description of both equilibrium structure and dynamics in disordered protein ensembles which we present here has been enabled by progressive refinement of all–atom MD force fields over the years so as to correct systematic biases that hampered earlier simulations.[63] Most notably, secondary structure biases that favored either α-helical

or β-sheet structure[64] and the tendency for unfolded proteins to be too collapsed[65] were obviously detrimental for simulations of disordered proteins. Guided by empirical data, however, it was possible to improve so as to match experiment.[25,66,67] The present set of 16 disordered linker IDRs presents a particularly stringent test of these improvements, as all are of the same length, and thus only sequence effects are responsible for the differences between them. The good agreement of the simulations with experiment, both in terms of overall dimensions and reconfiguration times strongly validates the force field improvements. Nonetheless, analysis of residual differences from experiment suggests that there are some discrepancies outside of the statistical noise, and these appear to be most highly correlated with the presence of oppositely charged residues. Simple reweighting procedures adding a correction term for salt bridge formation suggest that salt bridges, particularly those involving arginine, are slightly too strong in the force field, providing an avenue for future force field improvement.

The insights obtained here into the dynamics of disordered linkers may also have implications for their biological function. In signaling and regulation, preserving fast chain reconfiguration regardless of average compaction ought to facilitate conformational search, encounter–complex formation, and rebinding.[1] For example, in high–affinity disordered or "fuzzy" complexes, polyelectrolyte interactions can enable strong binding with rapid on/off kinetics, precisely the context where maintaining rapid reconfiguration is advantageous.[45,46] The dynamical buffering which conserves these rapid timescales even as configurational properties are modulated by sequence, also allows for robustness during evolution, reminiscent of the "conformational buffering" role proposed for the sequences of disordered linker regions.[7]

Methods

All-atom simulations

Extended all-atom configurations of all 16 IDR variants without dyes were generated using *tleap* from the Amber package.[68] The resulting structures were then placed in 14-nm rhombic dodecahedron boxes. To prevent self-interaction across periodic boundaries, short vacuum simulations were carried out for each IDR, and conformations with a maximum end-to-end distance below 12.5 nm were retained. Finally, for each of the 16 IDRs, three of these conformations were randomly chosen to initiate all-atom explicit-solvent simulations. The selected structures were energy-minimized using the steepest-descent algorithm. Each simulation box was then solvated with the TIP4P2005s water,[25] and the systems were again energy-minimized using the steepest-descent algorithm. Subsequently, the desired number of potassium and chloride ions were inserted to obtain the 172 mM KCl salt concentrations, matching the total ionic strength of the KCl salt and buffer used in the experiment. Subsequently, the system was again energy-minimized. The total number of atoms per simulation was approximately 250,000, varying slightly among different proteins. All simulations were performed using GROMACS[69] version 2021.5. Protein interactions were modeled using Amber99SBws,[25] with protein-dye and dye-dye interaction parameters described previously[26] and using optimized dye-water interaction parameters.[27] The integration time step was 2 fs. The temperature was kept constant at 295.15 K using velocity rescaling[70] ($\tau = 1$ ps), and the pressure was kept at 1 bar using the Parrinello-Rahman barostat[71] ($\tau = 5$ ps). Long-range electrostatic interactions were modeled using the particle-mesh Ewald method.[72] Dispersion interactions and short-range repulsion were described by a Lennard-Jones potential with a cutoff at 1 nm. H-bond lengths were constrained using the LINCS algorithm.[73] Each independent run was at least 2.05 μs, and the first 50 s were treated as equilibration and were omitted from the analysis.

For all 16 IDRs, the protein coordinates from each of the three independent runs without

dyes were extracted after 100 ns of simulation time and used as starting structures for simulations with dyes. The explicit Cy3B and CF660R dyes were attached using custom Python scripts. All subsequent steps—including energy minimization, solvation with TIP4P2005s water, and addition of potassium and chloride ions to reach 172 mM KCl—were identical to those used for the simulations without dyes. Each independent run was at least 2.02 μs long, with the first 20 ns treated as equilibration and omitted from the analysis.

The mean transfer efficiency, $\langle E \rangle$, was obtained from simulations with dyes as described previously.[26,74] In short, we computed the survival probability of the donor excited state with a time-dependent transfer rate and averaged it over all possible excitation times t_0 along each trajectory:

$$p_{D^*}(t) = \left\langle \exp\left(- \int_{t_0}^{t_0+t} \left[k_D + k_F(t') \right] \mathrm{d}t' \right) \right\rangle_{t_0}, \tag{2}$$

where

$$k_F(t') = k_D \frac{\kappa^2(t')}{2/3} \left(\frac{R_0}{r(t')} \right)^6. \tag{3}$$

Here, κ^2 is the orientational factor for the donor–acceptor dye pair, r is the instantaneous inter-dye distance, and R_0 is the Förster radius defined for $\kappa^2 = 2/3$. The dyes were treated as non-emissive whenever any donor–acceptor atom pair was closer than 0.4 nm (van-der-Waals contact). The average FRET efficiency was then computed by integrating the survival probability:

$$\langle E \rangle = 1 - k_D \int_0^{t_{\max}} p_{D^*}(t) \, \mathrm{d}t \tag{4}$$

with $t_{\max} = 20$ ns chosen such that $p_{D^*}(t)$ had effectively decayed to zero.

The inter-dye distance r was estimated from the simulations without dyes as

$$r = d \left(\frac{N+9}{N} \right)^{\nu} \tag{5}$$

where d is the Cα–Cα distance between the labeled residues). Here, N is the sequence

separation between the labeling sites, and we used a scaling exponent $\nu = 0.6$. This choice effectively accounts for the dye and linker lengths by adding nine "effective" residues to N. The estimate is only weakly sensitive to ν: varying ν by ± 0.1 changes the inferred transfer efficiency by roughly ± 0.01. For transfer efficiency calculations from simulations without explicit FRET dyes, we assumed $\kappa^2 = 2/3$.

The number of possible salt bridges (Fig. 4) was computed as the count of oppositely charged residue pairs within each IDR sequence, considering only pairs separated by at least three residues along the sequence.

Residue–residue contact fractions (Fig. S3) were evaluated using a transition/core-state approach.[75] Instead of a single cutoff separating bound and unbound states, distinct thresholds were applied for contact formation and breaking. For each residue pair, the distance metric was the minimum heavy-atom distance between the two residues. Starting from an unformed state, a contact was considered formed when this distance dropped below 0.38 nm; once formed, it remained in the contact state until the distance exceeded 0.8 nm.[75]

Force field parameters for Cy3B and CF660R

Parameters for the chromophores Cy3B and CF660R, and their cysteine-conjugated linker, were derived using the Antechamber program from the AmberTools package.[76] The labelled residue was split into a part corresponding to the linker-conjugated cysteine, and a second part corresponding to either the Cy3B or CF660R chromophore. Following the nomenclature of the AMBER-DYES force field,[77] the internal Cysteine and linker were denoted "C2R", while Cy3B and CF660R were respectively named "C3B" and "CF6". Because the C2R residue in the AMBER-DYES force field is much more polarized relative to the standard AMBER94 cysteine,[78] and this is known to be correlated with aberrant secondary structure formation,[24] we derived new charges for the C2R residue by constraining the backbone charges to be the same as the others in AMBER94-derived force fields, basing the calculations on surface electrostatic potentials determined with Gaussian 16 using Hartree-Fock with the

6-31G* basis set. The chromophore parameters were obtained using a similar process to those for C2R, constraining the charges of the amide conjugation to the linker to be the same as for the protein backbone, but using a 6-31+G* basis set (this having previously given better results for AlexaFluor dyes[26]). Scripts used to generate the new parameters and the parameters themselves are available on the accompanying Zenodo repository, as is the script for generating initial structures of the dye-conjugated proteins.

Reweighting to estimate effect of salt bridge strength

We assume that the the true energy of each salt bridge, considered for simplicity as a contact energy, is $E_{\mathrm{sb}} = E_{\mathrm{sb,ff}} + \epsilon_{\mathrm{sb}}$. Assuming all other terms in the energy function are unchanged, the total difference in energy between the correct salt bridge energy and the force field salt bridge energy would be $\Delta E = N_{\mathrm{sb}}\epsilon_{\mathrm{sb}}$ (in units of $k_{\mathrm{B}}T$). Each frame i of the simulation, having N_{sb} salt bridges, can then be reweighted using weight $w_i \propto \exp[-N_{\mathrm{sb}}(i)\epsilon_{\mathrm{sb}}]$, allowing observables such as FRET efficiency to be computed with the correction. A simple scan over values of ϵ_{sb} identifies the optimal value as that which best matches the reweighted simulation efficiencies to experiment.

We also considered the case where specific corrections are made for each of the four types of salt bridge, i.e. $\Delta E = N_{\mathrm{EK}}\epsilon_{\mathrm{EK}} + N_{\mathrm{ER}}\epsilon_{\mathrm{ER}} + N_{\mathrm{DK}}\epsilon_{\mathrm{DK}} + N_{\mathrm{DR}}\epsilon_{\mathrm{DR}}$. In this case the optimal parameters were determined by gradient descent using an adaptive step size.[79] The analytical gradients of χ^2 for each protein with respect to ϵ_{AB} are given by

$$\partial\chi^2/\partial\epsilon_{\mathrm{AB}} = 2(\langle E\rangle_{\mathrm{rw}} - E_{\mathrm{expt}})(\langle E\rangle_{\mathrm{rw}}\langle n_{\mathrm{AB}}\rangle_{\mathrm{rw}} - \langle En_{\mathrm{AB}}\rangle_{\mathrm{rw}}), \tag{6}$$

where the average $\langle\cdot\rangle_{\mathrm{rw}}$ indicates the reweighted average using the current contact energy parameters. The total gradient is the sum of Eq. 6 over all the proteins. In practice, we minimized the target function $L = \chi^2 + \frac{1}{2}\gamma(\epsilon_{\mathrm{EK}} + \epsilon_{\mathrm{ER}} + \epsilon_{\mathrm{DK}} + \epsilon_{\mathrm{DR}})^2$, where the constant γ, chosen to be 1.0, penalizes large values of the contact energies.

Bayesian reweighting

Simulations with dyes were reweighted using Bayesian inference of ensembles (BioEn)[80] to reach agreement with both the means and the variances of the transfer efficiency distributions observed experimentally. It is worth emphasizing, as already discussed in Ref. 43, that we use for this approach not the variance of the transfer efficiency histogram, but the variance of the transfer efficiency distribution that corresponds to the underlying distance distribution, which can be obtained from the deviations of the mean fluorescence lifetimes from the static FRET line.[56,81,82] In this way, the reweighting takes into account not only experimental information on the average intramolecular distance but also on the variance of the distance distribution. Instantaneous transfer efficiency and κ^2 were determined from simulations using the dye coordinates every 20 ps. The uncertainty of the experimental transfer efficiency was set to 0.03 and the uncertainty of the variance of the transfer efficiency distribution was set to 0.02. Reweighting parameters and results are shown in Table S2. Histograms of the initial and reweighted distance distributions are shown in Fig. S5.

SAXS calculations

To calculate SAXS curves (Fig. 2B), we saved protein structures every 5 ns from all runs of all simulations (1200 PDB files per IDR). Subsequently, we used CRYSOL[83] with default parameters to compute SAXS curves for each structure. We note that the experimental curve was not provided to CRYSOL, to avoid any overfitting that could arise from adjusting the hydration-layer density to better match the experimental SAXS profile.[84] For each IDR, the 1200 calculated curves were then averaged, and the error at every q point was determined as the standard deviation across the set of curves. To overlay experimental and computed SAXS curves, we scaled only the absolute intensity of the experimental curve, without applying any additional corrections for possible buffer mismatch.[85] We note that explicit consideration of the solvent environment typically yields protein scattering curves very similar to those from CRYSOL, likely because the solvation shell is highly correlated with the path of the protein

chain.[86] Experimental SAXS curves have been taken from the previous study.[22]

Computing reconfiguration times

Dynamics of the interchromophore distance was analyzed with a one-dimensional diffusion model.[38,54] Briefly, this distance coordinate was discretized into 30 equal-size, non-overlapping bins between the minimum and maximum sampled value for each trajectory. Histograms of observed transitions between bin i at time t and bin j at time $t+\Delta t$, $N_{ji}(\Delta t)$ were determined for different lag times Δt. Discretized free energies and position-dependent diffusion coefficients were optimized via Monte Carlo simulations using the likelihood function

$$lnL = \sum_{i,j} N_{ji}(\Delta t) \ln p(j, t+\Delta t|i,t) \tag{7}$$

where the propagators $p(j, t+\Delta t|i,t)$ describing the conditional probability of being in bin j a time Δt after having been in bin i are obtained from the discretized diffusion model as previously described:[38,54] in short, the discretized dynamics is mapped to a chemical kinetics scheme describing evolution of populations in the bins, $\dot{\mathbf{P}} = \mathbf{K}\mathbf{P}(t)$ where $\mathbf{P}(t)$ is the vector of the bin populations at time t, and $\mathbf{K}$ is a rate matrix derived from the diffusion coefficient(s) D_i and free energies F_i associated with each bin according to the scheme of Bicout and Szabo.[54,87] The propagators are then given by $p(j, t+\Delta t|i,t) = (\exp[\Delta t\mathbf{K}])_{ji}$. In estimating the most probable parameters from the data, a uniform prior is used for the diffusivities and free energies. We can compute normalized correlation functions

$$C(t) = \frac{\sum_{n=2}^{b} (\mathbf{r} \cdot \Psi_n^R)^2 e^{\lambda_n t}}{\sum_{n=2}^{b} (\mathbf{r} \cdot \Psi_n^R)^2}, \tag{8}$$

where the elements of r are the centers of each bin on the distance coordinate, Ψ_n^R is the nth right eigenvector of $\mathbf{K}$ (Ψ_1^R is the stationary eigenvector), and λ_n is the nth eigenvalue.

From this, the correlation times follow:

$$\tau = \frac{\sum_{n=2}^{b}(\mathbf{r} \cdot \Psi_n^R)^2 \lambda_n^{-1}}{\sum_{n=2}^{b}(\mathbf{r} \cdot \Psi_n^R)^2} \tag{9}$$

In determining the optimal parameters, we fitted simultaneously to lag times of 5, 10 and 20 ns, so as to include both short and long time correlations; the convergence of the relaxation times with respect to the lag time for fits to single lag times is shown in Fig. S7.

Single-molecule fluorescence spectroscopy

A set of 16 naturally occurring IDRs, each comprising 57 amino acid residues, was selected from the linker regions connecting folded domains in RNA-binding proteins.[22] For fluorescence labeling, cysteine residues were introduced at the the N- and C-termini of each IDR sequence. The constructs were recombinantly expressed in *Escherichia coli*, purified, and subsequently labeled with Cy3B and CF660R fluorophores via maleimide chemistry as described previously.[22] For single-molecule experiments, donor–acceptor–labeled peptides were diluted in 20 mM KH_2PO_4/K_2HPO_4, 125 mM KCl, pH 7.3, supplemented with 0.001% Tween-20 and 10 mM DTT. Measurements were performed at 22 °C either in chambered cover slides (μ-Slide, ibidi) at roughly 0.1 nM concentrations or, when possible, in zero-mode waveguides (ZMW) at high nM concentrations to reduce acquisition times.[60]

Measurements were carried out either on a MicroTime 200 (PicoQuant) or on a custom-built confocal single-molecule instrument. Excitation lasers were operated in continuous-wave mode. The MicroTime 200 was equipped with an Olympus UplanApo 60×/1.20 W objective. Donor excitation was achieved using an Oxxius laser at 532 nm attenuated to 50 μW at the back aperture of the objective. Fluorescence was collected through the same objective and separated from backscattered light by a triple-band mirror (ZT405/530/630RPC, Chroma Technology). Residual excitation light was further suppressed with a long-pass filter (LP532, Chroma Technology) before passing through a 100 μm pinhole. Emission was

split into four detection channels using a 50/50 or polarizing beam splitter and two dichroic mirrors (635DCXR, Chroma Technology). Donor and acceptor signals were subsequently filtered with an ET585/65M band-pass (Chroma Technology) and LP647RU long-pass filter (Chroma Technology) / 750 nm blocking edge BrightLine multiphoton short-pass emission filter (Semrock), respectively, and detected by four single-photon avalanche diodes (SPCM-AQR-15, PerkinElmer Optoelectronics). Photon arrival times were recorded with 16 ps resolution using a HydraHarp 400 TCSPC module (PicoQuant). In the custom-built instrument, the excitation source and optical filters were identical to those in the commercial instrument.[88]

For data analysis, we used only photons from bursts of molecules double-labeled with active donor and acceptor fluorophores, which reduces the contribution of donor-only and acceptor-only signal to the correlation. The selection was based on the transfer efficiency. All experimental transfer efficiencies reported here are from Holla et al.,[22] using a Förster radius of 6.0 nm. The correlation between two time-dependent intensity signals, $I_a(t)$ and $I_b(t)$, recorded on two detectors, a and b, is defined as

$$G_{ab}(\tau) = \frac{\langle \delta I_a(t)\, \delta I_b(t+\tau) \rangle}{\langle I_a(t) \rangle \, \langle I_b(t) \rangle}, \tag{10}$$

where the angle brackets denote time averaging. In our experiments, two donor detection channels and two acceptor detection channels were used, yielding the autocorrelation functions $G_{dd}(\tau)$ and $G_{aa}(\tau)$. In addition, photons detected in both donor and acceptor channels were used to calculate the cross-correlation function $G_{da}(\tau)$. The resulting nsFCS curves were computed and fitted with τ_{cd} as a global parameter over a linear range of lag times from 0 ns to +100 ns using the fit equation

$$G_{ab}(\tau) = \frac{1}{N} \left[c_{ab}\, e^{-\tau/\tau_{ab}} + c_{cd}\, e^{-\tau/\tau_{cd}} \right], \tag{11}$$

where N is a normalization factor. The first term, with amplitude c_{ab} and characteristic

timescale τ_{ab}, describes photon antibunching, and the second term, with amplitude c_{cd} and timescale τ_{cd}, describes chain dynamics.

We analyzed correlation times as described previously,[43,58] where the distance dynamics are treated as diffusive motion in the potential of mean force associated with the distance distribution. In this framework, for any distance-dependent observable, X, the correlation time, τ_X, is defined as

$$\tau_X = \frac{\int_0^\infty P(r)^{-1} \left[\int_0^r \delta X(\rho) P(\rho)\, d\rho\right]^2 dr}{D \int_0^\infty \delta X(\rho)^2 P(r)\, dr}, \tag{12}$$

where D is the effective dye-to-dye diffusion coefficient, and $\delta X(r) = X(r) - \langle X(r) \rangle_r$. The experimentally measured intensity correlation time from nsFCS, τ_E, corresponds to the observable $X = E$, the transfer efficiency. To obtain the chain reconfiguration time, τ_r, which corresponds to correlation time of the dye-to-dye distance, $X = r$, we computed the conversion factor $\vartheta = \tau_r/\tau_E$ by evaluating the general expression for the correlation time τ_X with the distance distribution approximating a self-avoiding walk model of a polymer chain, characterized by the Flory scaling exponent ν.[21] ϑ, which depends on the chosen distance distribution, the measured transfer efficiency, E, and the Förster Radius, R_0 (here 6.0 nm), was then applied to convert the measured τ_E into τ_r.

Acknowledgement

We thank Daniel Nettels for help with optimizing data analysis tools and Kresten Lindorff-Larsen for discussions based on Martini simulations. RB was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health (NIH). The contributions of the NIH author were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered Works of the United States Government. However, the findings and conclusions presented in this paper are those of the authors and do not

necessarily reflect the views of the NIH or U.S. Department of Health and Human Services. BS was supported by the Swiss National Science Foundation. We used the computational resources of Piz Daint, Alps, and Eiger at the CSCS Swiss National Supercomputing Center, and of the National Institutes of Health HPC Biowulf cluster (http://hpc.nih.gov).

Supporting Information Available

A supplementary file containing seven figures and two tables is provided with the manuscript.

References

(1) Holehouse, A. S.; Kragelund, B. B. The molecular basis for cellular function of intrinsically disordered protein regions. *Nature Reviews Molecular Cell Biology* **2024**, *25*, 187–211.

(2) Abyzov, A.; Blackledge, M.; Zweckstetter, M. Conformational dynamics of intrinsically disordered proteins regulate biomolecular condensate chemistry. *Chemical Reviews* **2022**, *122*, 6719–6748.

(3) Coskuner-Weber, O.; Mirzanli, O.; Uversky, V. N. Intrinsically disordered proteins and proteins with intrinsically disordered regions in neurodegenerative diseases. *Biophysical Reviews* **2022**, *14*, 679–707.

(4) Tycko, R. The evolving role of solid state nuclear magnetic resonance methods in studies of amyloid fibrils. *Current Opinion in Structural Biology* **2025**, *92*, 103043.

(5) Alberti, S.; Hyman, A. A. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nature reviews Molecular cell biology* **2021**, *22*, 196–213.

(6) Visser, B. S.; Lipiński, W. P.; Spruijt, E. The role of biomolecular condensates in protein aggregation. *Nature Reviews Chemistry* **2024**, *8*, 686–700.

(7) Gonzalez-Foutel, N. S.; Glavina, J.; Borcherds, W. M.; Safranchik, M.; Barrera-Vilarmau, S.; Sagar, A.; Estaña, A.; Barozet, A.; Garrone, N. A.; Fernandez-Ballester, G.; others Conformational buffering underlies functional selection in intrinsically disordered protein regions. *Nature structural & molecular biology* **2022**, *29*, 781–790.

(8) Sibille, N.; Bernado, P. Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. *Biochemical Society Transactions* **2012**, *40*, 955–962.

(9) Borgia, A.; Zheng, W.; Buholzer, K.; Borgia, M. B.; Schüler, A.; Hofmann, H.; Soranno, A.; Nettels, D.; Gast, K.; Grishaev, A.; Best, R. B.; Schuler, B. Consistent view of polypeptide chain expansion in chemical denaturants from multiple experimental methods. *J. Am. Chem. Soc.* **2016**, *138*, 11714–11726.

(10) Schuler, B.; Soranno, A.; Hofmann, H.; Nettels, D. Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins. *Annu. Rev. Biophys.* **2016**, *45*, 207–231.

(11) Best, R. B. Computational and theoretical advances in studies of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2017**, *42*, 147–154.

(12) Henriques, J.; Cragnell, C.; Skepö, M. Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment. *J. Chem. Theor. Comput.* **2015**, *11*, 3420–3431.

(13) Shrestha, U. R.; Juneja, P.; Zhang, Q.; Gurumoorthy, V.; Borreguero, J. M.; Urban, V.; Cheng, X.; Pingali, S. V.; Smith, J. C.; O'Neill, H. M.; Petridis, L. Generation of the

configurational ensemble of an intrinsically disordered protein from unbiased molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 20446–20452.

(14) Galvanetto, N.; Ivanović, M.; Chowdhury, A.; Sottini, A.; Nüesch, M. F.; Nettels, D.; Best, R. B.; Schuler, B. Extreme dynamics in a biomolecular condensate. *Nature* **2023**, *619*, 876–883.

(15) Marsh, J. A.; Forman-Kay, J. D. Sequence Determinants of Compaction in Intrinsically Disordered Proteins. *Biophys. J.* **2010**, *98*, 2383–2390.

(16) Das, R. K.; Ruff, K. M.; Pappu, R. V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 102–12.

(17) Best, R. B. Emerging consensus on the collapse of unfolded and intrinsically disordered proteins in water. *Curr. Opin. Struct. Biol.* **2020**, *60*, 27–38.

(18) Tesei, G.; Trolle, A. I.; Jonsson, N.; Betz, J.; Knudsen, F. E.; Pesce, F.; Johansson, K. E.; Lindorff-Larsen, K. Conformational ensembles of the human intrinsically disordered proteome. *Nature* **2024**, *626*, 897–904.

(19) Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; others A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **2018**, *174*, 688–699.

(20) Saar, K. L.; Morgunov, A. S.; Qi, R.; Arter, W. E.; Krainer, G.; Lee, A. A.; Knowles, T. P. Learning the molecular grammar of protein condensates from sequence determinants and embeddings. *Proceedings of the National Academy of Sciences* **2021**, *118*, e2019053118.

(21) Zheng, W.; Zerze, G.; Borgia, A.; Mittal, J.; Schuler, B.; Best, R. B. Inferring properties of disordered chains from FRET transfer efficiencies. *Biophysical Journal* **2018**, *114*, 367a.

(22) Holla, A.; Martin, E. W.; Dannenhoffer-Lafage, T.; Ruff, K. M.; Konig, S. L.; Nuesch, M. F.; Chowdhury, A.; Louis, J. M.; Soranno, A.; Nettels, D.; others Identifying Sequence Effects on Chain Dimensions of Disordered Proteins by Integrating Experiments and Simulations. *JACS Au* **2024**, *4*, 4729–4743.

(23) Nettels, D.; Gopich, I. V.; Hoffmann, A.; Schuler, B. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2655–2660.

(24) Best, R. B.; de Sancho, D.; Mittal, J. Residue-specific α-helix propensities from molecular simulation. *Biophys. J.* **2012**, *102*, 1462–1467.

(25) Best, R. B.; Zheng, W.; Mittal, J. Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theor. Comput.* **2014**, *10*, 5113–5124.

(26) Best, R. B.; Hofmann, H.; Nettels, D.; Schuler, B. Quantitative interpretation of FRET experiments via molecular simulation: force field and validation. *Biophys. J.* **2015**, *108*, 2721–2731.

(27) Nüesch, M.; Ivanović, M. T.; Nettels, D.; Best, R. B.; Schuler, B. Accuracy of distance distributions and dynamics from single-molecule FRET. *Biophysical Journal* **2025**,

(28) Lawrence, I.; Lin, K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **1989**, 255–268.

(29) Aznauryan, M.; Delgado, L.; Soranno, A.; Nettels, D.; Huang, J. R.; Labhardt, A. M.; Grzesiek, S.; Schuler, B. Comprehensive structural and dynamical view of an unfolded

protein from the combination of single-molecule FRET, NMR, and SAXS. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E5389–98.

(30) Riback, J. A.; Bowman, M. A.; Zmyslowski, A. M.; Knoverek, C. R.; Jumper, J. M.; Hinshaw, J. R.; Kaye, E. B.; Freed, K. F.; Clark, P. L.; Sosnick, T. R. Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science* **2017**, *358*, 238–241.

(31) Riback, J. A.; Bowman, M. A.; Zmyslowski, A. M.; Plaxco, K. W.; Clark, P. L.; Sosnick, T. R. Commonly used FRET fluorophores promote collapse of an otherwise disordered protein. *Proceedings of the National Academy of Sciences* **2019**, *116*, 8889–8894.

(32) Debiec, K. T.; Gronenborn, A. M.; Chong, L. T. Evaluating the strength of salt bridges: a comparison of current biomolecular force fields. *The Journal of Physical Chemistry B* **2014**, *118*, 6561–6569.

(33) Yoo, J.; Aksimentiev, A. Improved parameterization of amine–carboxylate and amine–phosphate interactions for molecular dynamics simulations using the CHARMM and AMBER force fields. *Journal of chemical theory and computation* **2016**, *12*, 430–443.

(34) Ahmed, M. C.; Papaleo, E.; Lindorff-Larsen, K. How well do force fields capture the strength of salt bridges in proteins? *PeerJ* **2018**, *6*, e4967.

(35) Panel, N.; Villa, F.; Fuentes, E. J.; Simonson, T. Accurate PDZ/peptide binding specificity with additive and polarizable free energy simulations. *Biophysical Journal* **2018**, *114*, 1091–1102.

(36) Mason, P. E.; Jungwirth, P.; Duboué-Dijon, E. Quantifying the strength of a salt bridge by neutron scattering and molecular dynamics. *The journal of physical chemistry letters* **2019**, *10*, 3254–3259.

(37) Matsuura, Y.; Joti, Y.; Bagautdinov, B.; Yutani, K. Evaluating the strengths of salt bridges in the CutA1 protein using molecular dynamic simulations: a comparison of different force fields. *FEBS Open bio* **2019**, *9*, 1939–1956.

(38) Galvanetto, N.; Ivanović, M. T.; Del Grosso, S. A.; Chowdhury, A.; Sottini, A.; Nettels, D.; Best, R. B.; Schuler, B. Material properties of biomolecular condensates emerge from nanoscale dynamics. *Proceedings of the National Academy of Sciences* **2025**, *122*, e2424135122.

(39) Das, R. K.; Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences* **2013**, *110*, 13392–13397.

(40) Sawle, L.; Ghosh, K. A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *The Journal of chemical physics* **2015**, *143*.

(41) Kofinger, J.; Stelzl, L. S.; Reuter, K.; Allande, C.; Reichel, K.; Hummer, G. Efficient ensemble refinement by reweighting. *Journal of chemical theory and computation* **2019**, *15*, 3390–3401.

(42) Crehuet, R.; Buigues, P. J.; Salvatella, X.; Lindorff-Larsen, K. Bayesian-maximum-entropy reweighting of IDP ensembles based on NMR chemical shifts. *Entropy* **2019**, *21*, 898.

(43) Nüesch, M. F.; Pietrek, L.; Holmstrom, E. D.; Nettels, D.; von Roten, V.; Kronenberg-Tenga, R.; Medalia, O.; Hummer, G.; Schuler, B. Nanosecond chain dynamics of single-stranded nucleic acids. *Nat. Commun.* **2024**, *15*, 6010.

(44) Leone, V.; Marinelli, F. From snapshots to ensembles: Integrating experimental data and dynamics. *Current Opinion in Structural Biology* **2025**, *95*, 103155.

(45) Sottini, A.; Borgia, A.; Borgia, M. B.; Bugge, K.; Nettels, D.; Chowdhury, A.; Heidarsson, P. O.; Zosel, F.; Best, R. B.; Kragelund, B. B.; Schuler, B. Polyelectrolyte interactions enable rapid association and dissociation in high-affinity disordered protein complexes. *Nat. Commun.* **2020**, *11*, 5736.

(46) Heidarsson, P. O.; Mercadante, D.; Sottini, A.; Nettels, D.; Borgia, M. B.; Borgia, A.; Kilic, S.; Fierz, B.; Best, R. B.; Schuler, B. Release of linker histone from the nucleosome driven by polyelectrolyte competition with a disordered protein. *Nat. Chem.* **2022**, *14*, 224–231.

(47) Chowdhury, A.; Nettels, D.; Schuler, B. Interaction Dynamics of Intrinsically Disordered Proteins from Single-Molecule Spectroscopy. *Annu. Rev. Biophys.* **2023**, *52*, 433–462.

(48) Ansari, A.; Jones, C. M.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. The role of solvent viscosity in the dynamics of protein conformational changes. *Science* **1992**, *256*, 1796–8.

(49) Hagen, S. J.; Qiu, L. L.; Pabit, S. A. Diffusional limits to the speed of protein folding: fact or friction? *J. Phys. Cond. Mat.* **2005**, *17*, S1503–S1514.

(50) Soranno, A.; Buchli, B.; Nettels, D.; Cheng, R. R.; Müller-Späth, S.; Pfeil, S. H.; Hoffmann, A.; Lipman, E. A.; Makarov, D. E.; Schuler, B. Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proceedings of the National Academy of Sciences* **2012**, *109*, 17800–17806.

(51) Borgia, A.; Wensley, B. G.; Soranno, A.; Nettels, D.; Borgia, M. B.; Hoffmann, A.; Pfeil, S. H.; Lipman, E. A.; Clarke, J.; Schuler, B. Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. *Nature communications* **2012**, *3*, 1195.

(52) Zheng, W.; Hofmann, H.; Schuler, B.; Best, R. B. Origin of internal friction in disordered

proteins depends on solvent quality. *The Journal of Physical Chemistry B* **2018**, *122*, 11478–11487.

(53) Soranno, A.; Holla, A.; Dingfelder, F.; Nettels, D.; Makarov, D. E.; Schuler, B. Integrated view of internal friction in unfolded proteins from single-molecule FRET, contact quenching, theory, and simulations. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E1833–E1839.

(54) Hummer, G. Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New Journal of Physics* **2005**, *7*, 34.

(55) Zerze, G. H.; Mittal, J.; Best, R. B. Diffusive dynamics of contact formation in disordered polypeptides. *Physical review letters* **2016**, *116*, 068102.

(56) Gopich, I. V.; Szabo, A. Theory of the energy transfer efficiency and fluorescence lifetime distribution in single-molecule FRET. *Proceedings of the National Academy of Sciences* **2012**, *109*, 7747–7752.

(57) Doi, M.; Edwards, S. F. *The Theory of Polymer Dynamics*; Oxford University Press, USA: New York, 1988.

(58) Gopich, I. V.; Nettels, D.; Schuler, B.; Szabo, A. Protein dynamics from single-molecule fluorescence intensity correlation functions. *J. Chem. Phys.* **2009**, *131*, 095102.

(59) Best, R. B.; Zheng, W.; Borgia, A.; Buholzer, K.; Borgia, M. B.; Hofmann, H.; Soranno, A.; Nettels, D.; Gast, K.; Grishaev, A.; others Comment on "Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water". *Science* **2018**, *361*, eaar7101.

(60) Nüesch, M. F.; Ivanovic, M. T.; Claude, J. B.; Nettels, D.; Best, R. B.; Wenger, J.;

Schuler, B. Single-molecule Detection of Ultrafast Biomolecular Dynamics with Nanophotonics. *J. Am. Chem. Soc.* **2022**, *144*, 52–56.

(61) Schuler, B.; Soranno, A.; Hofmann, H.; Nettels, D. Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. *Annual Review of Biophysics* **2016**, *45*, 207–231.

(62) Chowdhury, A.; Nettels, D.; Schuler, B. Interaction dynamics of intrinsically disordered proteins from single-molecule spectroscopy. *Annual review of biophysics* **2023**, *52*, 433–462.

(63) Shea, J.-E.; Best, R. B.; Mittal, J. Physics-based computational and theoretical approaches to intrinsically disordered proteins. *Current opinion in structural biology* **2021**, *67*, 219–225.

(64) Best, R. B.; Buchete, N.-V.; Hummer, G. Are current molecular dynamics force fields too helical? *Biophysical journal* **2008**, *95*, L07–L09.

(65) Nettels, D.; Müller-Späth, S.; Küster, F.; Hofmann, H.; Haenni, D.; Rüegger, S.; Reymond, L.; Hoffmann, A.; Kubelka, J.; Heinz, B.; others Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proceedings of the National Academy of Sciences* **2009**, *106*, 20740–20745.

(66) Best, R. B.; Hummer, G. Optimized molecular dynamics force fields applied to the helix- coil transition of polypeptides. *The journal of physical chemistry B* **2009**, *113*, 9004–9015.

(67) Piana, S.; Robustelli, P.; Tan, D.; Chen, S.; Shaw, D. E. Development of a force field for the simulation of single-chain proteins and protein–protein complexes. *Journal of chemical theory and computation* **2020**, *16*, 2494–2507.

(68) Case, D. A.; Darden, T.; Cheatham III, T. E.; Simmerling, C.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; others AMBER 9. *University of California, San Francisco* **2006**, *45*, 719.

(69) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

(70) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of chemical physics* **2007**, *126*.

(71) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **1981**, *52*, 7182–7190.

(72) Darden, T.; York, D.; Pedersen, L.; others Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *Journal of chemical physics* **1993**, *98*, 10089–10089.

(73) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: A linear constraint solver for molecular simulations. *Journal of computational chemistry* **1997**, *18*, 1463–1472.

(74) Nuesch, M. F.; Ivanovic, M. T.; Claude, J.-B.; Nettels, D.; Best, R. B.; Wenger, J.; Schuler, B. Single-molecule detection of ultrafast biomolecular dynamics with nanophotonics. *Journal of the American Chemical Society* **2021**, *144*, 52–56.

(75) Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of the National Academy of Sciences* **2013**, *110*, 17874–17879.

(76) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics and modelling* **2006**, *25*, 247–260.

(77) Graen, T.; Hoefling, M.; Grubmuller, H. AMBER-DYES: characterization of charge fluctuations and force field parameterization of fluorescent dyes for molecular dynamics simulations. *J. Chem. Theor. Comput.* **2014**, *10*, 5505–5512.

(78) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **1995**, *117*, 5179–5197.

(79) Wang, D.; Frechette, L. B.; Best, R. B. On the role of native contact cooperativity in protein folding. *Proceedings of the National Academy of Sciences* **2024**, *121*, e2319249121.

(80) Hummer, G.; Köfinger, J. Bayesian ensemble refinement by replica simulations and reweighting. *The Journal of chemical physics* **2015**, *143*.

(81) Kalinin, S.; Valeri, A.; Antonik, M.; Felekyan, S.; Seidel, C. A. Detection of structural dynamics by FRET: a photon distribution and fluorescence lifetime analysis of systems with multiple states. *The Journal of Physical Chemistry B* **2010**, *114*, 7983–7995.

(82) Barth, A.; Opanasyuk, O.; Peulen, T.-O.; Felekyan, S.; Kalinin, S.; Sanabria, H.; Seidel, C. A. Unraveling multi-state molecular dynamics in single-molecule FRET experiments. I. Theory of FRET-lines. *The Journal of Chemical Physics* **2022**, *156*.

(83) Svergun, D.; Barberato, C.; Koch, M. H. CRYSOL–a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Applied Crystallography* **1995**, *28*, 768–773.

(84) Chen, P.-c.; Hub, J. S. Structural properties of protein–detergent complexes from SAXS and MD simulations. *The Journal of Physical Chemistry Letters* **2015**, *6*, 5116–5121.

(85) Knight, C. J.; Hub, J. S. WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. *Nucleic acids research* **2015**, *43*, W225–W230.

(86) Zheng, W.; Borgia, A.; Buholzer, K.; Grishaev, A.; Schuler, B.; Best, R. B. Probing the action of chemical denaturant on an intrinsically disordered protein by simulation and experiment. *Journal of the American Chemical Society* **2016**, *138*, 11702–11713.

(87) Bicout, D.; Szabo, A. Electron transfer reaction dynamics in non-Debye solvents. *The Journal of chemical physics* **1998**, *109*, 2325–2338.

(88) Koenig, I.; Zarrine-Afsar, A.; Aznauryan, M.; Soranno, A.; Wunderlich, B.; Dingfelder, F.; Stüber, J. C.; Plückthun, A.; Nettels, D.; Schuler, B. Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. *Nat. Methods* **2015**, *12*, 773–779.