Article

Modest Influence of FRET Chromophores on the Properties of Unfolded Proteins

Gül H. Zerze,[1] Robert B. Best,[2] and Jeetain Mittal[1,*]
[1]Department of Chemical and Biomolecular Engineering, Lehigh University, Bethlehem, Pennsylvania; and [2]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland

ABSTRACT Single-molecule Förster resonance energy transfer (FRET) experiments are often used to study the properties of unfolded and intrinsically disordered proteins. Because of their large extinction coefficients and quantum yields, synthetic heteroaromatic chromophores covalently linked to the protein are often used as donor and acceptor fluorophores. A key issue in the interpretation of such experiments is the extent to which the properties of the unfolded chain may be affected by the presence of these chromophores. In this article, we investigate this question using all-atom explicit solvent replica exchange molecular dynamics simulations of three different unfolded or intrinsically disordered proteins. We find that the secondary structure and long-range contacts are largely the same in the presence or absence of the fluorophores, and that the dimensions of the chain with and without chromophores are similar. This suggests that, at least in the cases studied, extrinsic fluorophores have little effect on the structural properties of unfolded or disordered proteins. We also find that the critical FRET orientational factor κ^2, has an average value and equilibrium distribution very close to that expected for isotropic orientations, which supports one of the assumptions frequently made when interpreting FRET efficiency in terms of distances.

INTRODUCTION

Förster resonance energy transfer (FRET) (1,2) is a biophysical technique that can yield information on distance distributions and dynamics in unfolded proteins (3–9), including inside cells (10). Single-molecule FRET can even be used to investigate separately the dynamics for subpopulations of a sample, for example the unfolded and folded states of a protein (11,12). The principle behind the method is that the efficiency of nonradiative energy transfer between two fluorophores is dependent primarily on their separation in space (although orientational factors may need to be considered in certain cases, as we discuss below) (1,2,13). FRET and small-angle x-ray scattering (SAXS) studies have produced divergent views of the dimensions of the unfolded states of proteins in low denaturant (5,6,14–18). There are many possible explanations for this inconsistency, which is still unresolved, but one concern is that the inclusion of large organic chromophores conjugated to the protein may somehow interfere with the properties of the protein—for example, by stabilizing a collapsed state of the chain. The controversy has largely centered around protein L (5,6,14,18), which has been studied by both methods; however, qualitatively similar results have been obtained for the unfolded states of other proteins by either FRET (collapse as denaturant is diluted) (19) or SAXS (no collapse) (20).

A commonly used pair of fluorophores, due to their stability and favorable spectroscopic properties, consists of AlexaFluor 488 as "donor" and AlexaFluor 594 as "acceptor". These dyes have been shown to yield self-consistent information on distances in model systems (21,22). The effect of the dyes also appears to have at most a limited effect on protein stability, where this has been checked (23). However, the potential effects of these extrinsic fluorophores on the intrinsic properties of unfolded and intrinsically disordered proteins still needs to be addressed, something which has not been systematically done to date. Obtaining this information from experiment, while certainly possible, is challenging. Molecular simulations offer one avenue to compare the properties of the unfolded states of proteins with and without FRET chromophores. In this article, we use molecular simulations with a state-of-the-art protein force field and water model to address the effects of dyes on the formation of long-range contacts, distance distributions, and per-residue secondary structure propensities of the unfolded state of three different proteins. We find that the dyes have little effect on the structure formation and dimensions of the unfolded state, and that one of the key assumptions underlying quantitative distance calculations from FRET measurements is justified by our results.

METHODS

We use the AMBER ff03w force field (24) for the protein model and TIP4P/2005 as the water model (25) to perform simulations using GROMACS 4.5.3 (26). Parameters for the dyes were derived using the AnteChamber feature of AmberTools (27), with the charges derived from the electrostatic potential from an HF/6-31+G* calculation with the Gaussian program (28). Full parameters are available upon request from the authors. Temperature replica exchange was run with temperatures ranging from 285 to 493 K

Submitted October 10, 2013, and accepted for publication July 11, 2014.

*Correspondence: jeetain@lehigh.edu

Editor: Ashok Deniz.

0006-3495/14/10/1654/7 $2.00

http://dx.doi.org/10.1016/j.bpj.2014.07.071

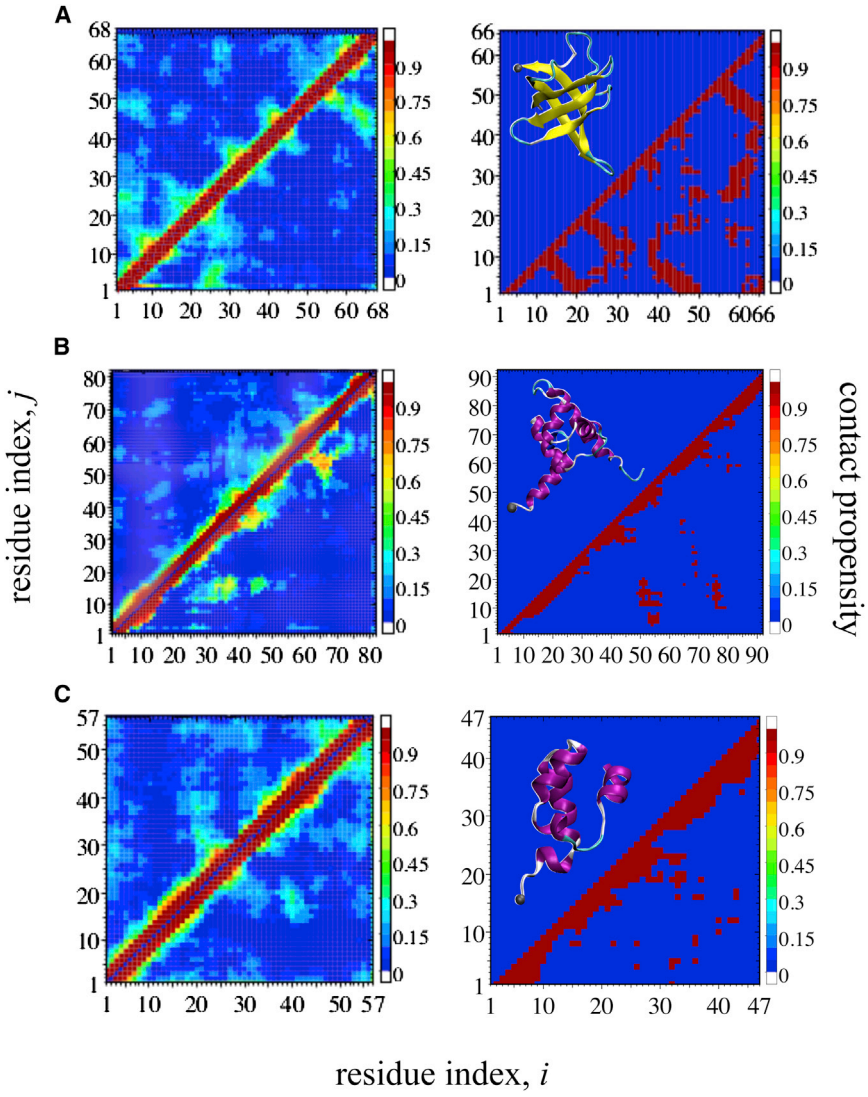

FIGURE 1 Contact maps of ensembles without (*upper left triangle*) and with (*lower right triangle*) extrinsic chromophores for three proteins (*left column*) along with their native contacts (*right column*). (*A*) CSP, (*B*) LR, and (*C*) IN. Folded structures of each protein are shown in the left corner of native contacts maps; (*yellow* and *magenta*) secondary structures; and (*gray sphere*) C-terminal α-carbon atom of each protein. For native contact maps and structures, the Protein Data Bank entries PDB:1G6P (41), PDB:1LMB (42), and PDB:1WJA (43) were used for CSP, LR, and IN, respectively. To see this figure in color, go online.

at a constant pressure of 1 bar, maintained using isotropic Parrinello-Rahman pressure coupling (29). The trajectory is propagated using Langevin dynamics with a 1 ps^{-1} friction coefficient. Electrostatic interactions are calculated using the particle-mesh Ewald method (30) with a real space cutoff of 0.9 nm. For van der Waals interactions, a 1.2-nm cutoff is used. A key feature of these simulations is the optimized protein model AMBER ff03w, which results in an improved reproduction of the properties of unfolded proteins relative to earlier force fields in conjunction with TIP3P water, as has been demonstrated in previous studies (31–33). Sizes of the systems have been shown in Table S2 in the Supporting Material by the number of water molecules present in simulation boxes. Initial unfolded configurations were generated from initial high temperature unfolding simulations, and were solvated in their simulation boxes. For such large systems, temperature replica exchange, which provides much more enhanced sampling of phase space compared to molecular dynamics, is not a common method because of its computational cost. In this work, that large system of each peptide has been simulated for 200 ns per replica using 56 replicas.

RESULTS AND DISCUSSION

To generalize how FRET dyes might affect the properties of unfolded polypeptides, we employ three different proteins with a variety of hydrophobic and electrostatic characteristics: namely the cold shock protein (CSP) from *Thermotoga maritima*, the DNA-binding domain of λ-repressor (LR), and the N-terminal domain of HIV integrase (IN) (7,8). CSP and LR are, respectively, 66-residue and 82-residue globular proteins, while IN is a 57 residue disordered protein, which folds in the presence of Zn^{2+} ions. The three proteins have quite different sequence and structural characteristics. CSP is an all-β fold, whereas LR and IN have all-α folds; the average hydrophobicities of the sequences, based on the Kyte-Doolittle score (34), are -0.55 for CSP, -0.25 for LR, and -0.49 for IN. The fraction of charged residues is 0.36, 0.26, and 0.32, respectively, for CSP, LR, and IN. In particular, both CSP and IN have been shown to collapse when denaturant is diluted (35) (although for IN there is a small reexpansion at the lowest concentrations of guanidinium chloride, due to the reduction in ionic screening).

In each case, we consider both the free protein and a variant with the FRET chromophores ligated to cysteine

residues added at or near the termini, with the N- and C-terminal cysteines respectively labeled with the chromophores AlexaFluor 488 and AlexaFluor 594. The labeled forms of each protein are denoted “CSP dyes”, “LR dyes”, and “IN dyes”. The full sequences of each system are shown in Table S1. Unfolded proteins are notoriously difficult to treat with standard force fields and water models, as evident by analysis of unfolded states in the atomistic folding simulations by Lindorff-Larsen et al. (36), Best et al. (37), and Piana et al. (38), as well as earlier comparisons with FRET data on unfolded proteins (8). To address this issue, we have adopted the most accurate four-site water model for describing the properties of liquid water (25), essential for accurately capturing the hydrophobic effect (39), and we have tuned a protein force field, AMBER ff03 (40), for compatibility with this model (24). Although not perfect, this combination has resulted in improved properties in simulations of disordered and unfolded proteins (31–33).

To compare the properties of all three proteins (Protein Data Bank (PDB) IDs are 1G6P (41) for CSP, 1LMB (42) for LR and 1WJA (43) for IN) with and without dyes, we have performed replica-exchange molecular-dynamics simulations of each system in explicit solvent. We focus on the 300 K replica, for which we use 150 ns of data, after neglecting an initial equilibration period of 50 ns. As a measure of global structure formation, we have computed contact maps (Fig. 1), in which the entry for each pair of residues (i,j) represents the fraction of time that at least one heavy atom of backbone from i and one from j are within 0.8 nm of each other, averaged over 150 ns.

Residue-residue contact formation for the simulation ensembles with and without dyes are shown in Fig. 1 (*left column*), which indicates that the trajectories with and without dyes are sampling quite similar conformations. Additionally, all contact maps reveal a general absence of long-range contacts, reflecting a disordered chain. The structure formed in the unfolded state essentially reflects nativelike contacts with small sequence separation (Fig. 1, *right column*).

To assess local structure formation, we have computed the average residual secondary structure formation over the trajectory. In Fig. 2 we show α-helix and β-sheet secondary structure assignments obtained using DSSP definitions (45). The remaining DSSP-assigned structures (β-bridge, bend, turn, 3_{10}-helix, 5-helix) are given in Fig. S1, Fig. S2, and Fig. S3 in the Supporting Material. This shows that both CSP and CSP dyes are sampling mostly coil conformations (Fig. 2). The per-residue populations of secondary structure are similar for both chains. One apparent exception is in the region spanning residues 64–67, giving higher propensity to be in β-sheet for the chain with dyes, which may reflect a small effect of the dye at the C-terminus. For α-helix, the populations with and without dyes are very similar, considering the error range, except for the region spanning residues 24–30, where α-helix propensity is higher for naturally occurring CSP than for CSP dyes. For LR, a difference appears for α-helical propensities in residues 59–65, and residues 6–11 near the N-terminus. These regions might be affected by the presence of the dyes; however, they are still small regions when compared to the total length of sequence. For IN, the only clear difference appears in helical propensity for a limited region, but despite the difference in numbers, both IN and IN dyes keep same trend for that region as well. Overall, the dyes at both termini have very little effect on the secondary

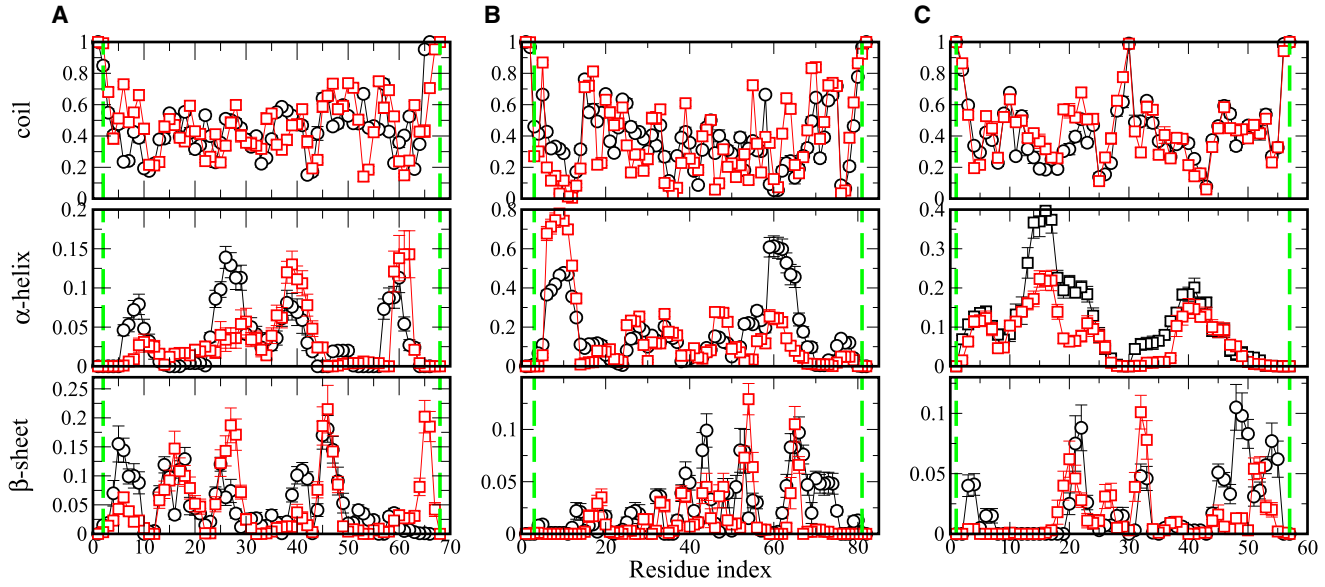

FIGURE 2 Per-residue secondary structure propensities calculated based on the DSSP definition. (*A*) CSP, (*B*) LR, (*C*) IN. (*Black curves*) Averages of ensembles without dyes; (*red curves*) averages of ensembles in presence of dyes. Standard errors are calculated using block averaging (44). (*Vertical green lines*) Residues to which the dyes are attached. To see this figure in color, go online.

structures populated, when viewed at a residue level. Although our replica-exchange molecular-dynamics sampling provides a representative ensemble for each system, the sampling is nonetheless limited and so small differences between ensembles should not be overinterpreted.

Because FRET primarily reports on interresidue distances and dimensions of the chain, we have also computed distributions of the radius of gyration (R_g) (Fig. 3, *top row*) and end-to-end (E2E) distance (Fig. 3, *second* and *third rows*), and dye-to-dye (D2D) distances (Fig. 3, *fourth row*) from the simulations (19,46). E2E distance distributions are calculated from the distances between the backbone nitrogen at the N-terminus to the backbone oxygen at the C-terminus. D2D distances are calculated from the distances between the centers of masses of the dyes in the simulations of proteins with dyes. Average values of R_g and E2E distance both with and without dyes as well as average values of D2D distance obtained from our data are summarized in Table 1. To have a clearer comparison of the results with and without dyes, each E2E and D2D distance distribution is fitted to an ideal (Gaussian) chain model (*red curve* in Fig. 3) and self-avoiding random walk model (47) (*green curve* in Fig. 3).

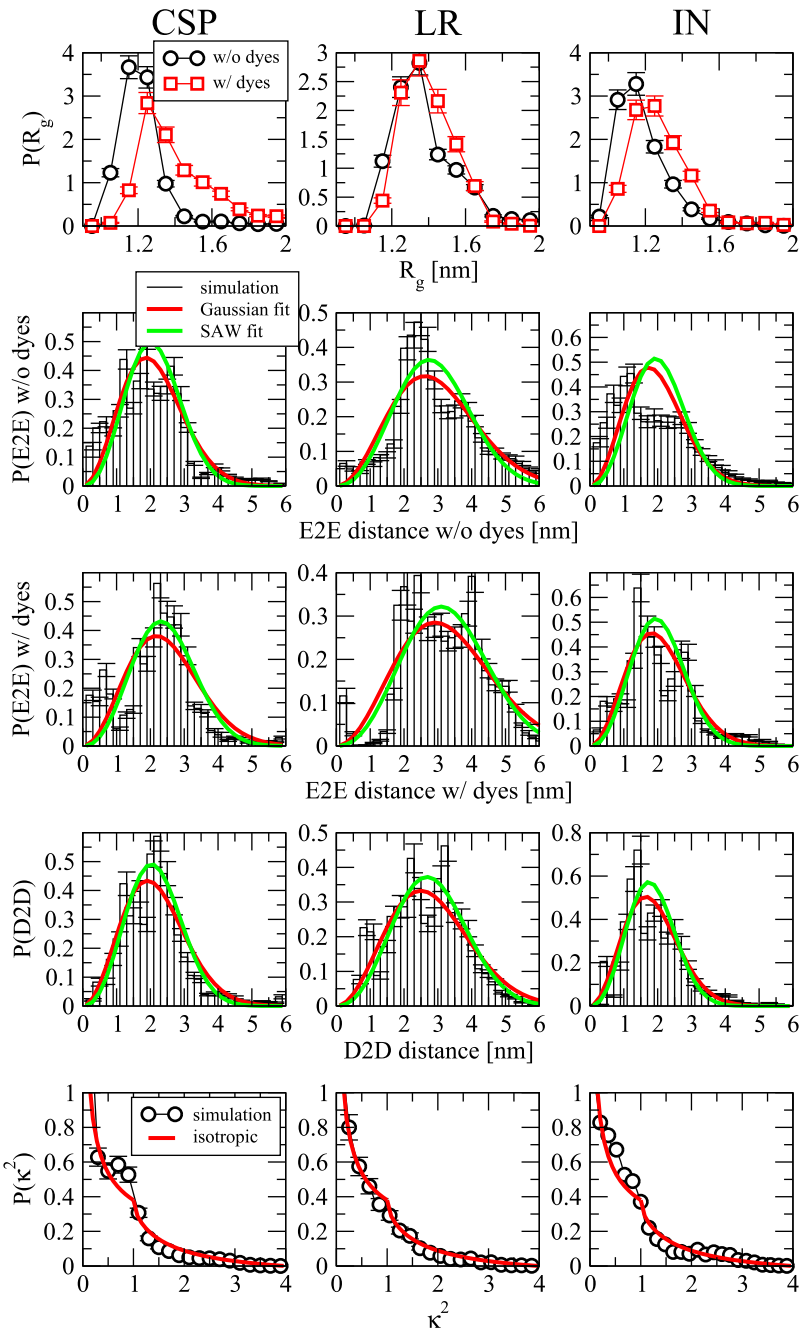

FIGURE 3 Distance distributions of proteins with and without dyes and orientational factor distribution of dyes. (From *top* to *bottom*) Radius of gyration (R_g) with and without dyes, End-to-end (E2E) distance without dyes, E2E distance with dyes, dye-to-dye (D2D) distance, and orientational factor κ^2 distributions of our data and the isotropic case. (*Red curves*) E2E and D2D distance distributions are Gaussian chain fits; (*green curves*) self-avoiding chain fits. To see this figure in color, go online.

TABLE 1 Average values of radius of gyration (R_g), end-to-end (E2E), and dye-to-dye (D2D) distances and orientational factor (κ^2) from simulations indicated in nanometers

Ensembles	$\langle R_g \rangle$ without dyes	$\langle R_g \rangle$ with dyes	$\langle E2E \rangle$ without dyes	$\langle E2E \rangle$ with dyes	$\langle D2D \rangle$	κ^2
CSP	1.23 ± 0.01	1.42 ± 0.02	2.06 ± 0.03	2.35 ± 0.09	2.34 ± 0.15	0.61 ± 0.05
LR	1.41 ± 0.04	1.40 ± 0.01	3.04 ± 0.08	3.24 ± 0.08	2.72 ± 0.09	0.58 ± 0.04
IN	1.19 ± 0.01	1.27 ± 0.01	1.19 ± 0.01	2.09 ± 0.09	1.93 ± 0.09	0.68 ± 0.03

The E2E distance distribution $P(x)$ for an ideal chain is given by the following equation,

$$P(x) = 4\pi x^2 \left(\frac{3}{2\pi N l_k^2}\right)^{\frac{3}{2}} \exp\left(-\frac{3}{2}\frac{x^2}{N l_k^2}\right), \qquad (1)$$

in which N is the number of residues and l_k is the segment or Kuhn length. The only free parameter in the ideal chain models is l_k, for which the fits yield very close numbers for E2E without and with dyes and D2D distributions for all proteins. All yielded fit parameters are summarized in Table 2. These segment lengths are slightly smaller than the α-carbon-to-α-carbon distance, which is always close to 0.38 nm; because the Cα-Cα-Cα angles are relatively inflexible, however, the shorter distances may reflect a projection of the Cα-Cα distance onto the local chain direction. The absolute values of l_k are comparable to, although slightly smaller than, those estimated from single molecule fluorescence experiments (35). Independent of the value of the parameters, their similarity indicates that the properties of the chain are relatively unaffected by the chromophores. Particularly, D2D distribution fittings reflect almost the same distribution as E2E distance without dyes. A more accurate representation of the chain should be given by a self-avoiding random walk (SAW). The E2E distance distribution for a SAW is given by

$$P(x) = \frac{a}{\langle x \rangle}\left(\frac{x}{\langle x \rangle}\right)^{2+\theta} \exp\left[-b\left(\frac{x}{\langle x \rangle}\right)^{\delta}\right], \qquad (2)$$

where θ, δ, a, and b are numerical parameters equal to 0.3, 2.5, 3.67853, and 1.23152, respectively (47). Thus, the average values of the end-to-end distances ($\langle x \rangle$) are the only fit parameters that are summarized in Table 1. Similar to Gaussian chain fitting, self-avoiding random walk fits show a close similarity between different distributions of proteins with and without dyes. Additionally, $\langle E2E \rangle$ distance values from fits (Table 2) and our data (Table 1) are not particularly different from each other, except for E2E distance distribution of IN.

Although the distributions of E2E distance are challenging to sample, because thorough exploration of essentially all internal degrees of freedom is required, sampling the R_g is easier because of a higher degree of self-averaging. The R_g distributions of all proteins with and without dyes are overall quite similar (Fig. 3, *top row*) with the average R_g value slightly higher in the presence of dyes for CSP and IN. The tendency of proteins with dyes to sample higher R_g values can be rationalized at the simplest level by the excluded volume effect of the dye molecules. For the case of CSP, the difference is a little more pronounced, most likely because of the difference in the total number of residues for CSP and CSP dyes. Note, however, that this would not help to explain the discrepancy between FRET and SAXS results, inasmuch as R_g estimates from FRET (with dyes) are generally smaller at low denaturant concentrations than those obtained by SAXS (without dyes). However, it may partially explain the fact that R_g computed from simulations is often smaller than that estimated from FRET.

To determine average distances from FRET efficiencies, one in principle has to consider the mutual orientation of the two chromophores, and not just their distance. The orientational contribution to the transfer efficiency is captured by the factor κ^2 in which κ is given by

$$\kappa = \hat{\mu}_D \cdot \hat{\mu}_A - 3(\hat{r} \cdot \hat{\mu}_A)(\hat{r} \cdot \hat{\mu}_D),$$

(22) where $\hat{r}$ is the unit vector between donor and acceptor and $\hat{\mu}_A$ and $\hat{\mu}_D$ are the unit vectors along the donor and acceptor transition dipoles, which are shown in Fig. S4. In an ideal situation in which the dyes are reorienting quickly with respect to each other and adopting an approximately isotropic distribution of relative orientations, an average κ^2 value of 2/3 can be assumed. If this assumption was invalid, it would complicate the interpretation of FRET efficiencies in terms of distances, even for qualitative applications. In experiment, anisotropy measurements can be used to

TABLE 2 Fit parameters l_k (nm) for Gaussian fitting and $\langle E2E \rangle$ (nm) for SAW fitting

Ensembles	Gaussian fitting l_k (E2E without dyes)	Gaussian fitting l_k (E2E with dyes)	Gaussian fitting l_k (D2D)	SAW fitting $\langle E2E \rangle$ without dyes	SAW fitting $\langle E2E \rangle$ with dyes	SAW fitting $\langle D2D \rangle$
CSP	0.28(0.92)	0.33(0.82)	0.29(0.93)	2.26(0.91)	2.60(0.86)	2.28(0.95)
LR	0.36(0.91)	0.40(0.90)	0.34(0.92)	3.08(0.92)	3.48(0.91)	3.01(0.92)
IN	0.28(0.80)	0.30(0.92)	0.27(0.93)	2.18(0.80)	2.18(0.91)	1.96(0.92)

Values indicated in parentheses are correlation coefficients of fits.

support $\kappa^2 \sim 2/3$, but the distribution cannot be measured directly in real samples.

As we run replica exchange simulations, we do not have direct access to the timescales of reorientation, but we can nonetheless compute the equilibrium average, and distribution of κ^2 from the simulations for comparison with their ideal isotropic counterparts (48) in addition to distance distributions. We find that the average values of κ^2, which are tabulated in Table 1, are within 1.5 standard deviations of the isotropic value (2/3). Comparison with the isotropic distribution (Fig. 3, *bottom row*) shows a close similarity between the isotropic and simulation distributions. Thus, at least at an equilibrium level, it seems that the interdye κ^2 factor may safely be assumed to be 2/3; however, the dynamics of the dyes would also ultimately have to be considered (i.e., does κ^2 average to 2/3 within the fluorescence lifetime?).

In conclusion, we have compared the unfolded ensembles of naturally occurring CSP, LR, and IN and a CSP, LR, and IN with AlexaFluor 488 and AlexaFluor 594 fluorophores attached to it, similar to that used in FRET experiments. We find that both ensembles with and without dyes yielded negligible long-range contacts for all three sequences, indicating that they are disordered, and that the distributions of dye-to-dye distance of sequences with dyes and end-to-end distance of sequences without dyes are remarkably similar, and that the dyes did not affect the average secondary structure sampled by each residue. The major effect of the chromophores was a very slight expansion of the chain. Overall, our findings that extrinsic chromophores have little effect on the chain properties (at least for sequences studied here), and that the average $\kappa^2 \sim 2/3$, lends confidence to the use of FRET chromophores to extract true dimensions for unfolded proteins. Even though we have studied a limited number of sequences, we believe results here can be generalized inasmuch as these three sequences accommodate a variety of different chain characteristics.

SUPPORTING MATERIAL

Parameterization of Chromophores, two tables, and four figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00847-9.

This work was supported by National Science Foundation grant No. CBET-1120399 (to J.M.). Use of the high-performance computing capabilities of the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation grant no. TG-MCB-120014, is also gratefully acknowledged. R.B.B was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health.

REFERENCES

1. Förster, T. 1948. Intermolecular energy migration and fluorescence [Zwischenmolekulare energiewanderung und fluoreszenz]. *Ann. Phys.* 2:55–75.
2. Förster, T. 1965. Delocalized excitation and excitation transfer. *In* Modern Quantum Chemistry: Istanbul Lectures. Part III, Action of Light and Organic Crystals. O. Sinanoglu, editor. Academic Press, New York.
3. Kuzmenkina, E. V., C. D. Heyes, and G. U. Nienhaus. 2005. Single-molecule Forster resonance energy transfer study of protein dynamics under denaturing conditions. *Proc. Natl. Acad. Sci. USA.* 102:15471–15476.
4. Kuzmenkina, E. V., C. D. Heyes, and G. U. Nienhaus. 2006. Single-molecule FRET study of denaturant induced unfolding of RNase H. *J. Mol. Biol.* 357:313–324.
5. Sherman, E., and G. Haran. 2006. Coil-globule transition in the denatured state of a small protein. *Proc. Natl. Acad. Sci. USA.* 103:11539–11543.
6. Merchant, K. A., R. B. Best, …, W. A. Eaton. 2007. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc. Natl. Acad. Sci. USA.* 104:1528–1533.
7. Nettels, D., A. Hoffmann, and B. Schuler. 2008. Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds. *J. Phys. Chem. B.* 112:6137–6146.
8. Nettels, D., S. Müller-Späth, …, B. Schuler. 2009. Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proc. Natl. Acad. Sci. USA.* 106:20740–20745.
9. Ziv, G., D. Thirumalai, and G. Haran. 2009. Collapse transition in proteins. *Phys. Chem. Chem. Phys.* 11:83–93.
10. Dhar, A., A. Samiotakis, …, M. S. Cheung. 2010. Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding. *Proc. Natl. Acad. Sci. USA.* 107:17586–17591.
11. Schuler, B., and H. Hofmann. 2013. Single-molecule spectroscopy of protein folding dynamics—expanding scope and timescales. *Curr. Opin. Struct. Biol.* 23:36–47.
12. Schuler, B., and W. A. Eaton. 2008. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* 18:16–26.
13. Scholes, G. D. 2003. Long-range resonance energy transfer in molecular systems. *Annu. Rev. Phys. Chem.* 54:57–87.
14. Plaxco, K. W., I. S. Millett, …, D. Baker. 1999. Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nat. Struct. Biol.* 6:554–556.
15. Scalley, M. L., S. Nauli, …, D. Baker. 1999. Structural transitions in the protein L denatured state ensemble. *Biochemistry.* 38:15927–15935.
16. Magg, C., and F. X. Schmid. 2004. Rapid collapse precedes the fast two-state folding of the cold shock protein. *J. Mol. Biol.* 335:1309–1323.
17. Magg, C., J. Kubelka, …, F. X. Schmid. 2006. Specificity of the initial collapse in the folding of the cold shock protein. *J. Mol. Biol.* 360:1067–1080.
18. Yoo, T. Y., S. P. Meisburger, …, K. Plaxco. 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.
19. Schuler, B., E. A. Lipman, and W. A. Eaton. 2002. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature.* 419:743–747.
20. Jacob, J., R. S. Dothager, …, T. R. Sosnick. 2007. Fully reduced ribonuclease A does not expand at high denaturant concentration or temperature. *J. Mol. Biol.* 367:609–615.
21. Schuler, B., E. A. Lipman, …, W. A. Eaton. 2005. Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proc. Natl. Acad. Sci. USA.* 102:2754–2759.
22. Best, R. B., K. A. Merchant, …, W. A. Eaton. 2007. Effect of flexibility and *cis* residues in single-molecule FRET studies of polyproline. *Proc. Natl. Acad. Sci. USA.* 104:18964–18969.

23. Chung, H. S., J. M. Louis, and W. A. Eaton. 2009. Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proc. Natl. Acad. Sci. USA.* 106:11837–11844.

24. Best, R. B., and J. Mittal. 2010. Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B.* 114:14916–14923.

25. Abascal, J. L., and C. Vega. 2005. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* 123:234505.

26. Hess, B., C. Kutzner, …, E. Lindahl. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–447.

27. Wang, J., R. M. Wolf, …, D. A. Case. 2004. Development and testing of a general amber force field. *J. Comput. Chem.* 25:1157–1174.

28. Frisch, M. J., G. W. Trucks, …, J. A. Pople. 2004. Gaussian 03, Rev. C.02. Gaussian, Inc., Wallingford, CT.

29. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52:7182–7190.

30. Essmann, U., L. Perera, …, L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.

31. Mittal, J., T. H. Yoo, …, T. M. Truskett. 2013. Structural ensemble of an intrinsically disordered polypeptide. *J. Phys. Chem. B.* 117: 118–124.

32. Knott, M., and R. B. Best. 2012. A preformed binding interface in the unbound ensemble of an intrinsically disordered protein: evidence from molecular simulations. *PLOS Comput. Biol.* 8:e1002605.

33. Miller, C., G. H. Zerze, and J. Mittal. 2013. Molecular simulations indicate marked differences in the structure of amylin mutants, correlated with known aggregation propensity. *J. Phys. Chem. B.* 117:16066–16075.

34. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105–132.

35. Müller-Späth, S., A. Soranno, …, B. Schuler. 2010. Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA.* 107:14609–14614.

36. Lindorff-Larsen, K., S. Piana, …, D. E. Shaw. 2011. How fast-folding proteins fold. *Science.* 334:517–520.

37. Best, R. B., G. Hummer, and W. A. Eaton. 2013. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. USA.* 110:17874–17879.

38. Piana, S., J. L. Klepeis, and D. E. Shaw. 2014. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struct. Biol.* 24:98–105.

39. Ashbaugh, H. S., N. J. Collett, …, J. A. Staton. 2010. Assessing the thermodynamic signatures of hydrophobic hydration for several common water models. *J. Chem. Phys.* 132:124504.

40. Duan, Y., C. Wu, …, P. Kollman. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24:1999–2012.

41. Kremer, W., B. Schuler, …, H. R. Kalbitzer. 2001. Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritima*. *Eur. J. Biochem.* 268:2527–2539.

42. Clarke, N. D., L. J. Beamer, …, C. O. Pabo. 1991. The DNA binding arm of λ-repressor: critical contacts from a flexible region. *Science.* 254:267–270.

43. Cai, M., R. Zheng, …, A. M. Gronenborn. 1997. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.* 4:567–577.

44. Flyvbjerg, H., and H. G. Petersen. 1989. Error estimates on averages of correlated data. *J. Chem. Phys.* 91:461–466.

45. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.

46. Stryer, L., and R. P. Haugland. 1967. Energy transfer: a spectroscopic ruler. *Proc. Natl. Acad. Sci. USA.* 58:719–726.

47. O'Brien, E. P., G. Morrison, …, D. Thirumalai. 2009. How accurate are polymer models in the analysis of Förster resonance energy transfer experiments on proteins? *J. Chem. Phys.* 130:124903.

48. Hoefling, M., N. Lima, …, H. Grubmüller. 2011. Structural heterogeneity and quantitative FRET efficiency distributions of polyprolines through a hybrid atomistic simulation and Monte Carlo approach. *PLoS ONE.* 6:e19791.

Supporting Information for "Modest influence of FRET chromophores on the properties of unfolded proteins"

Gül H. Zerze,[†] Robert B. Best,[‡] and Jeetain Mittal[*,†]

Department of Chemical Engineering, Lehigh University, Bethlehem, PA 18034, USA, and Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, MD 20892, USA

E-mail: jeetain@lehigh.edu

Parameterization of chromophores

Our priority in determining parameters was to obtain reasonable non-bonded interactions, as the most critical feature of the model is how the chromophores interact with each other and with the protein. We use standard Amber atom types for all of the atoms in the two chromophores, thus fixing the Lennard-Jones parameters; angle and torsion terms were added by analogy with similar terms in the AMBER force field. Charges were determined using the restrained electrostatic potential (RESP) fitting[1] as implemented in the Antechamber program. For consistency with AMBER charges, electrostatic potentials were determined with unrestricted Hartree-Fock, assuming all carboxylate groups to be deprotonated, i.e. net charge of -2 for each dye; the geometry of each molecule was first optimized with the same

[*]To whom correspondence should be addressed
[†]Lehigh University
[‡]National Institutes of Health

method. We used the 6-31+G* basis set in place of the standard 6-31G*, the diffuse basis functions being included due to the presence of the sulfur atoms in these molecules. We also found that using this basis set gave more reasonable results in the geometry optimizations. Gromacs parameters for the dyes are available upon request from the authors.

Table S1: Sequences of molecules simulated in this study. *Dyes are attached to cysteine residues given in red.

	Sequences
CSP	MRGKVKWFDS KKGYGFITKD EGGDVFVHWS AIEMEGFKTL KEGQVVEFEI QEGKKGPQAA HVKVVE
CSP Dyes*	MCRGKVKWFD SKKGYGFITK DEGGDVFVHW SAIEMEGFKT LKEGQVVEFE IQEGKKGPQA AHVKVVEC
LR	GPCLTQEQLE DARRLKAIYE KKKNELGLSQ ESVADKMGMG QSGVGALFNG INALNAYNAA LLAKILKVSV EEFSPSIARE CR
LR Dyes*	GPCLTQEQLE DARRLKAIYE KKKNELGLSQ ESVADKMGMG QSGVGALFNG INALNAYNAA LLAKILKVSV EEFSPSIARE CR
IN	CFLDGIDKAQ EEHEKYHSNW RAMASDFNLP PVVAKGIVAS CDKCQLKGEA MHGQVDC
IN Dyes*	CFLDGIDKAQ EEHEKYHSNW RAMASDFNLP PVVAKGIVAS CDKCQLKGEA MHGQVDC

Table S2: Number of non-protein molecules in simulated systems.

	CSP	**CSP Dyes**	**LR**	**LR Dyes**	**IN**	**IN Dyes**
Water	8339	8329	8294	8234	8422	8353
Sodium ions	10	10	9	11	12	16
Chloride ions	10	6	10	8	8	8

References

(1) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357–1377.

(2) Cross, S.; Kuttel, M. M.; Stone, J. E.; Gain, J. E. *J. Mol. Graph. Model.* **2009**, *28*, 131–139.

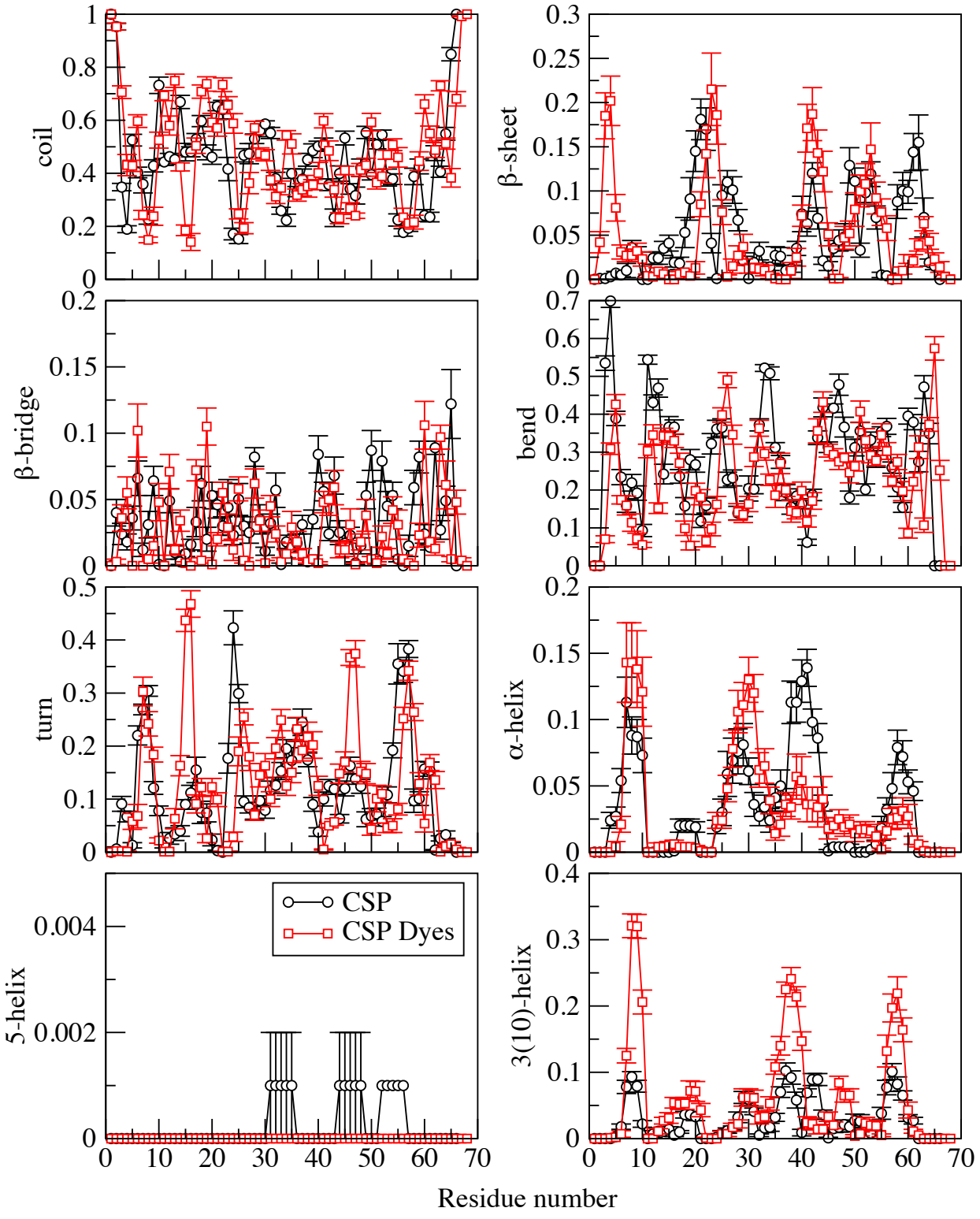

Figure S1: Per-residue secondary structure propensities calculated based on DSSP definition of CSP and CSP Dyes. Standard errors are calculated using block averaging.

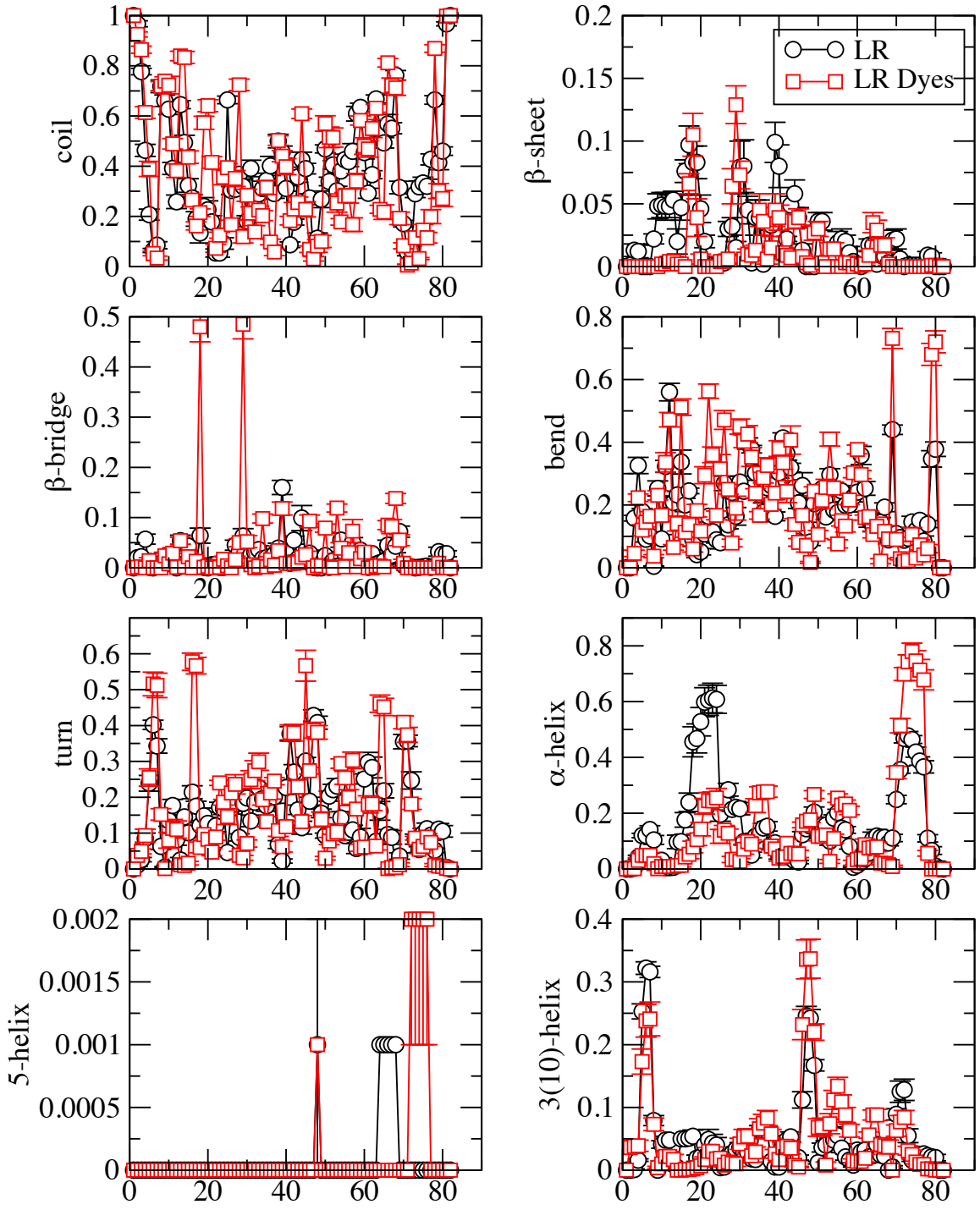

Figure S2: Per-residue secondary structure propensities calculated based on DSSP definition of LR and LR Dyes. Standard errors are calculated using block averaging.

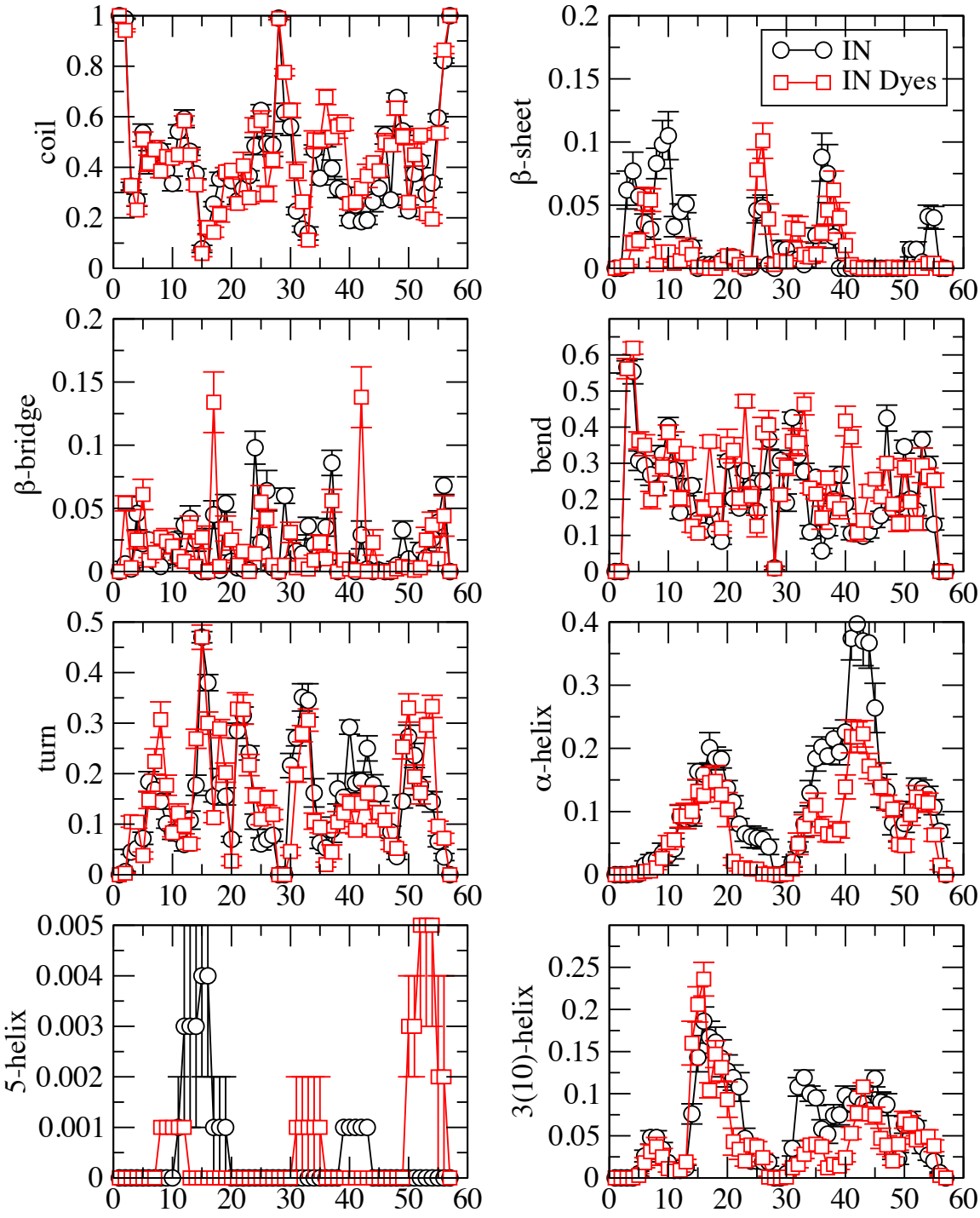

Figure S3: Per-residue secondary structure propensities calculated based on DSSP definition of IN and IN Dyes. Standard errors are calculated using block averaging.

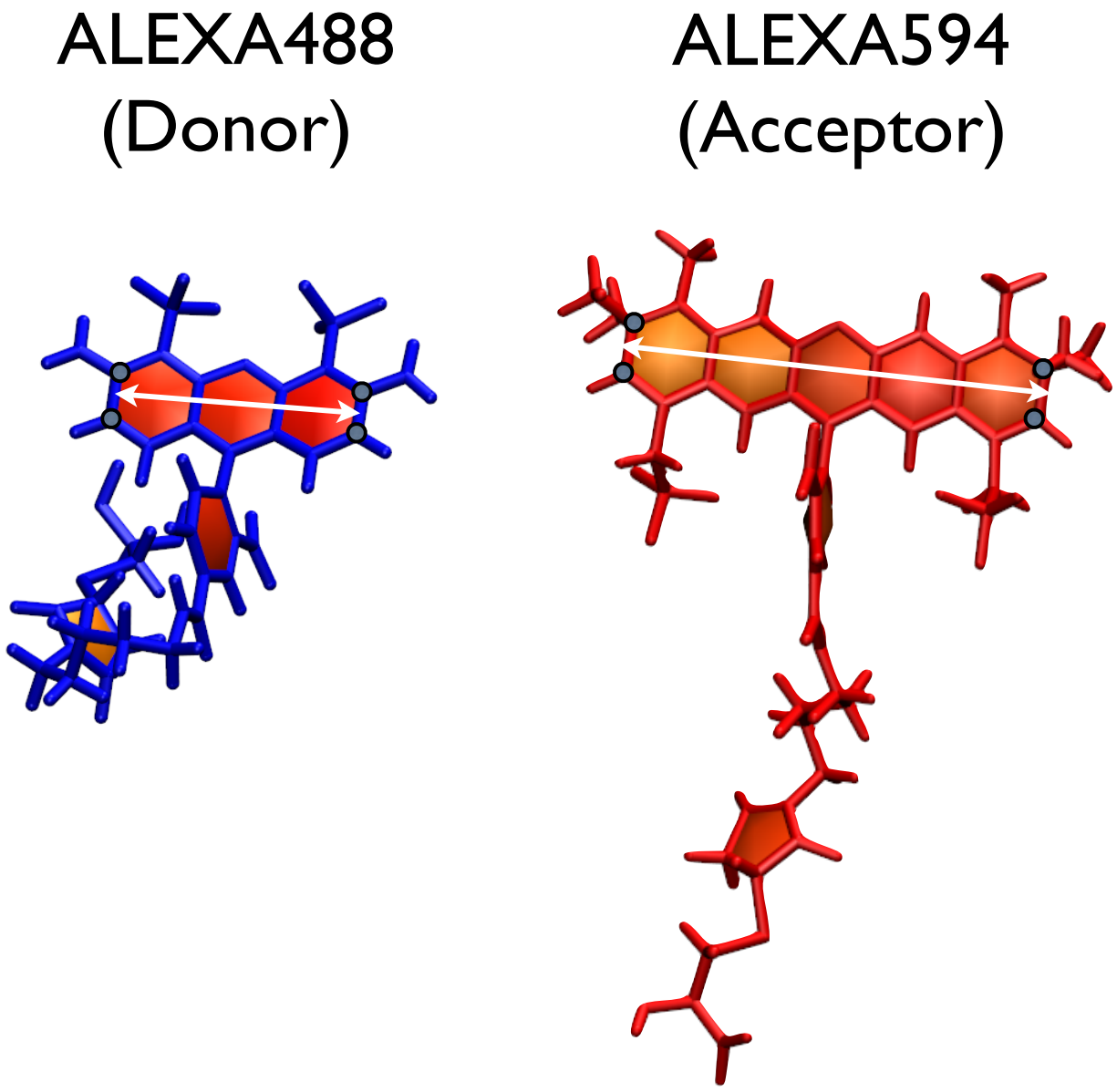

Figure S4: Blue molecule is the Alexa 488 attached at the N-terminus and red molecule is the Alexa 594 attached at the C-terminus. Both are shown with the paperchain representation in VMD.[2] Transition dipole vectors used for orientational factor κ^2 calculation are shown with white arrows.