An experimental study of GFP-based FRET, with application to intrinsically unstructured proteins.

Tomoo Ohashi, Stéphane D. Galiacy, Gina Briscoe, Harold P Erickson

An experimental study of GFP-based FRET, with application to intrinsically unstructured proteins

TOMOO OHASHI, STEPHANE D. GALIACY, GINA BRISCOE, AND HAROLD P. ERICKSON

Department of Cell Biology, Duke University Medical Center, Durham, North Carolina 27710, USA

Abstract

We have experimentally studied the fluorescence resonance energy transfer (FRET) between green fluorescent protein (GFP) molecules by inserting folded or intrinsically unstructured proteins between CyPet and Ypet. We discovered that most of the enhanced FRET signal previously reported for this pair was due to enhanced dimerization, so we engineered a monomerizing mutation into each. An insert containing a single fibronectin type III domain (3.7 nm end-to-end) gave a moderate FRET signal while a two-domain insert (7.0 nm) gave no FRET. We then tested unstructured proteins of various lengths, including the charged-plus-PQ domain of ZipA, the tail domain of α-adducin, and the C-terminal tail domain of FtsZ. The structures of these FRET constructs were also studied by electron microscopy and sedimentation. A 12 amino acid linker and the N-terminal 33 amino acids of the charged domain of the ZipA gave strong FRET signals. The C-terminal 33 amino acids of the PQ domain of the ZipA and several unstructured proteins with 66–68 amino acids gave moderate FRET signals. The 150 amino acid charged-plus-PQ construct gave a barely detectable FRET signal. FRET efficiency was calculated from the decreased donor emission to estimate the distance between donor and acceptor. The donor–acceptor distance varied for unstructured inserts of the same length, suggesting that they had variable stiffness (persistence length). We conclude that GFP-based FRET can be useful for studying intrinsically unstructured proteins, and we present a range of calibrated protein inserts to experimentally determine the distances that can be studied.

Keywords: GFP; FRET; unstructured proteins; worm-like chain; persistence length; CyPet; YPet

Fluorescence resonance energy transfer (FRET) has been developed as a powerful method for biochemical and biological studies. FRET can only occur when the donor and acceptor fluorophores are close to each other (in the range of from 1 to 10 nm, depending on the fluorophore combination). In addition to the donor–acceptor distance, the orientation of the two fluorophores influences the FRET signal (Wu and Brand 1994; Tsien 1998). Conformational changes within proteins or protein–protein interactions can be monitored in solution and in living cells. Recent advances in GFP (green fluorescent protein) technology have increased the usefulness of FRET, especially in vivo. For example, GFP-based FRET has been used as a proteolysis indicator (Heim and Tsien 1996; Mitra et al. 1996; Xu et al. 1998), a Ca^{2+} sensor (Miyawaki et al. 1997; Persechini et al. 1997), and a cAMP sensor (Zaccolo et al. 2000). Many GFP-based FRET studies have been published and novel GFP variants for FRET are still being developed (Nagai et al. 2002; Nguyen and Daugherty 2005; Ai et al. 2006).

The Förster distances, at which the transfer efficiency is 50%, have been estimated theoretically for various GFP-based FRET pairs (Patterson et al. 2000), but these

have only recently been tested experimentally by inserting proteins with a range of known sizes between the GFPs. Evers et al. (2006) designed various lengths of flexible peptide linkers to insert between two GFPs, and compared the observed FRET signal with a computer-modeled estimate.

Until recently, intrinsically unstructured proteins or domains have been considered to be merely flexible linkers. However, a number of reports have shown that unstructured proteins can have important biological functions when they interact with other proteins or nucleic acids (Dunker et al. 2002; Tompa 2002; Uversky 2002; Dyson and Wright 2005). Although unstructured proteins can be predicted from the amino acid sequence, it is difficult to study them experimentally. Protease sensitivity is a convenient probe. NMR and CD spectroscopy can also be used.

We previously developed a method for studying unstructured proteins by rotary shadowing electron microscopy (EM) (Li et al. 2001; Ohashi et al. 2002). Because unstructured proteins such as the charged-plus-PQ (proline-glutamine rich) domain of ZipA and the PEVK (proline-glutamate-valine-lysine rich) domain of titin were invisible in EM, we added small globular domains at both the N- and C-termini to measure the end-to-end distances of the unstructured segments between them. These end-to-end distances provide valuable information for understanding intrinsically unstructured proteins, in particular for estimating a persistence length of a corresponding worm-like chain. However, these EM measurements are most valuable for polypeptides longer than 150 amino acids.

In the present study, we have used a FRET-based assay to explore the end-to-end separation of presumed unstructured proteins shorter than 150 amino acids. In addition to the unstructured peptides, we have used several rigid inserts of known structure to provide calibration points. We will discuss the application of GFP-based FRET to the study of protein structure, especially its usefulness in estimating the end-to-end distance of unstructured proteins.

Results

We originally started this project using the FRET pair YFP$_{Venus}$–ECFP*, but when Nguyen and Daugherty (2005) reported the development of an improved FRET pair, YPet–CyPet, which had a sevenfold enhancement of the FRET signal, we decided to use this enhanced FRET pair for our project. However, an early application showed that a 66-amino acid adducin insert (AD66t1) gave a FRET signal almost identical to that with the 12-amino acid insert. We suspected that dimerization of the YPet–CyPet might be the problem, so we introduced the monomerizing mutation A206K (Zacharias et al. 2002), first in YPet, and then in both YPet and CyPet. As shown in

Table 1, which is a comparison of the FRET signals in various 12AA constructs, these two monomerizing mutations progressively decreased the FRET signal with the 12-amino acid insert. Moreover, the FRET signal from the AD66t1 construct was now much less than that of the 12AA construct when tested with the double-monomeric mutant (Table 2).

We then tested the effect of the monomerizing mutations in the YFP$_{VENUS}$–ECFP* pair. This FRET pair also had a tendency to form a dimer, but the interaction was much weaker than that of YPet–CyPet. Changes in the quantum yield and the molar extinction coefficient caused by the monomerizing mutation cannot explain the significant reduction of the FRET signals. The different FRET efficiencies between mYPet–mCyPet and mYFP$_{VENUS}$–mECFP* pairs are probably due to the different Föster distances (we estimated the Föster distance for the mYFP$_{VENUS}$–mECFP* pair to be 5.2 nm based on the molar extinction coefficient of mYFP$_{VENUS}$ at 514 nm [81,000 M^{-1} cm^{-1}] and the quantum yield of mECFP*, 0.58). The FRET signal from mYPet–mCyPet is slightly higher than that of mYFPvenus-mECFP* when using the emission ratio as the measure, and slightly lower when using the decrease in donor fluorescence, E_{DD}.

The bacterially expressed and purified FRET constructs containing various inserts were resolved on SDS-PAGE before and after trypsin digestion (Fig. 1). Several purified proteins showed a minor degradation product before trypsin digestion that could be an N-terminal mYPet fragment that was purified with the N-terminal His-tag. A minor acceptor contamination would not affect our results, because we used decrease in donor fluorescence as the measure of FRET. After mild trypsin digestion, mYPet ran as a monomeric GFP, and the mCyPet ran either as a monomer (when the insert was completely digested) or as a larger fragment when the insert remained attached (Fig. 1). It seems likely that K238, in the C-terminal flexible region of mYPet, is sensitive to trypsin digestion, and this cut disrupts the FRET pair. The emission intensities of the individually purified mYPet and mCyPet were not changed by mild trypsin digestion (data not shown). Previous studies

Table 1. *Comparison of monomeric mutations in various 12AA constructs*

Acceptor	Donor	E_{DD}	Emission ratio
YPet	CyPet	0.67	8.61
mYPet	CyPet	0.56	4.27
mYPet	mCyPet	0.51	3.53
Venus	ECFP*	0.61	3.36
mVenus	ECFP*	0.57	2.93
mVenus	mECFP*	0.56	2.84

E_{DD} is FRET efficiency calculated from decrease in donor fluorescence. Emission ratio is acceptor emission (528 nm)/donor emission (475 nm).

Table 2. *Estimated end-to end distance, FRET efficiency, and sedimentation coefficient for each construct*

	Insert size (aa)[a]	End-to-end distance of insert (nm)[b]	Flexible region (aa)[c]	r_{FRET} (nm)[d]	p (nm)[e]	E_{DD}	E_{DD} in urea	E_{DD} in Ficoll	S	Smax/S[f]
Unstructured										
12 AA	12	(2.6)	28	5.0	1.09	0.50	0.26	0.49	4.0	1.4
Ch33N	37	4.1	53	5.3	0.69	0.40	0.17	0.41	4.0	1.4
PQ33C	37	4.7	53	6.1	0.93	0.23	0.03	0.27	3.7	1.5
AD66t1	70	5.1	86	5.9	0.55	0.28	0.08	0.34	4.2	1.4
AD66t3	70	5.6	86	6.6	0.69	0.16	0.01	0.20	3.9	1.5
Ctt68	72	5.2	88	6.0	0.56	0.26	0.04	0.34	3.9	1.5
Ch66N	80	5.3	96	6.1	0.53	0.24	0.03	0.28	4.3	1.4
PQ66C	80	5.7	96	6.6	0.62	0.16	0.02	0.18	3.7	1.6
ChPQ150	164	7.7	180	8.2	0.53	0.05	−0.02	0.11	3.8	1.7
Structured										
GB1	60	2.6	16	6.4	-	0.19	0.04	0.21	4.1	1.4
FN10	98	3.7	16	6.4	-	0.19	0.12	0.24	4.3	1.4
FN7–8	188	7.0	16	-	-	0.01	−0.01	0.01	4.1	1.6

[a] Insert size is the number of amino acids in the domain inserted between the two GFPs, including the cloning site.
[b] For unstructured inserts the average end-to-end distance was calculated by Equation 6 for the worm-like-chain with the contour length (L) of the insert (0.34 nm per amino acid), and the estimated persistence length (p in column 5) set to match the r_{FRET}.
[c] Flexible region includes the unstructured 11 amino acids at the C terminus of mYPet and five amino acids at the N terminus of mCyPet.
[d] r_{FRET} is the donor–acceptor separation distance estimated from FRET efficiency (E_{DD}) by Equation 2.
[e] The persistence length was calculated from r_{FRET} by Equation 6.
[f] The ratio of the maximum sedimentation coefficient (S_{MAX}), calculated for an unhydrated sphere of protein of the same mass, to the measured sedimentation coefficient (S).

have also shown that trypsin digestion does not affect the fluorophores of GFP and its variants (Heim and Tsien 1996; Nguyen and Daugherty 2005; Shimozono et al. 2006). Unstructured domains such as Ch66N and Ctt68 were also digested with trypsin, although some unstructured segments such as PQ33C and ChPQ150 were not digested completely, probably due to their low lysine and arginine content. In contrast, structured domains such as FN10 and FN7–8 remained attached to mCyPet following trypsin digestion, suggesting that these domains were properly folded and not subject to digestion.

The emission spectra of selected FRET constructs before and after trypsin cleavage are shown in Figure 2. In all constructs except for FN7–8, donor (mCyPet) intensities increased and acceptor (mYPet) intensities decreased after trypsin cleavage, indicating that FRET occurs in these constructs. 12AA and Ch33N showed strong FRET signals. PQ33C, Ch66N, PQ66C, Ctt68, AD66t1, AD66t3, GB1, and FN10 had moderate FRET signals. The 150-amino acid unstructured segment ChPQ150 gave a barely detectable FRET signal. FN7–8 had no detectable FRET signal, as expected for a rigid insert 7.0 nm long. The FRET efficiency (E_{DD}) was calculated from the measured emission intensities of mCyPet, with and without trypsin digestion according to Equation 1 in Methods and Materials (Table 2). Based on the Förster distance for the mCyPet and mYPet pair and the E_{DD} calculated from the donor fluorescence change, the pseudo-average separation distances of fluorophores, r_{FRET}, were determined from Equation 2. The

reproducibility of our measurements was checked for a few constructs by repeating the experiments with protein that came from different purification batches. The error range of E_{DD} was about 2%, (compare the 12AA construct in Table 2, $E_{DD} = 0.50$, with its equivalent mYPet–mCyPet construct in Table 1, $E_{DD} = 0.51$).

We initially tried to interpret the results by using a single persistence length for all flexible constructs, and comparing r_{FRET} with the calculated fluorophore separation of the worm-like chain. However, it is clear from Table 2 that the experimentally determined r_{FRET} varied

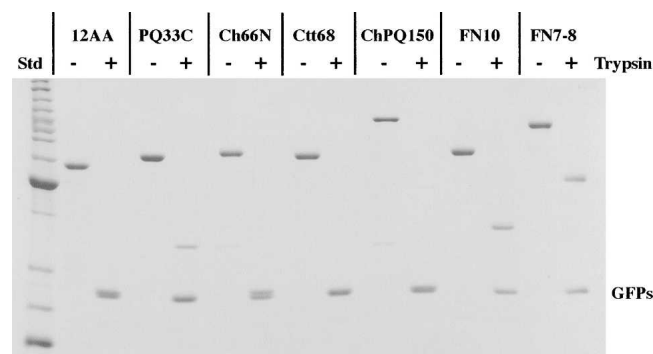**Figure 1.** FRET constructs before and after trypsin cleavage. Samples (1 μM) were analyzed by SDS-PAGE and stained with Coomassie Blue. As seen in the 12AA construct, trypsin digestion separated the two GFPs from each other. PQ33C, FN10, and FN7–8 remained linked to one of the GFPs, but the Ch66N, Ctt68, and ChPQ150 inserts were almost completely digested by trypsin. (Std) BenchMark protein ladder (Invitrogen).
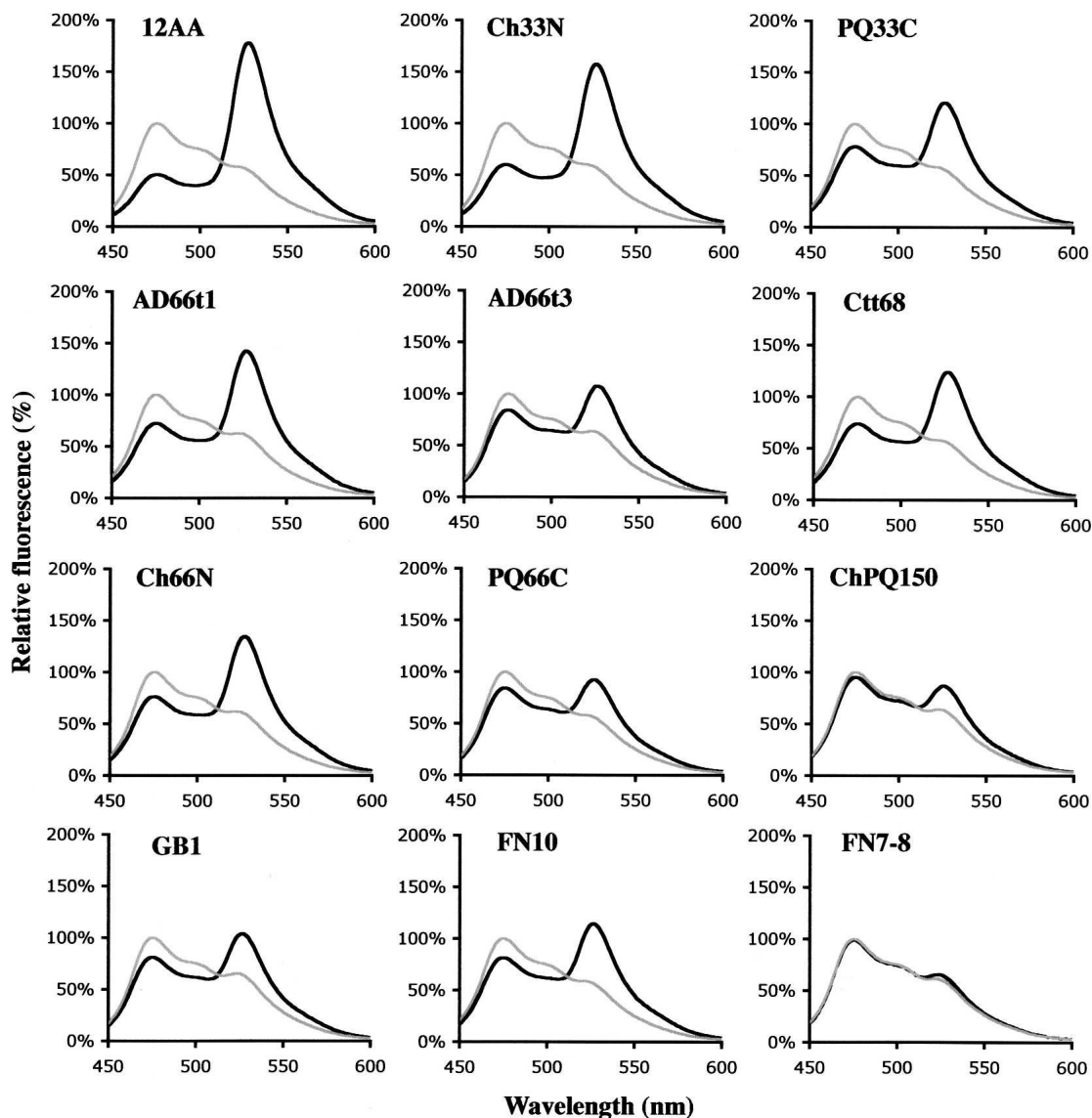
Figure 2. Emission spectra of the FRET constructs. Samples before (black line) and after (gray line) trypsin cleavage were excited at 433 nm and the relative emission intensities were measured at wavelengths from 450 to 600 nm. Strong FRET signals were detected in the 12AA and Ch33N constructs, a very weak signal in ChPQ150, and no signal in FN7–8; the remaining constructs showed moderate levels of FRET.

for flexible inserts of the same length. These segments apparently have a variable degree of residual order and stiffness. The calculated fluorophore separation of the worm-like chain is quite sensitive to the persistence length, so we decided that the most informative interpretation would be to determine the persistence length that would have the fluorophore separation of the worm-like chain match the experimental r_{FRET}. In this interpretation, the different "intrinsically unstructured" segments are considered to have a greater or lesser degree of stiffness.

For most inserts, the persistence lengths were in the range of from 0.5 to 0.7 nm. This is similar to the value of 0.66 nm found in our previous EM study of ZipA, and slightly above the values of 0.4–0.5 nm found in AFM studies of force-induced domain unfolding (Oberhauser et al. 1998; Dietz and Rief 2004). However, the PQ33C construct gave a lower FRET signal than expected, based on the length of this insert, and the estimated persistence length was 0.9 nm. This segment is apparently stiffer and more extended than the rest of the unstructured inserts. 12AA also had a longer persistence length, but we believe this arises not from the stiffness of the linker, but from a limitation on our calculations. With this short linker, many configurations would have the two GFPs bumping

into each other, and the end-to-end distance calculated for the worm-like chain cannot be achieved. Our estimate thus fails for very short flexible inserts, but seems fairly good for inserts of 30 amino acids or longer.

We examined the effect of 6 M urea on the FRET signals of our constructs (Fig. 3; Table 2). In all constructs examined, FRET signals were substantially reduced in the presence of urea. In several constructs, however, weak or moderate FRET signals were still detected. These constructs seem to have short or flexible inserts such as 12AA, Ch33N, AD66t1, and Ctt68, or chemically very stable structured inserts such as FN10. In control experiments, we observed that urea had no effect on the individually purified mYPet intensities, but slightly reduced the emission of mCyPet (data not shown). This is probably why there were a few negative E_{DD} values in the presence of urea (Table 2). We conclude that 6 M urea substantially increases the stiffness and extension of unstructured peptides.

We also tested whether macromolecular crowding affects the conformation of intrinsically unstructured protein, because macromolecular crowding is considered to stabilize protein structures (Ellis 2001; Chebotareva et al. 2004; Despa et al. 2005; Minton 2005). In the presence of the crowding agent Ficoll 70, the FRET signals were slightly increased in most constructs, including the structured ones (Table 2). This suggests that the FRET constructs form a slightly more compact conformation under crowding conditions. Ficoll did not affect the emission spectra of individually purified mYPet and mCyPet. The slightly increased refractive index (1.35) of Ficoll 70 affected the quantum yield but did not alter the Förster distance (see Equations 3 and 4).

The measured sedimentation coefficients of these FRET constructs are shown in Table 2. Smax/S is an indication of the relative extension of the proteins (Schürmann et al. 2001). A value of 1.6–1.9 is characteristic of a moderately extended protein such as TNfn1–5. This segment of tenascin contains five FN-III repeats in a rigid rod, 14.6 nm long by 2.5 nm diameter, and has Smax/S = 1.65 (Schürmann et al. 2001). Most of our constructs had Smax/S of 1.4–1.5, which suggests that they are monomeric proteins somewhat less elongated than TNfn1–5. This is reasonable for all constructs, based on the size of the inserts. Only three constructs violate this generalization. PQ66C, ChPQ150, and FN7–8 appear to be more elongated, which correlates with their low FRET signals.

Rotary shadowing EM images of several FRET constructs are shown in Figure 4. Two GFPs are seen as closely spaced globular domains. The unstructured domains connecting them are invisible in these images, as expected for unstructured proteins, while the structured inserts such as FN10 are seen as a short rod between the two GFPs. The two GFPs of the Ch33N constructs are relatively closer to each other than those of the PQ33C constructs, consistent with the FRET results.

Discussion

Using FRET to measure distances of rigid and flexible protein inserts

Figure 5 illustrates hypothetical conformations of the structured and unstructured constructs. The figure emphasizes an important point, that the separation of the fluorophores is not simply related to the size of the insert. Even for a rigid insert like FN7–8, the flexible C terminus of YFP and N terminus of CFP permit a substantial range of separations. In most configurations of the FN7–8 insert, the YFP and CFP are too far apart to generate FRET, but in some they can be folded back and even brought into contact. However, these close configurations are apparently rare because the overall FRET signal with FN7–8 was undetectable. With the shorter FN10 insert, we obtained a moderate FRET signal, which gave an r_{FRET} of 6.4 nm. This seems reasonable for the 3.7-nm rigid insert and the ensemble of conformations
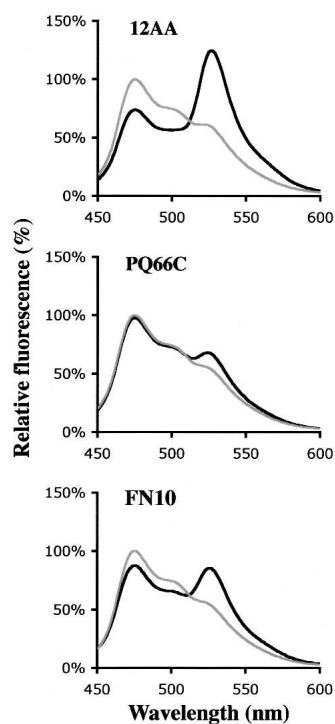

Figure 3. Emission spectra of the FRET constructs in urea. Samples before (black line) and after (gray line) trypsin cleavage in 6 M urea were excited at 433 nm and the relative emission intensities were measured at wavelengths from 450 to 600 nm. FRET signals were still detected in the 12AA and FN10 constructs, but were minimal in the PQ66C construct.
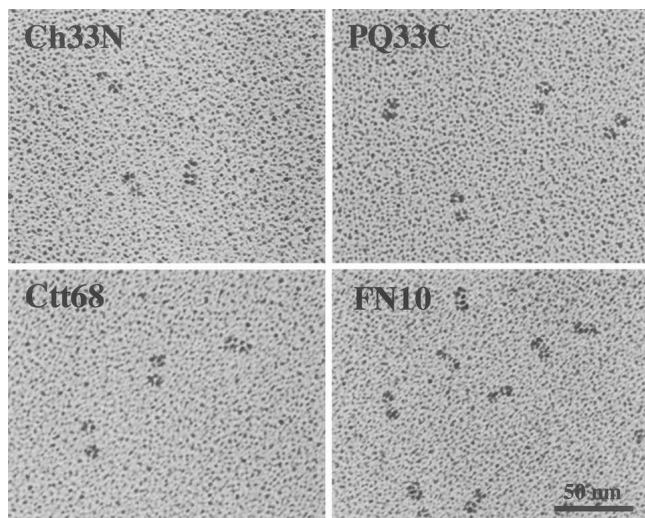
Figure 4. Rotary shadowing electron microscope images of the FRET constructs. The two GFPs are seen as closely spaced globular particles. The unstructured inserts, Ch33N, PQ33C, and Ctt68 between GFPs are invisible in EM images. The structured FN10 constructs show short rods connecting the two globular domains.

from the flexible ends (see below for discussion of ensembles). We had expected the shorter GB1 insert (2.6 nm) to give an increased FRET signal, but its FRET was identical to that of FN10. As in the case of the 12AA construct, the shorter GB1 insert may be limited in its FRET by the exclusion of configurations where the two GFPs would overlap. Another possibility is that GB1 (and perhaps FN10 as well) may interact with the YFP and CFP and restrict their rotational configurations.

For flexible inserts, we chose to determine a persistence length that would match the calculation of the worm-like chain to the measured r_{FRET}. The persistence length varied from 0.5 to 0.7 nm (Table 2) for most of the presumed unstructured inserts. Those with $P = 0.5$–0.55 are probably the most unstructured and flexible, and those with $P = 0.6$–0.7 are somewhat stiffer. PQ33C had $P = 0.93$, the highest value in the set, so we conclude that this peptide is definitely stiffer and more extended than the average unstructured segment. Note that PQ66C, which contains PQ33C, also had a higher $P = 0.62$.

In a study closely related to ours, Evers et al. (2006) prepared a series of constructs with engineered flexible linkers of 23–71 amino acids between ECFP and EYFP. Their inserts consisted of a variable number of repeats of GGSGGS, and they modeled this as a worm-like chain with a persistence length of 0.45 nm. This is somewhat less than the 0.5 nm of our most flexible inserts, which is reasonable, since GGSGGS should be at the extreme of flexibility. Evers et al. (2006) used computer modeling to calculate a large ensemble of possible configurations, using a worm-like chain model for the flexible segments.

They calculated the FRET signal for each configuration in the ensemble, and then determined the average FRET signal. An important advantage of this approach is that it accounts for the fact that configurations that bring the fluorophores close together contribute more to the net FRET than they do to the conformational average. Our approach, based on a single pseudo-average structural estimate, ignores this complication.

While Evers et al. (2006) used computer modeling that should accurately determine how the complete ensemble of configurations make up the FRET signal, our work suggests that these complex calculations are not needed for most applications. Our much simpler approach, using Equation 6 to calculate the fluorophore separation for a worm-like chain, and matching this to r_{FRET}, gives a fit almost as good. For example, if we use our calculation for Evers's 71-amino acid construct, we get a persistence length of 0.47 nm, which is close to the 0.45 nm that they used for their modeling. Also, our results show that intrinsically unstructured proteins actually vary in their stiffness, as discussed above.

We suggest that the most valuable application of FRET to characterize unstructured proteins is to estimate the persistence length and compare this to other proteins. Our simplified modeling approach appears to be satisfactory for this for inserts longer than 30 amino acids. Our set of eight unstructured and three structured inserts, covering the range from very short linkers with high FRET to longer ones with no FRET, provides a calibration set against which new proteins can be compared.

GFP dimerization

GFP and its variants are thought to form dimers, because GFPs are often seen as an antiparallel dimer in crystal structures (Ormo et al. 1996; Rekas et al. 2002). The dissociation constant for a GFP homodimer has been estimated to be ~0.1 mM (Phillips 1997; Zacharias et al. 2002). In our FRET experiments, the two linked GFPs would have a very high local concentration. For example, r_{FRET} is 8.2 nm in our longest linker, ChPQ150. If the GFPs averaged 8.2-nm apart in solution, their concentration would be ~3 mM, well above the estimated 0.1 mM K_D. However, in our earlier work using ECFP*–YFP$_{VENUS}$ we saw negligible FRET signal with this 150-amino acid insert (our unpublished observation), and we concluded that dimerization was not a problem. This would suggest that the K_D for dimerization is above 3 mM, i.e., much weaker than previously estimated.

Nguyen and Daugherty (2005) reported the development of an improved FRET pair, YPet–CyPet, which had a sevenfold enhancement of the FRET signal. However, we discovered that most of the signal enhancement was due to enhanced dimerization of YPet to CyPet within the
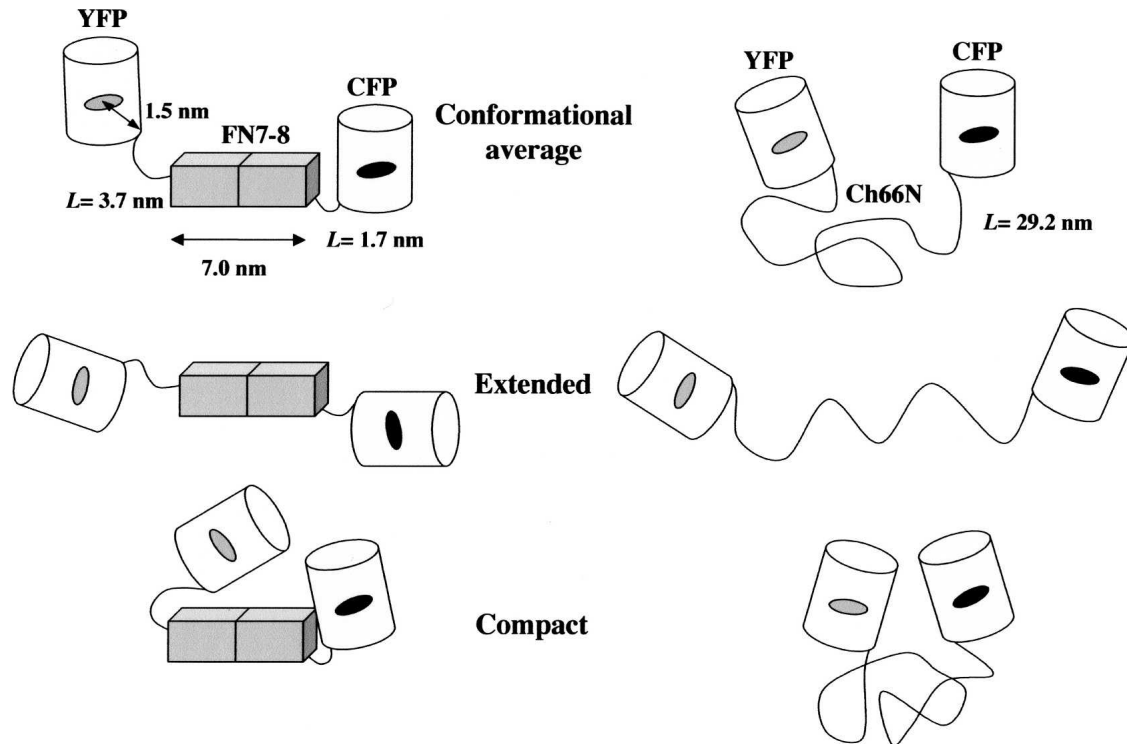
Figure 5. Hypothetical conformations of the FRET constructs. The diagram shows several possible conformations of the FRET constructs. The flexible C terminus of YFP and N terminus of CFP in a structured construct such as FN7–8 can bring two GFPs close together (as in the compact conformation) or space them further apart (as in the extended conformation). In the case of Ch66N, the unstructured polypeptide chain can also be in an extended or compact conformation.

tethered construct, and was substantially reduced when we incorporated the monomerizing mutation previously developed to prevent the weak dimerization of GFP (Zacharias et al. 2002). The problem was likely caused by the S208F mutation in YPet, which localizes in the dimer–dimer interface, and probably increases the affinity for dimerization. Nguyen and Daugherty (2005) noted that the enhancement in the CFP2–YFP3 pair was due to substitutions in YFP3, which was where S208F appeared. In a later application, You et al. (2006) used CyPet–YPet to assay peptide binding to ligands. The FRET assay indicated binding affinities three to 20 times higher than other assays, consistent with the possibility that GFP dimerization was enhancing the binding.

We then returned to the YFP$_{VENUS}$–ECFP* pair, and found that introducing the monomerizing mutations here also reduced the FRET signal, but the effect was much weaker than with YPet–CyPet. Even with shorter inserts, the dimer probably exists for only a small fraction of configurations. This could explain why Evers et al. (2006) saw no evidence for a dimer by fluorescence anisotropy. However, it does suggest that the FRET efficiencies measured by Evers et al. (2006) are slightly increased by the weak dimerization. As discussed above, this could be compensated in the modeling by using a slightly higher value for the persistence length.

Effects of urea and crowding agents on unstructured proteins

Several studies have demonstrated that protein backbone structures are extended or stiffened by denaturants (Liu et al. 2004; Mohana-Borges et al. 2004; Whittington et al. 2005). We tested this with all of our constructs. We found that 6 M urea substantially reduced the FRET signal for all flexible inserts, mostly to values that were barely measurable. This suggests that in the presence of urea, the unstructured proteins went from a fairly compact worm-like chain to a more rigidly extended chain. The FRET signal for the rigid insert FN10 was also substantially reduced. This was not due to denaturation of FN10, which is known to be chemically very stable (Plaxco et al. 1997; Cota et al. 2000) and does not denature with 6 M urea (our unpublished observation). The reduction of FRET signal is likely due to stiffening and extension of the N- and C-terminal flexible segments of the GFPs.

Macromolecular crowding is increasingly understood to affect protein structure and function inside the cell (Minton

and Wilf 1981; Minton 2005). Macromolecular crowding enhances the enzymatic activity of proteins as well as protein–protein interactions and polymerization (Ellis 2001; Chebotareva et al. 2004; Despa et al. 2005). The crowding effect seems also to increase protein stability. It has been reported that denatured apomyoglobin can be stabilized under crowding conditions (McPhie et al. 2006). On the other hand, crowding has been shown not to induce the formation of secondary structures in intrinsically unstructured proteins, e.g., c-Fos and p27[kip1] (Flaugh and Lumb 2001). In-cell NMR of the intrinsically unstructured protein α-synuclein indicated that while crowding could not induce structure, it did prevent the formation of aggregates (McNulty et al. 2006). Interestingly, in the case of FlgM, which is known to be unstructured in solution, the C-terminal region seems to be structured inside the cell (Dedmon et al. 2002). In the present study, we used Ficoll 70 to create crowding conditions to mimic the inside of the cell. Most unstructured constructs had a significantly more compact conformation in the presence of Ficoll. Thus, the end-to-end distances of intrinsically unstructured proteins inside the cell are probably shorter than the estimated lengths in solution.

Materials and Methods

Protein expression

For most of our work we used the recently developed FRET pair YPet (YFP variant) and CyPet (CFP variant) (Nguyen and Daugherty 2005). Constructs optimized for mammalian codon expression were kindly provided by Dr. Patrick Daugherty (University of California, Santa Barbara) and were mutated to eliminate a KpnI restriction enzyme site for our subsequent cloning. To generate monomeric forms of YPet and CyPet, alanine 206 was substituted with lysine, as reported previously for other GFP variants (Zacharias et al. 2002). Note that the residue numbering of GFP used in the present article does not include the second residue, valine, which is inserted in the mammalian codon-optimized GFPs. The PCR-amplified monomeric YPet (mYPet) and CyPet (mCyPet) fragments were cloned into pET15b (Novagen). The mYPet construct has the sequence: mgsshhhhhhssglvprgshmggrMVSK...(mYPet)... ELYK tsggr (the mYPet sequence is underlined, and the sequences in lowercase are derived from the cloning sites and linkers). The mCyPet construct has the sequence: mgsshhhhhhssglvprgshmg grsrtsgspglqefgtMVSK...(mCyPet)... ELYKggr. For creating the parent construct for FRET experiments, the mCyPet fragment was removed by HindIII and SpeI digestion and inserted into the mYPet construct. This construct has a 12-amino acid spacer, TSGSPGLQEFGT (whose DNA encodes SpeI–BamHI–SmaI–PstI–EcoRI–KpnI sites), between mYPet and mCyPet (Fig. 6). This construct, which we call 12AA, has the sequence: mgsshhhhhhssglvprgshmggrMVSK...(mYPet)...   ELYKtsgspg lqefgtMVSK...(mCyPet)... ELYKggr. We also generated YPet–CyPet (without the monomeric mutations), YFP_VENUS–ECFP*, and mYFP_VENUS–mECFP* FRET pairs for comparison. ECFP* indicates that our ECFP had two additional mutations, K26R/N164H,
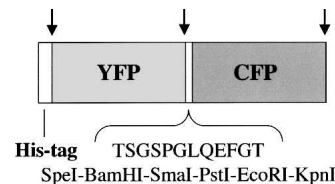

Figure 6. A diagram of the parental FRET construct (12AA). The structured and unstructured proteins were inserted into the SpeI and KpnI sites. Arrows indicate the putative trypsin sensitive sites.

which do not localize to the dimer–dimer interface, and appear to be innocuous.

Several segments of intrinsically unstructured proteins were inserted between mYPet and mCyPet. These included the charged and PQ domains of ZipA (Ohashi et al. 2002), the tail domain of α-adducin that has been characterized with CD spectroscopy (Hughes and Bennett 1995), and the C-terminal tail domain of FtsZ predicted to be unstructured (Erickson 2001). Specific constructs included the full-length 150 amino acid charged-plus-PQ domain (we designate this ChPQ150; the sequence at the linker site is -ts-DRPL...[charged-plus-PQ domain]...VMDK-ts), the N-terminal 66 amino acids of the charged domain (Ch66N; -ts-DRPL...[Ch]...QPRQ-ts), the N-terminal 33 amino acids of the charged domain (Ch33N; -ts-DRPL...[Ch]...RVHR-gt), the C-terminal 66 amino acids of the PQ domain (PQ66C; -ts-AQPV...[PQ]...VMDK-ts), and the C-terminal 33 amino acids of the PQ domain (PQ33C; -ts-SAPQ...[PQ])...VMDK-gt). Two different segments of 66 amino acids of the tail domain of α-adducin were tested: (AD66t1[440–505]; -ts-QQRE...[ADt]... MRNK-gt, AD66t3 [592–657]; -ts-EARE...[ADt]...GFPM-gt), and the C-terminal tail domain of FtsZ (Ctt68, -tsh-MDKR...[Ctt]...KQAD-gt). For comparison, we also tested different sized structured proteins between mYPet and mCyPet: the B1 immunoglobulin binding domain of streptococcal protein G (GB1; -ts-MQYK ...[GB1]... TVTE-gt-) (Franks et al. 2006) (PDB: 2GI9), FN-III domain 10 of fibronectin (FN10; -ts-VSDV...[FN10]... NYRT-gt-), and FN-III domains 7–8 (FN7–8; -ts-PLSP... [FN7–8]...RQKT-gt) (Leahy et al. 1996) (PDB: 1FNF).

Most of the constructs were insoluble following expression in *Escherichia coli* BL21 (DE3) at 37°C. When they were expressed at 20°C, a sufficient amount of soluble protein was obtained for experiments. We later learned that the solubility was further improved when they were expressed in *E. coli* C41 (DE3) at 20°C (Miroux and Walker 1996). The recombinant protein from the soluble fraction was purified with a cobalt-agarose column (TALON, Clontech) using standard procedures. Eluted proteins from the column were dialyzed against 20 mM Tris with 150 mM NaCl (TBS, pH 8.0) to remove imidazole. Protein concentration was estimated from the absorbance at 280 nm using the molar extinction coefficient of each protein calculated by the Protean computer program (DNAstar, Inc.). The molar extinction coefficients of mCyPet at 433 nm (26,000 $M^{-1}cm^{-1}$) and for mYPet at 514 nm (85,000 $M^{-1}cm^{-1}$) in TBS were estimated based on the protein concentrations determined at 280 nm. In these estimates, we assumed that all proteins were correctly folded and the fluorophores were properly formed. As a control to eliminate FRET, purified proteins (4 μM) were digested with trypsin (10 μg/mL) at room temperature for 1 h. Then, 4-μM samples with and without trypsin treatment were diluted four times with TBS, 8 M urea/TBS, or 20% Ficoll

70/TBS for fluorescence analysis. SDS-PAGE was performed using standard procedures.

FRET analysis

Fluorescence measurements were performed with a spectro-fluorophotometer (Shimadzu RF-5301-PC). Emission spectra were collected at 1-nm intervals from 450 to 600 nm with excitation at 433 nm using slit widths of 3 nm for excitation and 5 nm for emission. Spectra were acquired at room temperature for purified proteins at a concentration of 1 μM in TBS. At this concentration, triplicate measurements showed errors were within 5% and the absorbance at 433 nm was below 0.05, so that the inner filter effect should be insignificant. The relative intensity of emission spectra was normalized to the donor-alone intensity (after trypsin treatment) at 475 nm. Individually purified mYPet showed a small peak at 528 nm when excited at 433 nm, but no detectable emission at 475 nm. The normalized emission intensities at 475 nm for mCyPet were used to calculate the FRET efficiency as given by the decrease in donor intensity (E_{DD}) according to the equation (see Equation 13.14 in Lakowicz 1999):

$$E_{DD} = 1 - \frac{C_{FRET}}{C_{TRYPSIN}}, \tag{1}$$

where $C_{TRYPSIN}$ is the intensity of CFP after trypsin treatment and C_{FRET} is the intensity of CFP before trypsin digestion.

The steady-state FRET signal permits the calculation of a single estimate for the separation of the fluorophores, r_{FRET}, using the equation (see Equation 13.12 in Lakowicz 1999):

$$r_{FRET} = R_0 \sqrt[6]{\frac{1}{E_{DD}} - 1}, \tag{2}$$

where R_0 is the Förster distance. We calculated the Förster distance for the mYPet–mCyPet pair to be 5.0 nm from the following equation (see Equation 13.6 in Lakowicz 1999):

$$R_0 = \sqrt[6]{8.79 \times 10^{-11} [\kappa^2 n^{-4} Q_D J(\lambda)]} \text{ (in nm)}, \tag{3}$$

where κ^2 is the orientation factor between the donor and acceptor (typically equal to 2/3 for randomly oriented molecules, see also Evers et al. 2006), n is the refractive index of the medium (1.33 in water), Q_D is the quantum yield of the donor (determined below), and $J(\lambda)$ is the spectral overlap integral (determined below).

The Q_D for mCyPet was determined to be 0.44 from the following equation (see Equation 2.5 in Lakowicz 1999):

$$Q_D = Q_R \frac{I_D}{I_R} \cdot \frac{OD_R}{OD_D} \cdot \frac{n_D{}^2}{n_R{}^2}, \tag{4}$$

where I is the integrated intensity, OD is the optical density, and the subscripts D and R refer to the donor and reference fluorophore, respectively. We used fluorescein (Invitrogen) as a reference ($Q_R = 0.95$ in 0.1 N NaOH).

$J(\lambda)$ is the spectral overlap integral given by the following equation (see Equation 13.3 in Lakowicz 1999):

$$J(\lambda) = \frac{\int_0^\infty F_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda}{\int_0^\infty F_D(\lambda) d\lambda}, \tag{5}$$

where $F_D(\lambda)$ is the donor emission at a given wavelength λ, and $\varepsilon_A(\lambda)$ is the extinction coefficient of the acceptor at that wavelength. The emission spectrum of mCyPet and the absorbance of mYPet were measured from the individually expressed and purified proteins. Our estimated Förster distance for the mYPet–mCyPet pair (5.0 nm) was slightly larger than those reported for the CFP–YFP pair (4.8 and 4.9 nm) (Patterson et al. 2000; Evers et al. 2006), probably due to the higher quantum yield of mCyPet and molar absorptivity of mYPet.

We estimated the separation of the fluorophores, r_{FRET}, using the conventional Equation 2. We used this simple ''pseudo-average'' to calculate the persistence length of each flexible polypeptide chain using the following equation for a worm-like chain (WLC):

$$2pL\left(1 - \frac{p}{L}\left(1 - e^{-L/p}\right)\right) = \langle d_{WLC}^2 \rangle, \tag{6}$$

where p is the persistence length, and L is the contour length (Rivetti et al. 1996; Zhou 2004). The end-to-end separation of the flexible insert (d_{WLC}) was assumed to be equal to r_{FRET} to calculate the persistence length. The contour length was taken to be 0.34 nm per amino acid, which is the average length of an amino acid residue in an extended β-strand (see caption to Fig. 5F of Yang et al. 2000). An important additional step was to recognize that the crystal structures of GFP, and its variants show that approximately five amino acids at the N terminus and 11 amino acids at the C terminus are flexible (Ormo et al. 1996; Rekas et al. 2002). It has also been reported that the deletions of these flexible segments do not affect the GFP fluorophore (Shimozono et al. 2006). For the calculation of the contour length, we added these 16 amino acids to the length of the insert. The recent study of Evers et al. (2006) also included these segments as part of the flexible linker. We also added 3 nm to the contour length to account for the 1.5-nm distance of each fluorophore (in the center of the GFP) from the surface, as estimated with the PyMOL computer program (Delano Scientific). Although the 1.5-nm segment is actually rigid, this simplification is probably reasonable, because the GFP is attached to the peptide by a fully flexible joint.

Glycerol gradient sedimentation and electron microscopy

In order to estimate sedimentation coefficients, the purified proteins were sedimented at 20°C through a 15%–40% glycerol gradient in 0.2 M ammonium bicarbonate at 42,000 rpm for 16 h in a Beckman SW-55.1 rotor (Schürmann et al. 2001). The glycerol gradients were calibrated with standard proteins of known S value (catalase, 11.3 S, aldolase, 7.3; BSA, 4.6 S; ovalbumin, 3.5 S). For rotary shadowing, samples from the glycerol gradient fractions were sprayed onto freshly cleaved mica, dried in vacuum, and rotary shadowed with platinum (Fowler and Erickson 1979; Ohashi and Erickson 2004).

References

Ai, H.W., Henderson, J.N., Remington, S.J., and Campbell, R.E. 2006. Directed evolution of a monomeric, bright, and photostable version of Clavularia cyan fluorescent protein: Structural characterization and applications in fluorescence imaging. *Biochem. J.* **400:** 531–540.

Chebotareva, N.A., Kurganov, B.I., and Livanova, N.B. 2004. Biochemical effects of molecular crowding. *Biochemistry (Mosc.)* **69:** 1239–1251.

Cota, E., Hamill, S.J., Fowler, S.B., and Clarke, J. 2000. Two proteins with the same structure respond very differently to mutation: The role of plasticity in protein stability. *J. Mol. Biol.* **302:** 713–725.

Dedmon, M.M., Patel, C.N., Young, G.B., and Pielak, G.J. 2002. FlgM gains structure in living cells. *Proc. Natl. Acad. Sci.* **99:** 12681–12684.

Despa, F., Orgill, D.P., and Lee, R.C. 2005. Molecular crowding effects on protein stability. *Ann. N. Y. Acad. Sci.* **1066:** 54–66.

Dietz, H. and Rief, M. 2004. Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc. Natl. Acad. Sci.* **101:** 16192–16197.

Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M., and Obradovic, Z. 2002. Intrinsic disorder and protein function. *Biochemistry* **41:** 6573–6582.

Dyson, H.J. and Wright, P.E. 2005. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6:** 197–208.

Ellis, R.J. 2001. Macromolecular crowding: An important but neglected aspect of the intracellular environment. *Curr. Opin. Struct. Biol.* **11:** 114–119.

Erickson, H.P. 2001. The FtsZ protofilament and attachment of ZipA—structural constraints on the FtsZ power stroke. *Curr. Opin. Cell Biol.* **13:** 55–60.

Evers, T.H., van Dongen, E.M., Faesen, A.C., Meijer, E.W., and Merkx, M. 2006. Quantitative understanding of the energy transfer between fluorescent proteins connected via flexible peptide linkers. *Biochemistry* **45:** 13183–13192.

Flaugh, S.L. and Lumb, K.J. 2001. Effects of macromolecular crowding on the intrinsically disordered proteins c-Fos and p27(Kip1). *Biomacromolecules* **2:** 538–540.

Fowler, W.E. and Erickson, H.P. 1979. Trinodular structure of fibrinogen. Confirmation by both shadowing and negative stain electron microscopy. *J. Mol. Biol.* **134:** 241–249.

Franks, W.T., Wylie, B.J., Stellfox, S.A., and Rienstra, C.M. 2006. Backbone conformational constraints in a microcrystalline U-15N-labeled protein by 3D dipolar-shift solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **128:** 3154–3155.

Heim, R. and Tsien, R.Y. 1996. Engineering green fluorescent protein for improved brightness, longer wavelengths, and fluorescence resonance energy transfer. *Curr. Biol.* **6:** 178–182.

Hughes, C.A. and Bennett, V. 1995. Adducin: A physical model with implications for function in assembly of spectrin-actin complexes. *J. Biol. Chem.* **270:** 18990–18996.

Lakowicz, J.R. 1999. *Principles of fluorescence spectroscopy*, 2nd ed. Kluwer Academic, New York.

Leahy, D.J., Aukhil, I., and Erickson, H.P. 1996. 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell* **84:** 155–164.

Li, H., Oberhauser, A.F., Redick, S.D., Carrion-Vazquez, M., Erickson, H.P., and Fernandez, J.M. 2001. Multiple conformations of PEVK proteins detected by single-molecule techniques. *Proc. Natl. Acad. Sci.* **98:** 10682–10686.

Liu, Z., Chen, K., Ng, A., Shi, Z., Woody, R.W., and Kallenbach, N.R. 2004. Solvent dependence of PII conformation in model alanine peptides. *J. Am. Chem. Soc.* **126:** 15141–15150.

McNulty, B.C., Young, G.B., and Pielak, G.J. 2006. Macromolecular crowding in the *Escherichia coli* periplasm maintains α-synuclein disorder. *J. Mol. Biol.* **355:** 893–897.

McPhie, P., Ni, Y.S., and Minton, A.P. 2006. Macromolecular crowding stabilizes the molten globule form of apomyoglobin with respect to both cold and heat unfolding. *J. Mol. Biol.* **361:** 7–10.

Minton, A.P. 2005. Influence of macromolecular crowding upon the stability and state of association of proteins: Predictions and observations. *J. Pharm. Sci.* **94:** 1668–1675.

Minton, A.P. and Wilf, J. 1981. Effect of macromolecular crowding upon the structure and function of an enzyme: Glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry* **20:** 4821–4826.

Miroux, B. and Walker, J.E. 1996. Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260:** 289–298.

Mitra, R.D., Silva, C.M., and Youvan, D.C. 1996. Fluorescence resonance energy transfer between blue-emitting and red-shifted excitation derivatives of the green fluorescent protein. *Gene* **173:** 13–17.

Miyawaki, A., Llopis, J., Heim, R., McCaffery, J.M., Adams, J.A., Ikura, M., and Tsien, R.Y. 1997. Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature* **388:** 882–887.

Mohana-Borges, R., Goto, N.K., Kroon, G.J., Dyson, H.J., and Wright, P.E. 2004. Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. *J. Mol. Biol.* **340:** 1131–1142.

Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K., and Miyawaki, A. 2002. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20:** 87–90.

Nguyen, A.W. and Daugherty, P.S. 2005. Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat. Biotechnol.* **23:** 355–360.

Oberhauser, A.F., Marszalek, P.E., Erickson, H.P., and Fernandez, J.M. 1998. The molecular elasticity of the extracellular matrix protein tenascin. *Nature* **393:** 181–185.

Ohashi, T. and Erickson, H.P. 2004. The disulfide bonding pattern in ficolin multimers. *J. Biol. Chem.* **279:** 6534–6539.

Ohashi, T., Hale, C.A., De Boer, P.A., and Erickson, H.P. 2002. Structural evidence that the P/Q domain of ZipA is an unstructured, flexible tether between the membrane and the C-terminal FtsZ-binding domain. *J. Bacteriol.* **184:** 4313–4315.

Ormo, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y., and Remington, S.J. 1996. Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **273:** 1392–1395.

Patterson, G.H., Piston, D.W., and Barisas, B.G. 2000. Forster distances between green fluorescent protein pairs. *Anal. Biochem.* **284:** 438–440.

Persechini, A., Lynch, J.A., and Romoser, V.A. 1997. Novel fluorescent indicator proteins for monitoring free intracellular Ca^{2+}. *Cell Calcium* **22:** 209–216.

Phillips Jr., G.N. 1997. Structure and dynamics of green fluorescent protein. *Curr. Opin. Struct. Biol.* **7:** 821–827.

Plaxco, K.W., Spitzfaden, C., Campbell, I.D., and Dobson, C.M. 1997. A comparison of the folding kinetics and thermodynamics of two homologous fibronectin type III modules. *J. Mol. Biol.* **270:** 763–770.

Rekas, A., Alattia, J.R., Nagai, T., Miyawaki, A., and Ikura, M. 2002. Crystal structure of venus, a yellow fluorescent protein with improved maturation and reduced environmental sensitivity. *J. Biol. Chem.* **277:** 50573–50578.

Rivetti, C., Guthold, M., and Bustamante, C. 1996. Scanning force microscopy of DNA deposited onto mica: Equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J. Mol. Biol.* **264:** 919–932.

Schürmann, G., Haspel, J., Grumet, M., and Erickson, H.P. 2001. Cell adhesion molecule L1 in folded (horseshoe) and extended conformations. *Mol. Biol. Cell* **12:** 1765–1773.

Shimozono, S., Hosoi, H., Mizuno, H., Fukano, T., Tahara, T., and Miyawaki, A. 2006. Concatenation of cyan and yellow fluorescent proteins for efficient resonance energy transfer. *Biochemistry* **45:** 6267–6271.

Tompa, P. 2002. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27:** 527–533.

Tsien, R.Y. 1998. The green fluorescent protein. *Annu. Rev. Biochem.* **67:** 509–544.

Uversky, V.N. 2002. Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.* **11:** 739–756.

Whittington, S.J., Chellgren, B.W., Hermann, V.M., and Creamer, T.P. 2005. Urea promotes polyproline II helix formation: Implications for protein denatured states. *Biochemistry* **44:** 6269–6275.

Wu, P. and Brand, L. 1994. Resonance energy transfer: Methods and applications. *Anal. Biochem.* **218:** 1–13.

Xu, X., Gerard, A.L., Huang, B.C., Anderson, D.C., Payan, D.G., and Luo, Y. 1998. Detection of programmed cell death using fluorescence energy transfer. *Nucleic Acids Res.* **26:** 2034–2035.

Yang, G., Cecconi, C., Baase, W.A., Vetter, I.R., Breyer, W.A., Haack, J.A., Matthews, B.W., Dahlquist, F.W., and Bustamante, C. 2000. Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme. *Proc. Natl. Acad. Sci.* **97:** 139–144.

You, X., Nguyen, A.W., Jabaiah, A., Sheff, M.A., Thorn, K.S., and Daugherty, P.S. 2006. Intracellular protein interaction mapping with FRET hybrids. *Proc. Natl. Acad. Sci.* **103:** 18458–18463.

Zaccolo, M., De Giorgi, F., Cho, C.Y., Feng, L., Knapp, T., Negulescu, P.A., Taylor, S.S., Tsien, R.Y., and Pozzan, T. 2000. A genetically encoded, fluorescent indicator for cyclic AMP in living cells. *Nat. Cell Biol.* **2:** 25–29.

Zacharias, D.A., Violin, J.D., Newton, A.C., and Tsien, R.Y. 2002. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296:** 913–916.

Zhou, H.X. 2004. Polymer models of protein stability, folding, and interactions. *Biochemistry* **43:** 2141–2154.