Polyproline has recently been used as a spacer between donor and acceptor chromophores to help establish the accuracy of distances determined from single-molecule Förster resonance energy transfer (FRET) measurements. This work showed that the FRET efficiency in water is higher than expected for a rigid spacer and was attributed to the flexibility of the polypeptide. Here, we investigate this issue further, using a combination of single-molecule fluorescence intensity and lifetime measurements, NMR, theory, and molecular dynamics simulations of polyproline-20 that include the dyes and their linkers to the polypeptide. NMR shows that in water ~30% of the molecules contain internal cis prolines, whereas none are detectable in trifluoroethanol. Simulations suggest that the all-trans form of polyproline is relatively stiff, with persistence lengths of 9–13 nm using different established force fields, and that the kinks arising from internal cis prolines are primarily responsible for the higher mean FRET efficiency in water.

We show that the observed efficiency histograms and distributions of donor fluorescence lifetimes are explained by the presence of multiple species with efficiencies consistent with the simulations and populations determined by NMR. In calculating FRET efficiencies from the simulation, we find that the fluctuations of the chromophores, attached to long flexible linkers, also play an important role. A similar simulation approach suggests that the flexibility of the chromophore linkers is largely responsible for the previously unexplained high value of \( R_0 \) required to fit the data in the classic study of Stryer and Haugland.
molecule photon trajectories from continuous wave excitation of immobilized polyprolines of 8–24 residues. After correcting for shot noise, they concluded that there is a static distribution of different conformers of polyproline. Using a worm-like chain model, they calculated that a persistence length of 2.3 nm would be required to fit the observed mean efficiency. Watkins et al. (21) also suggested that the structure distribution resulted from the presence of cis prolines but made no estimate of their contribution.

Given the importance of experiments on polyproline for quantifying distance information in single-molecule FRET experiments, as well as the large variation in reported persistence lengths, we have investigated this problem further. We have used NMR spectroscopy to determine the fraction and location of cis prolines (Fig. 1B), measured single-molecule photon trajectories using pulsed, picosecond excitation of freely diffusing molecules to obtain accurate FRET efficiencies from the fluorescence decay curves of subpopulations, and performed molecular dynamics calculations to interpret the results. The single-molecule lifetime and intensity data are interpreted in terms of the NMR-determined populations of cis isomers and the efficiencies for these isomers obtained from molecular dynamics calculations that include the dyes and their linkers. To compare the FRET efficiency histograms with the predictions from the simulations, a theoretical model was used that accounts for both photon statistics and background noise. Measurements were also made in trifluoroethanol (TFE), believed to reduce the cis proline content relative to water (22).

Results

Fraction and Location of cis Residues in Polyproline Determined by NMR. The H− region of the 1H spectra of polyproline-8 and polyproline-20 [(Pro)8Gly and (Pro)20Gly] in D2O is shown in Fig. 2 A and B. The highest intensity peak at 4.73 ppm can be attributed to the major population of prolines; residues that are trans and followed by a trans peptide bond. The integrated peak intensities indicate that the smaller signals at 4.44, 4.64, and 4.78 ppm each correspond to ~1 residue. The resonances at 4.78 and 4.64 are assigned to residues 1 and 2 based on sequential H−H° NOEs, whereas at 4.44 lacks such NOEs, and instead shows an NOE to the amide proton of the C-terminal Gly (data not shown). Four additional, smaller resonances at 4.31, 4.53, 4.60, and 4.85 are also apparent for both peptides in D2O. Strong H°H− NOEs between the signals at 4.31 and 4.85 (broken lines in Fig. 2 D and E), and 4.53 and 4.60 (solid lines in Fig. 2 D and E), indicate two pairs of proline residues that are each separated by a cis peptide bond. The first pair is due to cis proline within the main chain because it scales with the number of residues and the signal must therefore be due to cis proline within the main chain. The intensity of the latter pair scales with the signals for the terminal residues and is attributed to a small population of cis proline at the C terminus. From peak intensity ratios, the fraction of cis proline at the C terminus is ~10% (10.3 ± 1.0% in polyproline-8 and 11.9 ± 2.0% in polyproline-20). The relative peak integrals indicate that the fraction of cis peptides within the chain is ~2% (1.8 ± 0.2% in polyproline-8 and 1.9 ± 0.2% in polyproline-20).

In the NOESY spectrum of polyproline in TFE, only a single H°H− NOE was observed (Fig. 2F). The absence of sequential H°H° NOEs to either signal confirms that this cross-peak is due to a cis proline at the C terminus. The corresponding peaks on the diagonal appear as shoulders on the larger peaks at 4.42 and 4.49 (indicated by asterisks in Fig. 2C). Because of this overlap, the cis fraction at the C terminus cannot be obtained from 1D integration, however, from an analysis of the intensities of the diagonal and cross peaks in the NOESY spectrum, we estimate that ~12.5 ± 3.0% of the C-terminal residues are cis. There is no detectable H−H° NOE signal for internal cis residues; assuming that the cis and trans signals are not exactly overlapped, this result indicates that the fraction of internal cis residues is <0.1%.

FRET Efficiency Histograms: Shot Noise, Bleaching, and Blinking. Photon trajectories were recorded in TFE for dilute solutions of freely diffusing polyproline-20 labeled at the N terminus with a FRET acceptor (Alexa Fluor 594) and at the C terminus with a FRET donor (Alexa Fluor 488) (Fig. 1). The trajectories were divided into “bins” of 2 ms and those bins having a sum of donor and acceptor photons of <25 were discarded. Before further analysis, the data were corrected for differences in donor/acceptor quantum yield and detection efficiency, by random deletion of acceptor photons (23) [see supporting information (SI) Text]. No background subtraction was performed; rather, the background was considered explicitly in later analysis. Fig. 3 shows the distribution of transfer efficiencies. The main peak in the distribution in TFE (Fig. 3 A) occurs at an efficiency of ~0.34; the additional low efficiency peak is due to molecules lacking an active acceptor dye. In the corresponding efficiency histogram for the same experiment in water (Fig. 3B), the main peak is shifted toward a higher efficiency of ~0.45.

The widths of the main peaks in the efficiency distributions in TFE and water are wider than expected from statistical “shot noise” arising from the finite number of detected photons. The expected shot noise width was calculated from σn = (∑n−1)((1−E))/N)N−1 (19, 24, 25), where N is the number of photons in each bin, yielding shot-noise limited widths of 0.088 and 0.095 for polyproline-20 in TFE and water, respectively. The actual standard deviation of the main peak in the efficiency distribution is 0.116 (32% wider than shot noise) in TFE and 0.146 (54% wider than shot noise) in water.

The histograms of time delays were calculated for donor photons belonging to time bins on the low and high side of the donor/acceptor FRET peak in Fig. 3. There is a small difference...
donor photons from the subpopulations with corresponding colors in the shot-noise-limited width of the distribution (24, 19). Solid blue line in Insets gives the expected efficiency distribution for a heterogeneous mixture of species containing cis proline, taking the relative populations from NMR and the efficiencies from simulation. (Insets) The donor fluorescence decays for donor photons from the subpopulations with corresponding colors in the efficiency histograms.

between the two curves in TFE (Fig. 3A Inset), but a marked difference in water (Fig. 3B Inset). This difference indicates that some of the width arises from structural heterogeneity of polyproline that is not averaged on a time scale comparable with or longer than the interval between detecting photons (50–100 μs). We note that additional width cannot arise from heterogeneity in labeling (i.e., exchange of donor and acceptor attachment points) because different chemistry is used to attach the donor and acceptor chromophores.

Additional width could also arise from bleaching and blinking of the acceptor dye. In Fig. 4A, we compare the efficiency distributions calculated from the first and second half of each bin. The similarity of the distributions suggests that photobleaching of the acceptor dye is not a significant effect, because it would tend to shift the distribution from the second half of the bin to lower efficiency.

We also address the issue of “blinking” of the acceptor chromophore to a nonfluorescent state with a poor spectral overlap with the donor fluorescence, and occurring on time scales longer than or comparable with the interphoton detection interval (24). For bins belonging to each interval of 0.1 in efficiency in the FRET histograms the frequency distribution of “strings” of consecutive donor and acceptor photons of different lengths were calculated, and normalized by the number of possible strings of each length, given the empirical distribution of the number of photons per time bin. The distribution for donor photons is shown in Fig. 4B and the distribution for acceptor photons is shown in SI Fig. 12. In each such interval, the acceptor fraction of the total photons, which is denoted by ε, should be approximately the same. If the order of detection of donor and acceptor photons is random, the probabilities of a sequence of nA consecutive acceptor photons or nD consecutive donor photons are simply given by $p(n_A) = \varepsilon^n_A$ and $p(n_D) = (1 - \varepsilon)^n_D$, respectively. Blinking of the acceptor chromophore (or photobleaching) would give rise to a more frequent observation of long strings of donor photons. However, we find that the distributions match what is expected for uncorrelated emission, within error (Fig. 4B).

Conformational Distributions, Dynamics, and Persistence Length from Molecular Simulations. We used an implicit solvent model for all-trans polyproline, because its structure is essentially determined by repulsive interactions. However, we used a five-proline fragment attached to each dye in explicit solvent to sample the distribution of the dye conformations, because they are attached to the polyproline by flexible linkers. This multiscale approach avoids costly simulations of the (very long) dye-labeled polyproline in explicit solvent.

Fig. 5 summarizes the relative contribution of the polyproline flexibility and range of motion of the linkers to the interdye distance distribution for polyproline-20. The contributions of the linkers to donor-acceptor distance fluctuations were approximated by projecting the distance from the polyproline terminus to the center of each chromophore onto the axis of the polyproline helix.

In SI Fig. 8, we present the various correlation functions for polypeptide, donor, and acceptor motions.

To determine the persistence length from the simulations, the average projection of the end-to-end vector onto the initial chain direction was calculated as a function of chain length (see SI Fig. 10C). The extrapolated limit of this projection for very long chains, $\sim 13$ nm, corresponds to the persistence length, $\xi$. A second approach, which allows an estimate of the persistence length for shorter chains, uses the analytical approximation of Thirumalai and Ha (26) to the radial probability distribution for a worm-like chain. Fitting their equation to the end-to-end distribution for polyprolines gives persistence lengths in the range 9–12 nm for polyprolines of more than $\sim 20$ residues, using a number of different force fields (see SI Fig. 10A and B).

Comparing Donor Fluorescence Decays with Simulations. In Fig. 6A are shown the histograms of time delays for polyproline-20 in TFE. The donor lifetime in the absence of an acceptor was
obtained by using donor photons from time bins belonging to the “donor-only” peak in the FRET histogram, defined conservatively as $E < 0.1$. The time delay histogram for the main efficiency peak was obtained from donor photons from time bins with $E > 0.2$. The fluorescence intensity decay $I(t)$ was calculated from simulation using the time-dependent rates determined from the trajectory, averaging over multiple initial points (27).

The mean efficiency was obtained from integration of $I(t)$ as described in SI Text. The calculated intensity decay for all-trans polyproline in TFE is given by the red curve in Fig. 6A. Inclusion of the $\sim 12.5\%$ cis proline at the C terminus found by NMR, gives a very similar result (solid blue curve in Fig. 6A). We note that calculation of lifetime distributions, or FRET efficiencies, in the limit of slow chain dynamics (Eq. S6 in SI Text) is a very good approximation to the full calculation.

In water, the experimental $I(t)$ lies below that calculated from simulation of all-trans polyproline (red curve in Fig. 6B). Fluorescence decays for polyprolines with internal cis prolines were generated from implicit solvent simulations of polyproline-20 molecules with a single cis residue at each possible position, as for the all-trans molecule. The average over all-trans polyproline and the cis proline species, weighted by the NMR populations is remarkably close to the experimental curve (blue curve in Fig. 6B).

**Comparing FRET Efficiency Distributions with Simulation.** We characterize the heterogeneity in FRET efficiencies by a model in which there are an assumed number $K$ of species with populations $w_i$ and efficiencies $e_i$. Each time bin is associated with only one species (i.e., species do not interconvert over the duration of the bin). We use the experimentally determined distribution of the sum of donor and acceptor photons, to bypass the problem of modeling diffusion through the laser spot (23, 25, 28). The joint probability of $n_A$ acceptor photons and $n_D$ donor photons in a time bin is then given by (see SI Text):

$$p(n_A, n_D) = p(n_A + n_D) \frac{(n_A + n_D)!}{n_A!n_D!} \times \sum_{i=1}^{K} w_i \left( e_i \left( 1 - \frac{b_A + b_D}{N} \right)^i + \frac{b_A}{N} \right)^{n_A} \times \left( 1 - e_i \right)^{n_D} \left( 1 - \frac{b_A + b_D}{N} \right)^{n_D} \left( 1 + \frac{b_D}{N} \right)^{n_D}$$

where $p(n_A + n_D)$ is the probability of a bin containing $n_A + n_D$ photons, $b_A$ and $b_D$ are the number of background counts per bin in the acceptor and donor channels, and $N$ is the mean number of photons per bin (all of which can be obtained from experimental data).

For polyproline-20 in TFE, a three-species model was used (donor-only, all-trans, and C-terminal cis). The parameters $w_i$ and $e_i$ were optimized by maximizing the joint likelihood of all observed bursts, where Eq. 1 gives the likelihood of the observation in an individual time bin. The efficiency histogram back-calculated from the optimal parameters is plotted in Fig. 3A. The largest fitted population in TFE has an efficiency of 0.34 and a population of 81% (excluding donor-only). The other species (apart from donor-only) has an efficiency of 0.53 and a population of 19%. These populations are close to the NMR values of 87.5 ± 3.0% for all-trans polyproline, and 12.5 ± 3.0% for molecules with a C-terminal cis proline. The corresponding efficiencies from simulation, determined by integrating $I(t)$, are 0.40 and 0.61.

In water, the NMR data show that there is the possibility of many different polyproline conformations with various combinations of cis prolines (38 in total, neglecting the small population of molecules expected to have more than one internal cis...
proline), so any fit would be highly underdetermined. We therefore calculated the FRET efficiency histogram with the populations taken from the NMR analysis and the efficiencies calculated from the simulations for each of the 38 isomers, assuming a uniform distribution of internal cis residues. We find that the prediction of the histogram is in remarkably good agreement with the measured histograms.

Discussion

FRET has been extensively used for obtaining qualitative distance information in single-molecule experiments on biomolecules (14, 29, 30). More recently, experiments have suggested that despite the large chromophores and long linkers, it should be possible to obtain accurate quantitative distance information in both proteins and nucleic acids (15, 16, 19, 21, 31). However, quantitative analysis of single-molecule FRET experiments using polyproline of varying lengths as spacers between donor and acceptor dyes has raised some doubt. Using continuous wave excitation, Schuler et al. (15), and later Watkins et al. (21), found that the mean FRET efficiency was much higher than expected for polyproline acting as a rigid rod spacer and that the width of the FRET efficiency distribution is much greater than expected from shot noise alone (Figs. 3 and 6). Can we explain these results, as well as the previous work on proteins unfolded by NMR and assumed an equal probability for a single cis residue at each internal position. This enumeration results in a total of 38 isomers. The internal cis residues have a much bigger effect than the C-terminal cis residues, and result in a large decrease in fluorescence lifetime compared with the all-trans molecule. The calculated fluorescence decay curves are in remarkably good accord with the observed histogram of donor time delays.

Calculation of FRET efficiency histograms requires careful consideration of both the shot noise and background fluorescence. We first ruled out any significant contribution to the width of the distribution from bleaching or blinking of the acceptor dye by comparing the FRET efficiency in the first and second half of each burst of photons, and by examining the distribution of continuous strings of donor photons (Fig. 4).

We calculated FRET efficiency histograms using a model that allows for multiple species with different FRET efficiency (see Eq. 1). In the case of TFE, maximum likelihood was used to determine optimal populations and efficiencies for the species, consistent with the distribution of donor and acceptor photons in the individual bursts as the molecules diffuse through the detection volume. The efficiency histogram computed from the optimal parameters explains the very small excess width above that expected from shot noise. The optimal populations are similar to those determined by NMR and the efficiencies are only slightly less than obtained from the simulations (Fig. 3). In the case of water, direct calculation of the efficiency histogram from 38 different isomers with populations determined from the NMR experiments and mean efficiencies from simulations is remarkably similar to the observed histogram. Although internal cis prolines bring the ends of the molecule closer together, the mean efficiency is only modestly increased because the dynamics associated with the additional flexibility is effectively slow relative to the donor lifetime (see SI Fig. 8).

We find that linker dynamics plays an important role: the donor dye is conjugated to the C-terminal cysteine residue by a very flexible five-carbon linker and undergoes large projected fluctuations (rms of 0.9 nm; Fig. 5B), compared with much smaller fluctuations of the acceptor, attached to the C-terminal glycine by a two carbon linker (0.25 nm Fig. 5C). Thus, the overall donor-acceptor distance distribution in Fig. 5D largely reflects the mobility of the donor chromophore. Whereas this mobility has the advantage that the orientational contribution to FRET, $\kappa^2$, is close to the isotropic value of 2/3, it also means that the contribution of linker dynamics to donor-acceptor separation needs to be carefully considered when determining distance information from FRET.

A related question that arose in the course of this work was whether donor/acceptor dynamics might explain the discrepancy mentioned earlier concerning the $R_0$ in the classic study of Stryer and Haugland (1). We found that simulations that included the dynamics of a naphthalenyl donor and dansyl acceptor resulted in increased FRET efficiency, explaining a large part of the difference between the calculated curves using the experimentally determined $R_0$ and the fitted $R_0$ (see SI Fig. 11).

The consistency of the results from NMR, single-molecule lifetime and intensity measurements, and molecular dynamics simulations indicates that, despite the structural complexity, it is indeed possible to understand single-molecule FRET experiments with polyproline spacers in quantitative detail. These results, as well as the previous work on proteins unfolded by chemical denaturants (6–20), suggests that single-molecule FRET will become an increasingly powerful tool in investigations of structure distributions in protein folding and related problems.
Materials and Methods

NMR Spectroscopy. $^1$H NOESY spectra were acquired for (Pro)$_{10}$Gly and (Pro)$_{20}$Gly in D$_2$O, and for (Pro)$_{20}$Gly in deuterated TFE (Cambridge Isotope Laboratories, Andover, MA) at 8°C to separate the water and H$^2$O signals; similar results were obtained at 20°C. All spectra were acquired on an 800-MHz Bruker spectrometer.

Single-Molecule Instrument. Single-molecule measurements were carried out with a Picocount Microtime 200 confocal fluorescence microscope (Berlin, Germany). A 470-nm pulsed diode laser (20-MHz repetition rate, 90 ps FWHM, 35 μW average power) was used to excite the donor chromophore, and donor and acceptor fluorescence were detected by single-photon avalanche photodiodes. A TimeHarp200 card was used to record the detection channel (donor, acceptor), the absolute arrival time, and fluorescence lifetime (37 ps resolution) of each photon.

Molecular Dynamics Simulations. Polyproline dynamics were investigated by using Langevin simulations of polyproline peptides of sequence Gly-(Pro)$_n$-Cys with the EEF1 implicit solvent force-field (32). The dye-linker dynamics were studied by using all-atom simulations (CHARMM27 force-field) of polyproline-dye fragments consisting of the peptide Gly-(Pro)$_3$-Cys linked either to a “donor” (attached to Cys using maleimide chemistry) or an “acceptor” (attached to Gly by peptide chemistry) dyes. Simulations were run for 20 ns in explicit TIP3P water with periodic boundary conditions (4.67-nm box size) at constant pressure (1 atm) using NAMD (33). To calculate time-dependent transfer rates $k_{ET}(t)$, composite trajectories were assembled from the three simulations by choosing random, independent time origins from each and evaluating $k_{ET}(t)$ over the subsequent portions of the simulations (differences in units of time because of friction as described above were also accounted for). Further details of simulations and efficiency calculations are in SI Text.

Note Added in Proof. A recent study by Doose et al. (34) provides evidence for internal cis prolines in aqueous solutions from short-range (sub-nanometer) fluorescence quenching by photoinduced electron transfer in ensemble and FCS experiments on polyprolines up to 10 residues in length but does not quantify either the fraction of internal cis residues or their effect on the FRET efficiency.

We thank Dennis Torchia and Attila Szabo for stimulating discussions and Wai-Ming Yau for synthesizing the polyproline peptides. This work was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health.