Quantitative Understanding of the Energy Transfer between Fluorescent Proteins Connected via Flexible Peptide Linkers

Toon H. Evers, Elisabeth M. W. M. van Dongen, Alex C. Faesen, E. W. Meijer, and Maarten Merkx*

Laboratory of Macromolecular and Organic Chemistry, Department of Biomedical Engineering, Eindhoven University of Technology, Post Office Box 513, 5600 MB Eindhoven, The Netherlands

Received June 27, 2006; Revised Manuscript Received August 29, 2006

ABSTRACT: The fusion of different protein domains via peptide linkers is a powerful, modular approach to obtain proteins with new functions. A detailed understanding of the conformational behavior of peptide linkers is important for applications such as fluorescence resonance energy transfer (FRET)-based sensor proteins and multidomain proteins involved in multivalent interactions. To investigate the conformational behavior of flexible glycine- and serine-containing peptide linkers, we constructed a series of fusion proteins of enhanced cyan and yellow fluorescent proteins (ECFP−linker−EYFP) in which the linker length was systematically varied by incorporating between 1 and 9 GGSGGS repeats. As expected, both steady-state and time-resolved fluorescence measurements showed a decrease in energy transfer with increasing linker length. The amount of energy transfer observed in these fusion proteins can be quantitatively understood by simple models that describe the flexible linker as a worm-like chain with a persistence length of 4.5 Å or a Gaussian chain with a characteristic ratio of 2.3. The implications of our results for understanding the properties of FRET-based sensors and other fusion proteins with Gly/Ser linkers are discussed.

The fusion of protein domains via flexible peptide linkers is a powerful approach to rationally design proteins with new functions. Examples of this approach include green fluorescent protein (GFP)[1] fusion proteins used in cellular localization studies (*1*), new antibody types such as multivalent antibodies and single-chain antibodies (*2*−*7*), artificial restriction enzymes consisting of zinc-finger and nuclease domains (*8*, *9*), and genetically encoded fluorescent sensor proteins consisting of two fluorescent protein domains and a ligand-binding domain (*10*). In some cases, it is sufficient if the peptide linker spatially separates the two proteins and does not interfere with the functioning and folding of both protein domains. However, for certain applications, more precise control of the spatial separation between the protein domains and the flexibility of the peptide linker are essential for proper functioning of the fusion protein.

Fluorescence resonance energy transfer (FRET) between donor and acceptor fluorescent protein domains is an effective way to translate conformational changes in a sensor domain into a ratiometric fluorescent signal. Following the pioneering work of Tsien and co-workers on the calcium sensor Cameleon (*11*, *12*), similar sensor proteins have been developed that allow for the imaging of a variety of other small molecules and cell-signaling events (*10*, *13*−*17*). In many FRET sensors, fluorescent proteins are fused to two protein domains that interact in the presence or absence of an analyte and are connected via a flexible peptide linker (*18*−*21*). Other FRET sensors use a flexible polypeptide that adopts a more rigid conformation upon binding the analyte (*22*−*24*) or contain linker peptides with protease-specific sequences (*25*−*28*). Because the energy-transfer efficiency is highly distance-dependent, knowledge of the conformational behavior of the flexible peptide linkers that separate the donor and acceptor would allow the prediction of the change in energy transfer and thus aid in the rational engineering of these biosensors (*29*). Information about the distance distribution between protein domains is also important for the rational design of fusion proteins in which domains interact with each other (to calculate effective local concentrations) (*30*) and for fusion proteins that interact with a target via multiple domains (*31*). Examples of the latter include multivalent antibody formats (*3*−*5*) and multivalent DNA-binding proteins (*32*, *33*). In this case, understanding the conformational behavior of peptide linkers is required to calculate the optimal linker length to match the spacing between binding sites on the target (*34*).

Despite their importance in determining the functional properties of fusion proteins, relatively little attention has been given to the design of peptide linkers. Only a few studies have experimentally tested the effects of linker length and linker constitution on the properties of the fusion proteins (*29*, *32*, *33*). In many cases, peptide sequences consisting of flexible and hydrophilic residues (arbitrary repeats of glycine and serine residues) are used because they are assumed to form a random coil and do not interact with (the folding of) the protein domains (*6*, *35*, *36*). Although various models have been developed to describe the conformational behavior of random coils in proteins and peptides (*37*−*43*), there is little experimental verification of their predictions for peptide

* To whom correspondence should be addressed. Telephone: (+31) 40-2474728. Fax: (+31) 40-2451036. E-mail: m.merkx@tue.nl.

[1] Abbreviations: CBD, chitin-binding domain; CLY*x*, ECFP−linker−EYFP protein with a linker containing *x* GGSGGS repeats; ECFP, enhanced cyan fluorescent protein; EDTA, ethylenediaminetetraacetic acid; EYFP, enhanced yellow fluorescent protein; FRET, fluorescence resonance energy transfer; GFP, green fluorescent protein; WLC, worm-like chain.

10.1021/bi061288t CCC: $33.50
Published on Web 10/12/2006

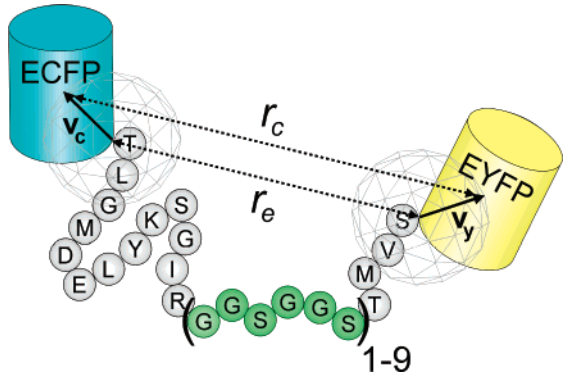

FIGURE 1: Schematic representation of our model to predict the energy transfer between ECFP and EYFP. The residues that are assumed to be part of the peptide linker are shown as well as the spherical distribution of orientations that are possible for the vector between the end of the linker and the chromophore ($\mathbf{v_c}$ and $\mathbf{v_y}$). The end to end distance of the linker (r_e) is translated into a chromophore−chromophore distance (r_c) by averaging over all possible orientations of $\mathbf{v_c}$ and $\mathbf{v_y}$.

linkers between folded protein domains (*34*). We have constructed a series of fusion proteins consisting of enhanced cyan fluorescent protein (ECFP) and enhanced yellow fluorescent protein (EYFP), linked by a flexible peptide containing 1−9 GGSGGS repeats (Figure 1). The energy transfer between ECFP and EYFP was used to determine the effect of the linker length on the distance distribution between both protein domains. We show that the energy-transfer efficiency can be quantitatively understood by describing the flexible linkers as random coils using either a worm-like chain (WLC) model or a Gaussian chain model. The implications of our results for understanding the properties of FRET-based sensors and other fusion proteins with Gly/Ser linkers are discussed.

EXPERIMENTAL PROCEDURES

Construction of pTXB1-ECFP, pTXB1-EYFP, and pET28-CLY9 Expression Vectors. All primers were purchased from MWG-Biotech AG. The vectors pECFP-C1 and pEYFP-N1 (Clontech) were used as a source of the ECFP and EYFP genes. To construct the expression vectors for ECFP and EYFP, the genes encoding ECFP and EYFP were amplified by polymerase chain reaction (PCR) using a 5′-GGTGGT-CATATGGTGAGCAAGGGCGAG-3′ forward and 5′-GGTGGTGAATTCCTTGTACAGCTCGTCCATGC-3′ reverse primer, hereby introducing an *Nde* I site directly upstream and an *Eco*R I site directly downstream from the amplified gene. The amplicons were inserted between the *Nde* I and *Eco*R I sites of the pTXB1 vector of the IMPACT system (New England Biolabs), resulting in the plasmids pTXB1-ECFP and pTXB1-EYFP. These plasmids encode fusion proteins of ECFP and EYFP with a chitin-binding domain (CBD) and an intervening intein sequence, ECFP−intein−CBD and EYFP−intein−CBD. The plasmid-encoding CLY9 was constructed from a plasmid available in our lab, pET28-CA-9-WY (E. M. W. M. van Dongen et al., manuscript in preparation). This plasmid contained the sequences of ECFP and EYFP with an intervening sequence, cloned between the *Nde* I and *Not* I sites of pET-28a(+) (Novagen). The DNA encoding the $(GGSGGS)_9$ linker, flanked by an *Eco*R I and a *Nco* I site, was purchased from GenScript Corp. The linker was inserted between the *Eco*R I and *Nco* I sites of pET28-CA-9-WY, thereby removing the intervening sequence and resulting in pET28-CLY9 (Figure 1 and Figures S1 and S2 in the Supporting Information).

Partial Digestion and Religation. A total of 2 μg of pET28-CLY9 DNA was digested for 1 h at 37 °C with 2−8 units of *Bam*H I endonuclease (New England Biolabs) in the supplied buffer in a total volume of 12 μL. Linear DNA was separated from nondigested DNA on a 1% (w/v) agarose gel and extracted in 30 μL of MilliQ water using the QIAquick gel extraction kit (Qiagen). A total of 8 μL of the linear DNA mixture was ligated overnight at 14 °C with 4 units of T4 DNA ligase (Invitrogen). The ligation mixture was used to transform *Escherichia coli* NovaBlue cells (Novagen), and single colonies were screened for the length of the linker by PCR. Each colony was boiled for 10 min in 30 μL of MilliQ water and centrifuged at 13 000 rpm. The supernatant was used as a template for a PCR reaction using HotStart polymerase (Qiagen) and the following primers: 5′-CGTCGCCGTCCAGCTCGACCAG-3′ and 5′-CATGGTC-CTGCTGGAGTTCGTG-3′. The amplicons, ranging from 163 to 307 bp, were analyzed on a 3% (w/v) agarose gel. Plasmid DNA was obtained for colonies displaying different linker lengths. DNA sequencing confirmed correct digestion and religation and identified plasmids for all different linker lengths (pET28-CLY*x*, where $x = 1-9$).

Expression and Purification of CLYx Proteins. LB medium used for culturing cells was supplemented with 30 μg/mL kanamycin to select for cells containing the pET28-CLY*x* plasmids. *E. coli* BL21(DE3) competent cells (Novagen) were transformed with the pET28-CLY*x* plasmids and plated on LB agar plates. Single colonies were used to inoculate 5 mL of LB medium and were grown overnight at 37 °C/250 rpm. Overnight cultures were diluted 1:100 in LB medium and grown at 37 °C/250 rpm until an OD_{600} of 0.6 was reached. Protein expression was induced by the addition of 0.1 mM isopropyl-β-D-thiogalactopyranoside. Cells were then grown overnight at 25 °C/250 rpm and harvested by centrifugation at 10000*g* and 4 °C for 10 min. The soluble protein fraction was isolated by resuspending the cells in BugBuster reagent (Novagen) following the instructions of the manufacturer. CLY*x* proteins were purified with nickel-affinity chromatography using His-bind resin (Novagen) according to the supplied instructions. After elution of the CLY*x* protein using 1 M imidazole, the buffer was exchanged for 50 mM Tris/HCl and 100 mM NaCl (pH 7.5) using Amicon UltraFree-4 centrifugal filter units (Millipore, 10 kDa cutoff). Aliquots of ~100 μM CLY*x* were frozen in liquid nitrogen and stored at −80 °C.

Expression and Purification of ECFP and EYFP. Production and isolation of ECFP and EYFP were performed essentially as described for CLY*x*, except that LB medium was supplemented with 100 μg/mL ampicillin to select for cells containing the pTXB1-ECFP or pTXB1-EYFP plasmids. Protein expression was induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside, and protein production was performed overnight at 15 °C/250 rpm. ECFP−intein−CBD and EYFP−intein−CBD were purified by applying the supernatant to a chitin beads column (New England Biolabs), equilibrated with 10 column volumes of column buffer [20 mM sodium phosphate and 500 mM NaCl (pH 6.0)]. The loaded column was washed with 10 volumes of column

buffer and subsequently with 2 column volumes of cleavage buffer (column buffer supplemented with 100 mM sodium 2-mercaptoethanesulphonate). After overnight incubation at room temperature, the cleaved protein was eluted from the column with cleavage buffer. To increase the yield of cleaved protein, the incubation and elution steps were repeated. The cleavage buffer was exchanged for 20 mM Tris/HCl and 150 mM NaCl (pH 7.8) using Amicon UltraFree-4 centrifugal filter units (Millipore, 10 kDa cutoff), and aliquots of ~100 μM ECFP and EYFP were frozen in liquid nitrogen and stored at −80 °C.

Spectroscopy. UV−vis spectra were recorded on a Shimadzu Multispec 1501 photodiode array spectrometer. Protein concentrations were determined using molar extinction coefficients of 84 000 M^{-1} cm^{-1} at 514 nm for EYFP and CLY*x* (*44*) and 29 000 M^{-1} cm^{-1} at 434 nm for ECFP (*45*). Because we occasionally observed some protein aggregation upon long-term storage of the fusion proteins, 10 μM protein solutions were briefly treated with 4 M urea. After two dialysis steps against 250 volumes of 50 mM Tris/HCl and 100 mM NaCl (pH 8.0) using 200 μL dialysis buttons (Hampton research) fitted with a 6−8 kDa membrane (Spectropore), protein solutions were diluted to ~0.2 μM in 50 mM Tris/HCl, 100 mM NaCl, 20 μM ethylenediaminetetraacetic acid (EDTA), and 10% glycerol (pH 8.0). Fluorescence spectra were recorded using an Edinburgh Instruments FS920 double-monochromator spectrometer with a Peltier-cooled photomultiplier. All spectra were corrected for wavelength-dependent efficiency of the instrument. Emission spectra were recorded using 420 nm excitation and corrected for the protein concentration. The Förster distance R_0 of ECFP and EYFP (48 Å) was calculated according to

$$R_0^6 = \frac{9000(\ln 10)\kappa^2 Q_D}{128\pi^5 N n^4} J(\lambda) \qquad (1)$$

using the values of 1.4 for n and 0.4 for Q_D (*46*). The overlap integral $J(\lambda)$ was calculated using the normalized ECFP emission spectrum and the EYFP excitation spectrum, normalized to 84 000 M^{-1} cm^{-1} at 514 nm. For κ^2, a value of $^2/_3$ was used, i.e., assuming dynamic averaging of the relative orientations of the ECFP emission and EYFP absorption dipole moments. Random but static orientations of the donor and acceptor would imply the use of $\kappa^2 = 0.476$ (*47*), but anisotropy measurements indicated that our system is not truly static. Furthermore, a thorough treatment of κ^2 should include the fact that κ^2 is correlated to the relative orientations of the protein domains in the calculation of $\langle E \rangle$. Because of uncertainties in the exact orientations of the transition dipoles and contributions of dynamic averaging, such a treatment is difficult to implement and is beyond the scope of our simple models. Because of this uncertainty and the fact that our anisotropy measurements indicated a certain degree of dynamic averaging, the use of $\kappa^2 = {}^2/_3$ seems reasonable. Energy-transfer efficiencies were calculated according to eq 2, in which the enhanced acceptor fluorescence F_{AD} was taken from the emission spectra at 528 nm after the subtraction of the donor contribution by deconvolution of the emission spectra using the emission spectra of ECFP and EYFP. Emission anisotropy measurements of EYFP and CLY*x* were performed using vertically polarized excitation light of 514 nm and recording the emission from 527 to 529 nm with 0.05 nm steps through a horizontally and vertically oriented polarizer. For each measurement, the G factor was measured accordingly using horizontally polarized excitation light and anisotropy values were calculated by averaging the anisotropy values of each wavelength step. Time-resolved fluorescence decays were recorded using an Edinburgh Instruments time-correlated single-photon counting Lifespec-PS spectrometer, consisting of a 400 nm picosecond laser (PicoQuant PDL 800B) operated at 2.5 MHz, an emission monochromator, and a Peltier-cooled Hamamatsu microchannel plate photomultiplier (R3809U-50). Decays were recorded at an emission wavelength of 475 nm in 4096 channels, divided over a 50 ns window with a 4 ns delay until 10 000 counts in the peak channel were obtained.

Modeling. Equations 5 and 6 describe the distribution functions $P_G(r_e)$ and $P_{WLC}(r_e)$ for the end to end distance r_e according to the Gaussian chain and WLC models, respectively. Linker residues were defined as depicted in Figure 1, with r_e being the distance between the C_α atoms of ECFP Ile229 and EYFP Lys3 (conventional residue numbering). Values for $\langle r_e \rangle$ were calculated by the integration of $P_G(r_e)*r_e$ or $P_{WLC}(r_e)*r_e$. To translate the end to end distance into the distance between the ECFP and EYFP chromophores, we calculated the average distance between the chromophores $\langle r_c \rangle$ for each end to end distance. The vectors, connecting the end of the linker and the chromophore of the linked protein, $\mathbf{v_c}$ and $\mathbf{v_y}$, with $|\mathbf{v_c}| = 20$ Å and $|\mathbf{v_y}| = 24$ Å (determined from an alignment of several GFP crystal structures from the RCSB Protein Data Bank, including 1CV7 and 1YFP), were both integrated over a sphere, as depicted in Figure 1. Using MATLAB 7.1 (The MathWorks), the spheres were created by starting with an octahedron and bisecting each edge. The edges of the created triangles were bisected again, and each vertex was moved to the surface of a sphere with a radius of $|\mathbf{v_c}|$ or $|\mathbf{v_y}|$. These approximations resulted in 66 points, nearly uniformly distributed over a sphere. Finer sampling of the possible conformations did not influence the results of the further calculations. The centers of the spheres were separated at r_e, and for each orientation of $\mathbf{v_c}$, r_c and the corresponding energy-transfer efficiency E were calculated for all orientations of $\mathbf{v_y}$. These 4356 values for r_c and E were averaged, resulting in a value for $\langle r_c \rangle$ and $\langle E \rangle$ for each r_e (Figure 4B). To calculate the ensemble average energy-transfer efficiency $\langle E \rangle_{ensemble}$ for each CLY*x*, the probabilities of each r_e were multiplied by their corresponding $\langle E \rangle$, creating the curves displayed in Figure 4C and Figure S3C in the Supporting Information. The integration of these curves yielded $\langle E \rangle_{ensemble}$. Values for $\langle r_c \rangle_{ensemble}$ (Figure 5 and Table 1) were calculated in the same manner by multiplying the probabilities of each r_e with the corresponding $\langle r_c \rangle$ and integrating the resulting curves.

RESULTS

Construction of ECFP−Linker−EYFP Proteins Using a Partial Digestion and Religation Strategy. A series of $(His)_6$-ECFP−linker−EYFP fusion proteins was constructed in which the linker length was systematically varied by incorporating between 1 and 9 GGSGGS repeats. These fusion proteins are referred to as CLY*x*, where *x* indicates the number of GGSGGS repeats. A total of 13 residues are located between the last residue of ECFP and the first glycine

Table 1: Overview of Experimental Properties and Modeling Parameters for CLY1−CLY9 and EYFP

protein	linker residues	E_{obs}[a]	r_{app} (Å)[b]	$\langle r_e \rangle$ (Å)[c,d]	$\langle r_c \rangle_{ensemble}$ (Å)[c,e]	$\langle E \rangle_{ensemble}$[c,f]	anisotropy[g]
EYFP							0.331 ± 0.002
CLY1	23	0.71	41.3	26.0/26.0	39.1/39.0	0.69/0.69	0.370 ± 0.003
CLY2	29	0.65	43.5	29.1/29.1	41.2/41.2	0.65/0.65	0.352 ± 0.003
CLY3	35	0.63	44.0	31.9/32.0	43.2/43.2	0.62/0.62	0.351 ± 0.003
CLY4	41	0.59	45.4	34.4/34.6	45.1/45.2	0.59/0.58	0.350 ± 0.003
CLY5	47	0.56	46.1	36.8/37.0	46.9/47.0	0.56/0.55	0.350 ± 0.003
CLY6	53	0.52	47.5	38.9/39.3	48.6/48.8	0.53/0.52	0.343 ± 0.003
CLY7	59	0.51	47.9	41.0/41.4	50.2/50.5	0.50/0.50	0.348 ± 0.003
CLY8	65	0.48	48.8	42.8/43.3	51.6/52.0	0.48/0.47	0.341 ± 0.003
CLY9	71	0.43	50.3	44.5/45.0	53.0/53.4	0.46/0.45	0.332 ± 0.003

[a] Experimentally determined energy-transfer efficiency. [b] The apparent distance r_{app} was calculated from the observed energy-transfer efficiency E_{obs} using the Förster equation and $R_0 = 48$ Å. [c] The first number was calculated using the Gaussian chain model with $C_\infty = 2.3$, and the second number was calculated using the WLC model with $l_p = 4.5$ Å. [d] Average distance between the ends of the peptide linker. [e] Average distance between the donor and acceptor chromophores. [f] Modeled energy-transfer efficiency. [g] Anisotropy of the EYFP emission determined at 527−529 nm using linearly polarized excitation light of 514 nm.

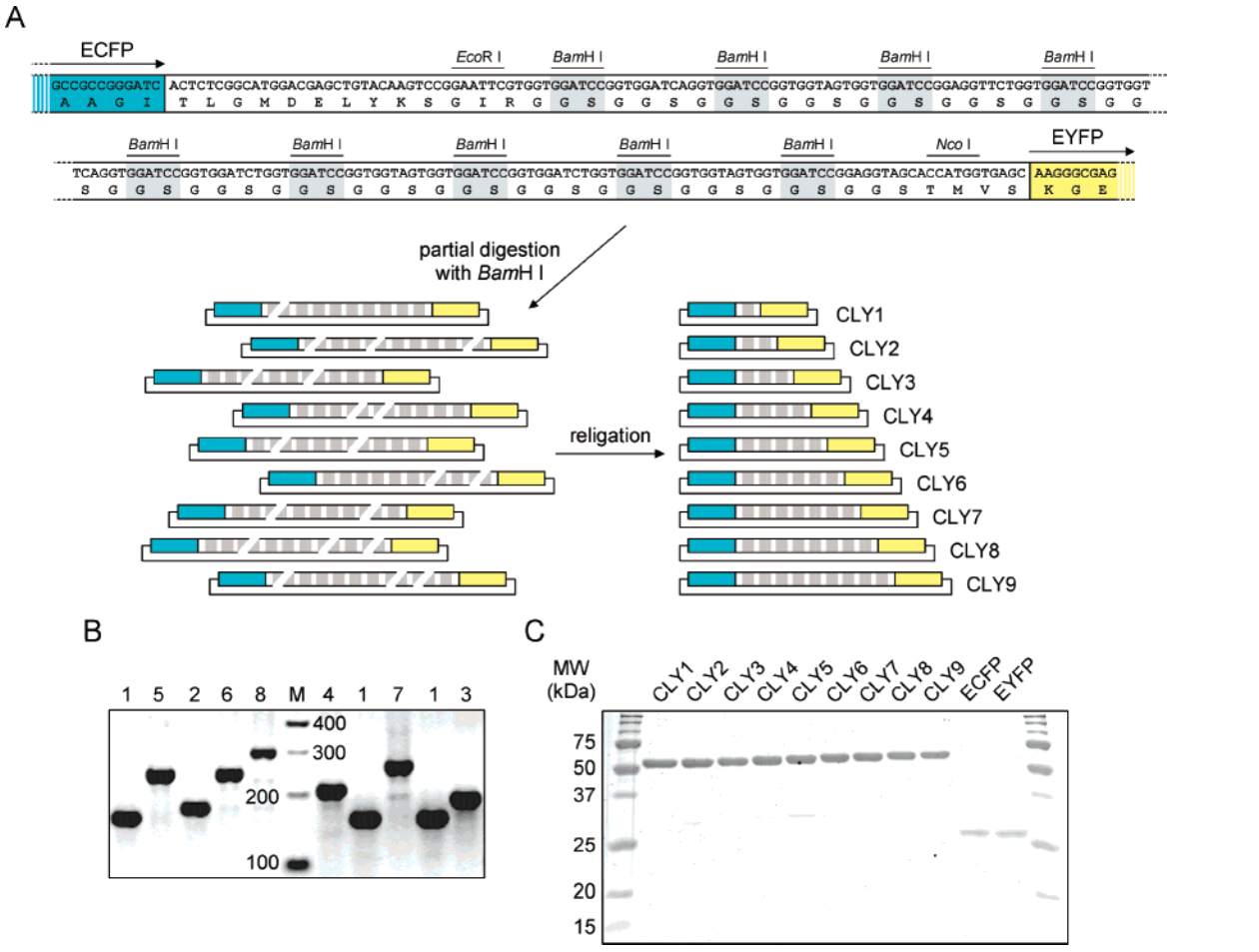

FIGURE 2: Construction of $(His)_6$-ECFP−linker−EYFP proteins. (A) Cloning strategy to obtain expression plasmids for CLY1−CLY9 using partial digestion and religation. A synthetic linker containing 9 GGSGGS repeats was cloned between ECFP and EYFP using the *Eco*R I and *Nco* I restriction sites. Partial digestion using *Bam*H I followed by religation yields a mixture of CLY*x* constructs with all possible linker lengths. (B) Transformation of the ligation mixture followed by colony PCR using primers adjacent to the linker allows for the identification of the linker length for each clone. The number of GGSGGS repeats of the clone is indicated above each lane. For comparison, a 100 base-pair marker (M) was included. (C) SDS−PAGE analysis of ECFP, EYFP, and the CLY1−CLY9 proteins obtained after expression in *E. coli* BL21(DE3) and purification.

residue of the linker, and 4 residues are located between the last serine residue of the linker and the first residue of EYFP (Figure 1). Thus, the flexible linkers cover a broad range from 23 residues for CLY1 to 71 residues for CLY9. A new and efficient cloning strategy was used to generate fusion constructs with different linker lengths that avoids the tedious cloning of each linker construct separately. This procedure, schematically depicted in Figure 2A, is based on partial digestion of the longest linker followed by religation to generate shorter linkers. A synthetic DNA sequence, containing 9 GGSGGS repeats separated by *Bam*H I restriction sites, was cloned between ECFP and EYFP in pET28. Upon partial digestion with *Bam*H I, the DNA was cleaved one or more times at different sites, resulting in a mixture of linearized plasmids with different lengths. After religation and transformation, colony PCR was used to identify clones for each

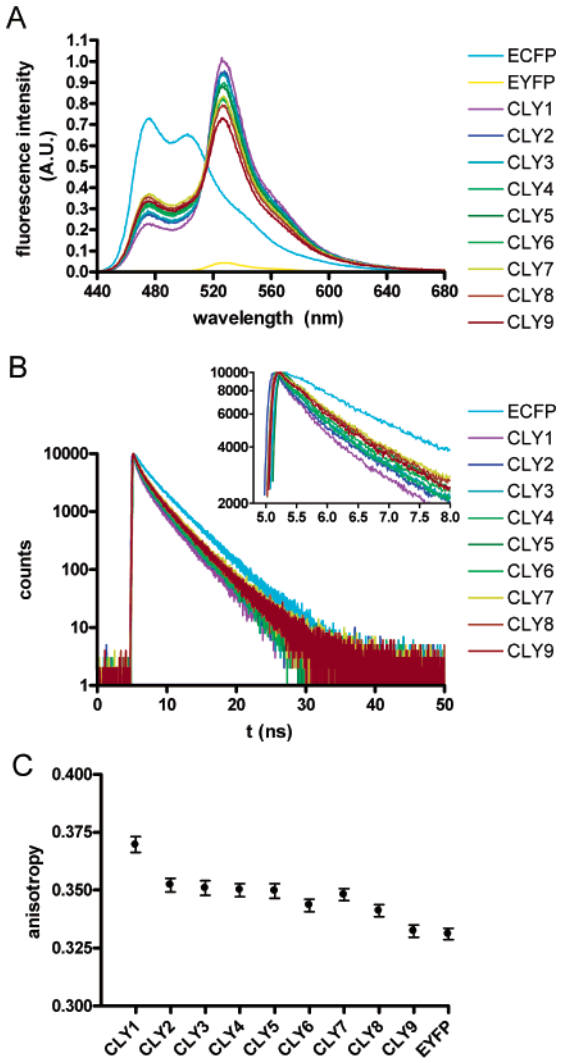

FIGURE 3: Characterization of the CLY*x* proteins using steady-state fluorescence spectroscopy, donor lifetime measurements, and fluorescence anisotropy measurements. All measurements were performed at room temperature in 50 mM Tris-HCl, 100 mM NaCl, 20 μM EDTA, and 10% (v/v) glycerol (pH 8.0) using a protein concentration of 0.2 μM. (A) Emission spectra were obtained using 420 nm excitation and corrected for the protein concentration. (B) Time-resolved fluorescence decays were recorded at 475 nm using a 400 nm picosecond laser for excitation. The inset shows a magnification of the first few nanoseconds following excitation. (C) The emission anisotropy of EYFP was determined using 514 nm vertically polarized excitation and recorded between 527 and 529 nm.

of the 9 different ECFP−linker−EYFP constructs (Figure 2B). All CLY*x* proteins were overexpressed in *E. coli* and purified in a single step by Ni-affinity chromatography (Figure 2C). Very similar absorption spectra were measured for all nine fusion proteins, showing the absorption bands typical of ECFP and EYFP in the expected 1:1 ratio (data not shown).

Influence of Linker Length on Energy Transfer. Fluorescence emission spectra of the CLY*x* proteins excited at 420 nm showed a substantial decrease in ECFP emission and an increase in EYFP emission compared to the spectra of ECFP and EYFP alone, indicating efficient energy transfer (Figure 3A). A clear trend was observed, showing less efficient energy transfer for proteins containing longer linkers. The fluorescence lifetimes of the ECFP domains in the CLY*x* proteins were also markedly decreased compared to the lifetime of ECFP alone and showed the same trend with increasing linker length (Figure 3B). The energy-transfer efficiency E_{obs} was calculated from the enhanced acceptor fluorescence according to

$$E_{obs} = \frac{\epsilon_A(\lambda_D^{ex})}{\epsilon_D(\lambda_D^{ex})}\left[\frac{F_{AD}(\lambda_A^{em})}{F_A(\lambda_A^{em})} - 1\right] \quad (2)$$

where $\epsilon_A(\lambda_D^{ex})$ and $\epsilon_D(\lambda_D^{ex})$ are the molar extinction coefficients of the acceptor and donor at the donor excitation wavelength and $F_{AD}(\lambda_A^{em})$ and $F_A(\lambda_A^{em})$ are the intensities of the acceptor in the presence and absence of the donor, respectively[2] (*46*). Values for E_{obs} varied from 0.71 for CLY1 to 0.43 for CLY9 (Table 1). Because conformational differences can be detected with the greatest sensitivity around $E_{obs} = 0.5$, the range of linker lengths present in CLY1−CLY9 was well-chosen.

Quantitative Understanding of Energy-Transfer Efficiency by Modeling Peptide Linkers as Gaussian or WLC Random Coils. The dependence of the energy-transfer efficiency E on the distance between two chromophores r_c is decribed by the Förster equation (*46*)

$$E = \frac{R_0^6}{R_0^6 + r_c^6} \quad (3)$$

The Förster distance R_0 is characteristic of a pair of chromophores. We determined an R_0 of 48 Å for the combination of ECFP and EYFP, which is in good agreement with literature values (*48*). Although a clear trend of increased energy transfer with decreasing linker length was observed, the difference in energy transfer between the shortest and longest linker appeared to be relatively small. Direct application of the Förster equation to translate E_{obs} values into the distance between the two chromophores yielded values ranging from 41 Å for CLY1 to 50 Å for CLY9 (r_{app}, Table 1), which is a surprisingly small difference considering the difference in the linker length of 48 amino acids between CLY1 and CLY9. However, because of the flexible nature of the linker, a distribution of interchromophore distances will exist in solution, instead of a single distance. Because the small distances in the ensemble contribute more to E_{obs} than the larger distances, the distances calculated from E_{obs} will not reflect the real distances between the fluorescent proteins.

To investigate whether the observed energy-transfer efficiencies can be quantitatively understood, we used two models to describe the peptide linker between the fluorescent protein domains as random-coil structures: the Gaussian chain model developed by Flory (*37*) and the Kratky−Porod WLC model (*38*). The Gaussian chain model describes the linker as a freely jointed chain of C_α atoms. According to

[2] E_{obs} is often determined using the decrease in donor emission. However, SDS−PAGE analysis of the purified proteins showed the presence of small but varying amounts of a 30 kDa protein, which is probably ECFP with a N-terminal His tag resulting from proteolytic degradation of the linker. Deconvolution of the UV−vis spectra showed that all CLY*x* protein samples contained slightly more ECFP than EYFP (data not shown). Because the presence of even small amounts of free ECFP could affect the accurate calculation of E_{obs}, it was decided to use the acceptor emission.

random-walk statistics, the mean-square end to end distance of a chain consisting of n peptide bonds is given by

$$\langle r_e^2 \rangle_G = C_n n b_0^2 \quad (4)$$

where b_0 is the average distance between adjacent C_α atoms (3.8 Å) and C_n is the characteristic ratio, a correction factor for the nonideal behavior of real chains (*49*). C_n is the only variable in this model and depends on the chain length for small n. However, C_n adopts a constant value C_∞ for long or very flexible chains, which is dependent on the amino acids present in the linker (*37*, *49*−*51*). For poly(Gly), C_∞ is 2.16, but C_∞ assumes values of 9.27 for poly(Ala) or 116 for poly(Pro). For simplicity, we tried to model the linkers using a single value for C_∞. These chains have a Gaussian probability density for the displacement vector between any two residues in the chain, and the distribution function $P_G(r_e)$ for the end to end distance r_e is given by (*39*)

$$P_G(r_e) = 4\pi r_e^2 (3/2\pi\langle r_e^2 \rangle)^{3/2} \exp(-3r_e^2/2\langle r_e^2 \rangle) \quad (5)$$

The Gaussian chain model using a single value for C_n might not be able to adequately describe the behavior of the shortest linkers used here. A different approach is to model the linker using the WLC model, which describes a polypeptide as a semiflexible tube. This model has the advantage that it can also be used for stiff short polymers and has been shown to accurately describe the distribution of end to end distances of flexible loops in proteins and predict the effect of the peptide linker length on the effective local concentration of protein domains (*39*). The distribution function of a WLC with a total contour length l_c is given by (*39*, *52*)

$$P_{WLC}(r_e) = 4\pi r_e^2 (3/4\pi l_p l_c)^{3/2} \exp(-3r_e^2/4l_p l_c)(1 - 5l_p/4l_c - 2r_e^2/l_c^2 + 33r_e^4/80l_p l_c^3 + 79l_p^2/160l_c^2 + 329r_e^2 l_p/120l_c^3 - 6799r_e^4/1600l_c^4 + 3441r_e^6/2800l_p l_c^5 - 1089r_e^8/12800l_p^2 l_c^6) \quad (6)$$

The total contour length of a chain is equal to $b_0 n$, where n is the number of peptide bonds and b_0 is the average distance between adjacent C_α atoms, 3.8 Å. This model takes into account that the direction of the chain has a persistent memory at a short contour length, reflected by the persistence length l_p. The persistence length is the only adjustable parameter of the WLC model and, like C_∞, reflects the stiffness of the polypeptide chain. Figure 4A shows the end to end distribution functions calculated for the linkers present in CLY1 ($n = 24$) to CLY9 ($n = 72$) using a persistence length of 4.5 Å, a value consistent with literature values, as discussed below. The distribution functions appear Gaussian because the chains are relatively long and are nearly identical to those of the Gaussian chain model with a C_∞ of 2.3 (Figure S3A in the Supporting Information). The average end to end distance of the peptide linker $\langle r_e \rangle$ varies from 26.0 to 44.5 Å for the Gaussian chains and from 26.0 to 45.0 Å for the WLC model (Table 1). This calculation illustrates that only a modest increase in the average distance between the ends of a flexible peptide chain is expected even if the number of amino acids is tripled. However, to be able to compare the predictions of the random-coil models to our experiments, the end to end distances have to be translated into distances

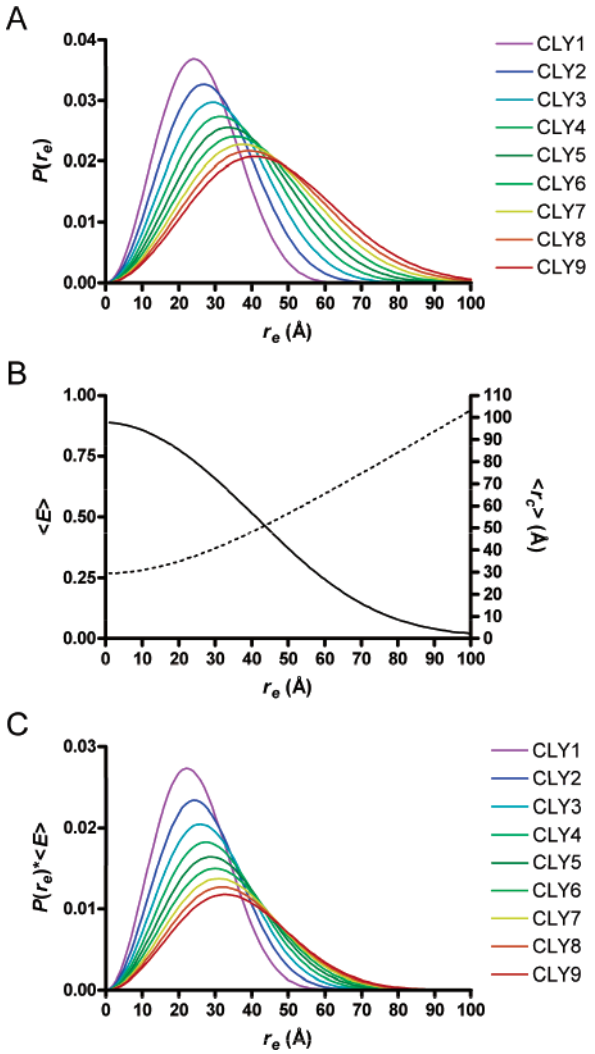

FIGURE 4: Modeling the energy-transfer efficiency for CLYx using the WLC model to describe the peptide linker. (A) Distribution functions showing the probability $P(r_e)$ for each end to end distance of the peptide linker for CLY1−CLY9, calculated using the WLC model assuming a persistence length of 4.5 Å. (B) Plot showing the average energy transfer ($\langle E \rangle$, —) and the average interchromophore distance ($\langle r_c \rangle$, - - -) as a function of r_e. (C) Contribution of each r_e to the overall energy transfer $\langle E \rangle_{ensemble}$, calculated by multiplying the probability of each r_e with the corresponding $\langle E \rangle$.

between the ECFP and EYFP chromophores, r_c. For a given end to end distance of the linker, the ECFP and EYFP protein domains can adopt many orientations; i.e., for every r_e, many different values for r_c are possible (Figure 1). To account for the different orientations of the proteins, the average distance between the chromophores $\langle r_c \rangle$ was calculated for each r_e by averaging r_c over all possible orientations of the vectors connecting the end of the linker and the chromophore, $\mathbf{v_c}$ and $\mathbf{v_y}$ (Figures 1 and 4B). It is important to realize that the average energy transfer $\langle E \rangle$ for each end to end distance cannot simply be calculated from $\langle r_c \rangle$, however, because orientations of the protein domains in which the chromophores are closer together contribute more to $\langle E \rangle$ than orientations in which they are further apart. Therefore, $\langle E \rangle$ was calculated in the same manner as $\langle r_c \rangle$ by calculating a value for E for every r_c using eq 3 and averaging over all possible orientations of $\mathbf{v_c}$ and $\mathbf{v_y}$. This yielded a value for $\langle E \rangle$ for every r_e, resulting in the curve shown in Figure 4B. Note that at $r_e = 0$ the average distance between the

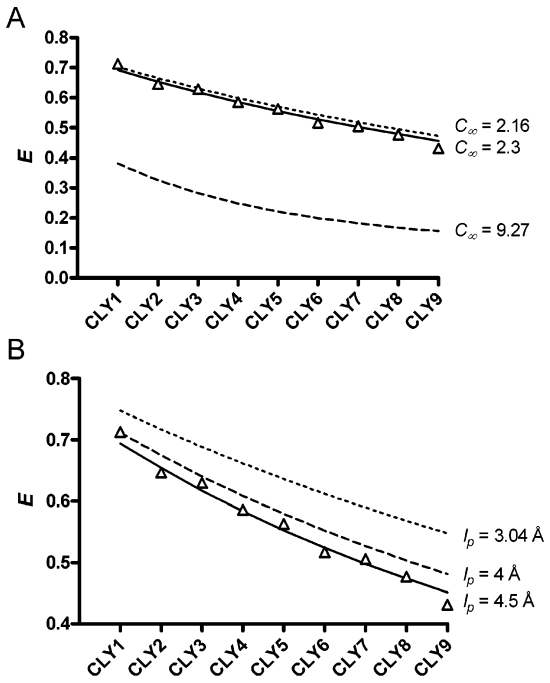

FIGURE 5: Comparison of the experimentally determined energy-transfer efficiencies (E_{obs}, △) with values predicted by the random-coil models. (A) $\langle E \rangle_{ensemble}$ predicted by the Gaussian chain model using values for C_∞ of 2.16, 2.3, and 9.27. (B) $\langle E \rangle_{ensemble}$ predicted by the WLC model using persistence lengths of 3.04, 4, and 4.5 Å.

chromophores is still ∼30 Å and $\langle E \rangle \sim 0.90$, which is due to the fact that the fluorophores are located at the center of the β barrel at some distance from the linker ends. When the $\langle E \rangle$ curve was multiplied with the distribution functions of the WLC model in Figure 4A and of the Gaussian chain model in Figure S3A in the Supporting Information, the contributions of each r_e to the ensemble energy transfer $\langle E \rangle_{ensemble}$ could be calculated for each linker (Figure 4C and Figure S3C in the Supporting Information). Finally, integration of these curves yielded the predicted energy-transfer efficiency $\langle E \rangle_{ensemble}$ for each of the CLY proteins (Table 1 and Figure 5). The calculated values of $\langle E \rangle_{ensemble}$ agree surprisingly well with the experimentally determined energy-transfer efficiencies over the entire range of linker lengths. For a comparison, the predicted values for $\langle E \rangle_{ensemble}$ using different values for l_p or C_∞ are also shown.

Effect of the Linker Length on the Rotational Diffusion of Protein Domains. The success of both random-coil models in predicting the energy-transfer efficiency between ECFP and EYFP suggests that the GGSGGS-peptide linker indeed assumes a random-coil structure. In our model, the two protein domains do not interact and a random distribution of orientations and distances exists between ECFP and EYFP. To investigate whether linking the fluorescent proteins with a flexible linker had an effect on the rotational freedom of the protein domains, we compared the fluorescence anisotropy of the EYFP domains of all CLY*x* proteins with the anisotropy of EYFP (Figure 3C and Table 1). The anisotropy of 0.33 determined for EYFP is in excellent agreement with values reported in the literature (*53*, *54*). The EYFP domains of the CLY*x* proteins showed a weak trend of slower rotational diffusion with shorter linkers, with the EYFP domain of CLY9 being essentially unaffected by the linker and the rotational diffusion of the EYFP domain of CLY1 slowing down to a larger extent.

DISCUSSION

The random-coil behavior of flexible peptides and unfolded proteins has been described using a variety of theoretical models that were originally developed for synthetic polymers (*37*−*43*). These models have mainly been used to successfully predict the biophysical properties of proteins under denaturing conditions but also of peptides and protein loops under native conditions (*49*, *55*). For example, various FRET studies have probed the conformational behavior of synthetic peptides under physiological conditions (*56*−*60*). Similar studies have not been reported for flexible peptide linkers that are part of a fusion protein, however. The systematic variation of linker lengths between ECFP and EYFP reported in this work allowed for the first time a quantitative understanding of the energy transfer between fluorescent proteins as a function of linker behavior. Some earlier studies also varied the linker length or composition to some extent (*27*, *35*), but none of them systematically varied the length of flexible linkers over a wide range using identical peptide repeats. Two simple models describing the peptide linkers as random coils in solution were shown to describe the observed energy transfer satisfactory over the entire range of linker lengths. Although glycine- and serine-rich peptides are generally assumed to adopt a random-coil structure, our results provide experimental evidence that these linkers indeed behave as random coils when fused between two protein domains.

Random-Coil Model. Our calculations for the expected energy-transfer efficiencies are based on a simple model using the following assumptions: (1) the peptide linker has a random-coil conformation; (2) the distribution of distances between the ends of the flexible peptide linkers can be described using a Gaussian or WLC model; (3) the protein domains can rotate freely around the end of the linker; (4) the energy transfer at a given end to end distance can be calculated by averaging the contribution of all possible orientations of ECFP and EYFP; and (*5*) the energy-transfer efficiency for the ensemble can be calculated by multiplying the energy transfer for each end to end distance by the probability of occurrence of this distance. We realize that our model contains some simplifications. When calculating the distribution of the distances between the chromophores, we did not take steric hindrance between the two proteins into account. Although steric hindrance is expected for small values of r_e, many different conformations are still possible even at these small separations. Furthermore, even if we would account for disallowed conformations for small values for r_e, the difference in $\langle E \rangle$ for this r_e would be small as a result of the relative insensitivity of $\langle E \rangle$ at distances shorter than 0.7 R_0. Also the probabilities of these small values for r_e are low, minimizing their effect on the overall energy transfer $\langle E \rangle_{ensemble}$. Nonetheless, when using shorter linkers than used here, the effects will become more important and molecular dynamics or Monte Carlo simulations may be necessary to accurately describe the interchromophore distance.

It should be noted that, except for l_p and C_∞, all parameters were determined independently of the model; i.e., they were

not adjusted to obtain a better fit. The ends of the peptide linker were defined as the first or last residue with observable electron density in the X-ray structures of GFPs. The distance between the end of the linker and the middle of the chromophore that was used in calculating r_c was also determined from X-ray structure analysis. The Förster distance was calculated to be 48 Å based on our own measurements of the spectra of ECFP and EYFP and agreed well with the Förster distance of 49 Å previously reported (*48*). We initially chose to model the linkers as WLCs because the WLC model has been successful in describing the behavior of random coils in nondenaturing conditions (*39*). Because the linkers used in this study covered a broad range in length and could all be accurately described with the same persistence length of 4.5 Å, the choice for the WLC model seemed to be valid. Furthermore, this persistence length agrees with other studies on flexible peptide chains. For example, the mechanic unfolding of single proteins with AFM could be understood by modeling the unfolded proteins as WLCs with persistence lengths ranging from 3.3 to 4.2 Å, although a few studies reported values of 6.5−8 Å (*39*). Furthermore, it was recently shown that the random-coil loops in structurally characterized proteins can also be very well-described by a WLC (*39*, *61*). For the ends of these loops that were directly connected to folded parts of the same domain, a persistence length of 3.04 Å was determined, whereas distances between residues within the flexible peptide linkers were accurately described using a persistence length of 4 Å. Although the ends of the linkers in our fusion proteins are also connected to a stable secondary structure, these ends have the same conformational freedom as the intrachain residues through the rotation of the protein domains. Therefore, it is expected that an l_p of 4 Å would better describe our linkers than an l_p of 3.04 Å, as shown in Figure 5B. The slightly higher value of 4.5 Å for the best fit could indicate more pronounced excluded volume effects of the folded protein domains.

The near-Gaussian distribution functions of the WLC model prompted us to also model the linkers using the simpler Gaussian chain model. In its simplest form, the Gaussian chain model also contains only one real variable, the characteristic ratio, which corrects for the stiffness of real chains. Figure 5A shows that the decrease in energy-transfer efficiency can also be fit by modeling the linkers as Gaussian chains with a C_∞ of 2.3. In fact, C_∞ and l_p are related. In the limit of long, flexible chains ($l_c \gg l_p$), eq 6 reduces to eq 5 with $l_p = C_\infty b_0/2$, which is exactly what we observe here. The value of 2.3 for C_∞ is very close to the value of C_∞ for poly(Gly), 2.16 (*51*), indicating that the high content of glycine residues in our linker governs the behavior of the linker. This result is consistent with theoretical calculations that predicted that a small percentage of glycine residues will have a large effect on the chain dimensions of random-coil polypeptides (*62*). Although C_n is dependent on the chain length, C_n is equal to C_∞ for poly(Gly) chains with $n > 10$ (*50*). Because our linkers range from $n = 24$ to 72 and are nearly as flexible as poly(Gly) chains, this justifies modeling all linkers using the same value. From Figure 5A, it is clear that these $(GGSGGS)_x$ linkers are much more flexible than poly(Ala) linkers ($C_\infty = 9.27$).

Because GFPs have been reported to associate at high protein concentrations (*63*), the observed energy-transfer efficiencies might also be explained by a model in which the linker promotes this association by increasing the local concentration. The good agreement between the energy-transfer efficiency predicted for a random coil and the experimentally determined energy transfer argues against the presence of substantial amounts of ECFP−EYFP complex, however. The fluorescence anisotropy measurements are also in agreement with the random-coil model. The anisotropies found for CLY9 and EYFP are identical and consistent with the rotational diffusion that is expected for a 30 kDa protein. If the linker would induce dimerization between ECFP and EYFP in CLY9, the increase in size should result in a higher anisotropy. Although slightly higher anisotropies were observed for the fusion proteins with shorter linker lengths, a more plausible explanation is that the rotational diffusion of EYFP is affected by the presence of the linker and the ECFP domain. Even with a very flexible linker, the dynamics of changes between different conformations may be slower, while the number of possible conformations in the ensemble is unaffected.

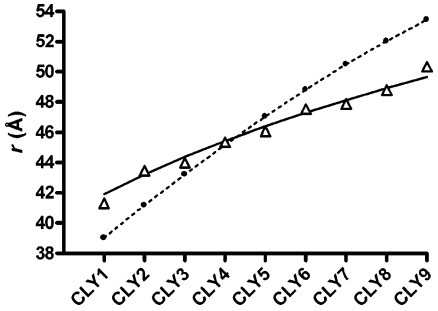

FIGURE 6: Comparison of the "true" average interchromophore distance obtained from modeling ($\langle r_c \rangle_{ensemble}$, - - -) with the apparent interchromophore distance obtained by direct application of the Förster equation to the observed energy-transfer efficiencies E_{obs} (r_{app}, △) or the modeled energy-transfer efficiencies $\langle E \rangle_{ensemble}$ (—). Both $\langle r_c \rangle_{ensemble}$ and $\langle E \rangle_{ensemble}$ were calculated using the WLC model with an l_p of 4.5 Å.

Implications for FRET Sensor Design. Our analysis of the linker length dependency of energy transfer between ECFP and EYFP has important implications for the design of FRET-based sensor proteins. First, it is often not realized that long flexible peptide linkers can still give rise to high energy-transfer efficiencies. Although the longest peptide linker used in this study can span a distance of 252 Å in an extended conformation, the most abundant end to end distance is only 41 Å, showing that random-coil structures are surprisingly compact. Another important lesson is that the energy transfer between fluorescent proteins separated by a flexible linker actually reflects a broad distribution of distances and not a single distance. Therefore, one should be careful when calculating a single distance or even an average distance from a single energy-transfer efficiency. To illustrate this, we have calculated an "apparent distance" r_{app} from the observed energy-transfer efficiencies by direct application of the Förster equation and plotted these together with the true calculated average distance between the chromophores, $\langle r_c \rangle_{ensemble}$ (Figure 6 and Table 1). Because of the nonlinear dependency of energy transfer on the distance, a calculated distance can be easily underestimated for long linkers or overestimated for short linkers. For example, the relatively high FRET efficiencies previously observed for EBFP and EGFP separated by linkers containing

3 or 4 GGGGS motifs and the small difference in FRET for both linkers were interpreted as evidence for the backfolding of the peptide linker onto the protein domains (*35*, *64*). This study shows that the high energy-transfer efficiency and the small difference between the two linkers can be explained assuming a fully random-coil peptide linker without backfolding. The metal-binding-induced folding of metallothionein has been used in FRET sensors using either synthetic (*65*, *66*) or protein fluorescent groups (*23*). Again, the relatively small difference between the unbound and bound states was used to argue that the unbound state contained residual structure (*65*, *66*). This is not necessarily true, however. First, the average end to end distance of a random-coil peptide can be very close to the distance between the same residues in a folded structure. Moreover, the distribution of distances present for the unfolded structure includes conformations with a very short donor−acceptor distance, which could contribute relatively strongly to the overall energy transfer of the ensemble of conformations.

In conclusion, we have shown that the conformational distribution of two protein domains connected via flexible Gly/Ser linkers can be described by modeling the linkers as Gaussian chains with a characteristic ratio of 2.3 or as WLCs with a persistence length of 4.5 Å. As described above, this knowledge could aid in the engineering of FRET-based sensor proteins by predicting changes in energy-transfer efficiency. Because the same models can be used to calculate effective local concentrations of protein domains separated by flexible peptide linkers, our findings could also be applied in the design of multivalent proteins, e.g., in antibody engineering, and the calculation of intramolecular domain interactions within multidomain proteins (*39*). Experimental verification of our findings for other fusion proteins could benefit from the new cloning method introduced here that facilitates systematic variation of linker lengths.

ACKNOWLEDGMENT

The authors thank M. Elschot and I. Kaashoek (Spinoza Research Institute, Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands) for production and purification of the CLY*x* proteins, M. Rensen for the cloning and production of the ECFP and EYFP proteins, and K. Pieterse for fruitful discussions.

SUPPORTING INFORMATION AVAILABLE

Three figures showing the vector map of pET28-CLY9, the open-reading frame encoding CLY9, the end to end distance distribution functions for the Gaussian chain model $P_G(r_e)$, and the energy-transfer efficiency distribution functions for the Gaussian chain model $P_G(r_e)*\langle E\rangle$. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. O'Rourke, N. A., Meyer, T., and Chandy, G. (2005) Protein localization studies in the age of "Omics", *Curr. Opin. Chem. Biol. 9*, 82−87.
2. Presta, L. (2003) Antibody engineering for therapeutics, *Curr. Opin. Struct. Biol. 13*, 519−525.
3. Deyev, S. M., Waibel, R., Lebedenko, E. N., Schubiger, A. P., and Plückthun, A. (2003) Design of multivalent complexes using the barnase-barstar module, *Nat. Biotechnol. 21*, 1486−1492.
4. Plückthun, A., and Pack, P. (1997) New protein engineering approaches to multivalent and bispecific antibody fragments, *Immunotechnology 3*, 83−105.
5. Todorovska, A., Roovers, R. C., Dolezal, O., Kortt, A. A., Hoogenboom, H. R., and Hudson, P. J. (2001) Design and application of diabodies, triabodies and tetrabodies for cancer targeting, *J. Immunol. Methods 248*, 47−66.
6. Neri, D., Momo, M., Prospero, T., and Winter, G. (1995) High-affinity antigen binding by chelating recombinant antibodies (CRAbs), *J. Mol. Biol. 246*, 367−373.
7. Zhang, J., Tanha, J., Hirama, T., Khieu, N. H., To, R., Tong-Sevinc, H., Stone, E., Brisson, J.-R., and MacKenzie, C. R. (2004) Pentamerization of single-domain antibodies from phage libraries: A novel strategy for the rapid generation of high-avidity antibody reagents, *J. Mol. Biol. 335*, 49−56.
8. Kim, Y.-G., Cha, J., and Chandrasegaran, S. (1996) Hybrid restriction enzymes: Zinc finger fusions to *Fok* I cleavage domain, *Proc. Natl. Acad. Sci. U.S.A. 93*, 1156−1160.
9. Kim, J.-S., and Pabo, C. O. (1998) Getting a handhold on DNA: Design of poly-zinc finger proteins with femtomolar dissociation constants, *Proc. Natl. Acad. Sci. U.S.A. 95*, 2812−2817.
10. Miyawaki, A. (2003) Visualization of the spatial and temporal dynamics of intracellular signaling, *Dev. Cell 4*, 295−305.
11. 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.
12. Miyawaki, A., Griesbeck, O., Heim, R., and Tsien, R. Y. (1999) Dynamic and quantitative Ca^{2+} measurements using improved cameleons, *Proc. Natl. Acad. Sci. U.S.A. 96*, 2135−2140.
13. Chudakov, D. M., Lukyanov, S., and Lukyanov, K. A. (2005) Fluorescent proteins as a toolkit for in vivo imaging, *Trends Biotechnol. 23*, 605−613.
14. Sato, M., Hida, N., Ozawa, T., and Umezawa, Y. (2000) Fluorescent indicators for cyclic GMP based on cyclic GMP-dependent protein kinase Iα and green fluorescent proteins, *Anal. Chem. 72*, 5918−5924.
15. Nagai, Y., Miyazaki, M., Aoki, R., Zama, T., Hagiwara, M., Inouye, S., Hirose, K., and Iino, M. (2000) A fluorescent indicator for visualizing cAMP-induced phosphorylation in vivo, *Nat. Biotechnol. 18*, 313−316.
16. Fehr, M., Frommer, W. B., and Lalonde, S. (2002) Visualization of maltose uptake in living yeast cells by fluorescent nanosensors, *Proc. Natl. Acad. Sci. U.S.A. 99*, 9846−9851.
17. DiPilato, L. M., Cheng, X., and Zhang, J. (2004) Fluorescent indicators of cAMP and Epac activation reveal differential dynamics of cAMP signaling within discrete subcellular compartments, *Proc. Natl. Acad. Sci. U.S.A. 101*, 16513−16518.
18. 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.
19. Ting, A. Y., Kain, K. H., Klemke, R. L., and Tsien, R. Y. (2001) Genetically encoded fluorescent reporters of protein tyrosine kinase activities in living cells, *Proc. Natl. Acad. Sci. U.S.A. 98*, 15003−15008.
20. Lin, C.-W., and Ting, A. Y. (2004) A genetically encoded fluorescent reporter of histone phosphorylation in living cells, *Angew. Chem., Int. Ed.* 43, 2940−2943.
21. Lin, C.-W., Jao, C. Y., and Ting, A. Y. (2004) Genetically encoded fluorescent reporters of histone methylation in living cells, *J. Am. Chem. Soc. 126*, 5982−5983.
22. Cicchetti, G., Biernacki, M., Farquharson, J., and Allen, P. G. (2004) A ratiometric expressible FRET sensor for phosphoinositides displays a signal change in highly dynamic membrane structures in fibroblasts, *Biochemistry 43*, 1939−1949.
23. Pearce, L. L., Gandley, R. E., Han, W., Wasserloos, K., Stitt, M., Kanai, A. J., McLaughlin, M. K., Pitt, B. R., and Levitan, E. S. (2000) Role of metallothionein in nitric oxide signaling as revealed by a green fluorescent fusion protein, *Proc. Natl. Acad. Sci. U.S.A. 97*, 477−482.
24. Kohn, J. E., and Plaxco, K. W. (2005) Engineering a signal transduction mechanism for protein-based biosensors, *Proc. Natl. Acad. Sci. U.S.A. 102*, 10841−10845.
25. 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.

26. Xu, X., Gerard, A. L. V., Huang, B. C. B., 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.
27. Felber, L. M., Cloutier, S. M., Jichlinski, P., Leisinger, H.-J., Deperthes, D., Kündig, C., Brossard, V., and Kishi, T. (2004) Evaluation of the CFP−substrate−YFP system for protease studies: Advantages and limitations, *BioTechniques 36*, 878−885.
28. Zhang, B. (2004) Design of FRET-based GFP probes for detection of protease inhibitors, *Biochem. Biophys. Res. Commun. 323*, 674−678.
29. Lissandron, V., Terrin, A., Collini, M., D'alfonso, L., Chirico, G., Pantano, S., and Zaccolo, M. (2005) Improvement of a FRET-based indicator for cAMP by linker design and stabilization of donor−acceptor interaction, *J. Mol. Biol. 354*, 546−555.
30. Zhou, H.-X. (2001) Single-chain versus dimeric protein folding: Thermodynamic and kinetic consequences of covalent linkage, *J. Am. Chem. Soc. 123*, 6730−6731.
31. Zhou, H.-X. (2001) The affinity-enhancing roles of flexible linkers in two-domain DNA-binding proteins, *Biochemistry 40*, 15069−15073.
32. Robinson, C. R., and Sauer, R. T. (1998) Optimizing the stability of single-chain proteins by linker length and composition mutagenesis, *Proc. Natl. Acad. Sci. U.S.A. 95*, 5929−5934.
33. van Leeuwen, H. C., Strating, M. J., Rensen, M., van der Vliet, P. C., and De Laat, W. (1997) Linker length and composition influence the flexibility of Oct-1 DNA binding, *EMBO J. 16*, 2043−2053.
34. Zhou, H.-X. (2003) Quantitative account of the enhanced affinity of two linked scFvs specific for different epitopes on the same antigen, *J. Mol. Biol. 329*, 1−8.
35. Arai, R., Ueda, H., Kitayama, A., Kamiya, N., and Nagamune, T. (2001) Design of the linkers which effectively separate domains of a bifunctional fusion protein, *Protein Eng. 14*, 529−532.
36. Alfthan, K., Takkinen, K., Sizmann, D., Soderlund, H., and Teeri, T. T. (1995) Properties of a single-chain antibody containing different linker peptides, *Protein Eng. 8*, 725−731.
37. Flory, P. J. (1969) *Statistical Mechanics of Chain Molecules*, Interscience Publishers, New York.
38. Kratky, O., and Porod, G. (1949) Röntgenuntersuchung gelöster fadenmoleküle, *Recl. Trav. Chim. Pays−Bas. 68*, 1106−1122.
39. Zhou, H.-X. (2004) Polymer models of protein stability, folding, and interactions, *Biochemistry 43*, 2141−2154.
40. Chan, H. S., and Dill, K. A. (1991) Polymer principles in protein structure and stability, *Annu. Rev. Biophys. Biophys. Chem. 20*, 447−490.
41. Dill, K. A., and Shortle, D. (1991) Denatured states of proteins, *Annu. Rev. Biochem. 60*, 795−825.
42. Fitzkee, N. C., and Rose, G. D. (2005) Sterics and solvation winnow accessible conformational space for unfolded proteins, *J. Mol. Biol. 353*, 873−887.
43. Kohn, J. E., Plaxco, K. W., Dillon, T. M., Cingel, N., Millett, I. S., Doniach, S., Zagrovic, B., Pande, V. S., Jacob, J., Dothager, R. S., Sosnick, T. R., Thiyagarajan, P., Seifert, S., Hasan, M. Z., and Ruczinski, I. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins, *Proc. Natl. Acad. Sci. U.S.A. 101*, 12491−12496.
44. Patterson, G. H., Day, R. N., and Piston, D. (2001) Fluorescent protein spectra, *J. Cell Sci. 114*, 837−838.
45. Rizzo, M. A., Springer, G. H., Granada, B., and Piston, D. W. (2004) An improved cyan fluorescent protein variant useful for FRET, *Nat. Biotechnol. 22*, 445−449.
46. Lakowicz, J. R. (1999) *Principles of Fluorescence Spectroscopy*, Kluwer Academic/Plenum Publishers, New York.
47. Maksimov, M. Z., and Rozman, I. M. (1962) On energy transfer in solid solutions, *Opt. Spectrosc. 12*, 606−609.
48. Patterson, G. H., Piston, D. W., and Barisas, B. G. (2000) Förster distances between green fluorescent protein pairs, *Anal. Biochem. 284*, 438−440.
49. Wang, L., Rivera, E. V., Benavides-Garcia, M. G., and Nall, B. T. (2005) Loop entropy and cytochrome *c* stability, *J. Mol. Biol. 353*, 719−729.
50. Schimmel, P. R., and Flory, P. J. (1967) Conformational energy and configurational statistics of poly-L-proline, *Proc. Natl. Acad. Sci. U.S.A. 58*, 52−59.
51. Brant, D. A., Miller, W. G., and Flory, P. J. (1967) Conformational energy estimates for statistically coiling polypeptide chains, *J. Mol. Biol. 23*, 47−65.
52. Yamakawa, H., and Stockmayer, W. H. (1972) Statistical mechanics of wormlike chains. II. Excluded volume effects, *J. Chem. Phys. 57*, 2843−2854.
53. Schaffer, J., Volkmer, A., Eggeling, C., Subramaniam, V., Striker, G., and Seidel, C. A. M. (1999) Identification of single molecules in aqueous solution by time-resolved fluorescence anisotropy, *J. Phys. Chem. A 103*, 331−336.
54. Borst, J. W., Hink, M. A., Visser, A. J. W. G., and van Hoek, A. (2005) Effects of refractive index and viscosity on fluorescence and anisotropy decays of enhanced cyan and yellow fluorescent proteins, *J. Fluoresc. 15*, 153−160.
55. Nagi, A. D., and Regan, L. (1997) An inverse correlation between loop length and stability in a four-helix-bundle protein, *Folding Des. 2*, 67−75.
56. Haas, E., Wilchek, M., Katchalski-Katzir, E., and Steinberg, I. Z. (1975) Distribution of end-to-end distances of oligopeptides in solution as estimated by energy transfer, *Proc. Natl. Acad. Sci. U.S.A. 72*, 1807−1811.
57. De Souza, E. S., Ito, A. S., Hirata, I. Y., and Juliano, L. (2000) End-to-end distance distribution in bradykinin observed by Förster resonance energy transfer, *Biochim. Biophys. Acta 1474*, 251−261.
58. Tucker, M. J., Oyola, R., and Gai, F. (2005) Conformational distribution of a 14-residue peptide in solution: A fluorescence resonance energy transfer study, *J. Phys. Chem. B 109*, 4788−4795.
59. Schuler, B., Lipman, E. A., Eaton, W. A., Steinbach, P. J., and Kumkell, M. (2005) Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence, *Proc. Natl. Acad. Sci. U.S.A. 102*, 2754−2759.
60. Sahoo, H., Roccatano, D., Zacharias, M., and Nau, W. M. (2006) Distance distributions of short polypeptides recovered by fluorescence resonance energy transfer in the 10 Å domain, *J. Am. Chem. Soc. 128*, 8118−8119.
61. Zhou, H.-X. (2001) Loops in proteins can be modeled as worm-like chains, *J. Phys. Chem. B 105*, 6763−6766.
62. Miller, W. G., Brant, D. A., and Flory, P. J. (1967) Random coil configurations of polypeptide copolymers, *J. Mol. Biol. 23*, 67−80.
63. Phillips, G. N. (1997) Structure and dynamics of green fluorescent protein, *Curr. Opin. Struct. Biol. 7*, 821−827.
64. Arai, R., Nagamune, T., Nishikawa, Y., Fujisawa, T., and Wriggers, W. (2004) Conformations of variably linked chimeric proteins evaluated by synchrotron X-ray small-angle scattering, *Proteins: Struct., Funct., Genet. 57*, 829−838.
65. Hong, S.-H., Hao, Q., and Maret, W. (2005) Domain-specific fluorescence resonance energy transfer (FRET) sensors of metallothionein/thionein, *Protein Eng. Des. Sel. 18*, 255−263.
66. Hong, S.-H., and Maret, W. (2003) A fluorescence resonance energy transfer sensor for the β-domain of metallothionein, *Proc. Natl. Acad. Sci. U.S.A. 100*, 2255−2260.

BI061288T