Folding of Green Fluorescent Protein and the Cycle3 Mutant†

Hiroyuki Fukuda, Munehito Arai, and Kunihiro Kuwajima*

Department of Physics, School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Received March 8, 2000; Revised Manuscript Received June 19, 2000

ABSTRACT: Although the correct folding of green fluorescent protein (GFP) is required for formation of the chromophore, it is known that wild-type GFP cannot mature efficiently in vivo in *Escherichia coli* at 37 °C or higher temperatures that the jellyfish in the Pacific Northwest have never experienced. Recently, by random mutagenesis by the polymerase chain reaction (PCR) method, a mutant called Cycle3 was constructed. This mutant had three mutations, F99S, M153T, and V163A, on or near the surface of the GFP molecule and was able to mature correctly even at 37 °C [Crameri et al. (1996) *Nat. Biotechnol. 143*, 315−319]. In the present study, we investigated the differences in their folding behavior in vitro. We observed the folding and unfolding reactions of both wild-type GFP and the Cycle3 mutant by using green fluorescence as an indicator of the formation of the native structure and examining hydrogen-exchange reactions by Fourier transform infrared spectroscopy. Both proteins showed unusually slow refolding and unfolding rates, and their refolding rates were almost identical under the native state at 25 and at 35 °C. On the other hand, aggregation studies in vitro showed that wild-type GFP had a strong tendency to aggregate, while the Cycle3 mutant did not. These results indicated that the ability to mature efficiently in vivo at 37 °C was not due to the improved folding and that reduced hydrophobicity on the surface of the Cycle3 mutant was a more critical factor for efficient maturation in vivo.

Green fluorescent protein (GFP)[1] from *Aequorea victoria* is a soluble globular protein that consists of 238 residues. The crystal structure of GFP, resolved a few years ago by two groups working independently (*1*, *2*), was found to be an 11 stranded β-barrel wrapped around a single central helix (Figure 1). The advantageous feature of GFP is that it can emit a visible green fluorescence without the use of any substrate, because it spontaneously forms a chromophore by the autocatalytic cyclization of the polypeptide backbone. For wild-type GFP, the excitation peaks are at 395 and 475 nm, and the emission peak is at 508 nm. GFP has been used as a useful gene expression marker in a wide variety of organisms (*3*−*6*), and GFP tagged with a target protein reports the time and location of expression of the target protein by its bioluminescence. Recently, many mutants with useful characters such as enhanced fluorescence intensity (*7*), shifted emission wavelength (*1*, *8*), and pH-sensitive fluorescence change (*9*) have been discovered.

In addition to these biological studies, many physicochemical studies of GFP have been carried out with respect to chromophore formation and fluorescence emission. Makino et al. (*10*) have also shown that GFP is useful for studying chaperonin-mediated folding; its green fluorescence is an indicator of formation of the native structure. However, rather little is known about the detailed mechanism of the folding of GFP itself. To date, only pioneering studies in the early 1980s by Ward and Bokman (*11*, *12*) reporting basic properties such as temperature and pH stability and the efficiency of refolding from the acid-, base-, and guanidine hydrochloride-(GdnHCl-) induced unfolded states, have appeared. Their results showed that GFP was refolded with a more than 50% yield.

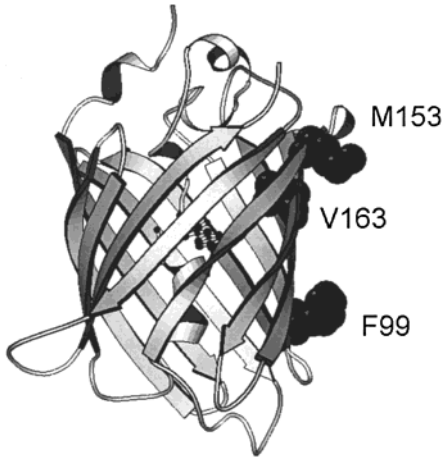

FIGURE 1: X-ray crystal structure of wild-type GFP (PDB code 1GFL). The mutation sites of the Cycle3 mutant are shown as a space-filling union of spheres model and the chromophore is shown as a ball-and-stick model. The drawing was generated by the program MOLSCRIPT (*35*).

In the present study, we investigate the in vitro folding of GFP and its mutant, Cycle3. Although GFP is a highly useful protein, as noted above, wild-type GFP does not mature efficiently in vivo at 37 °C or higher temperatures that the jellyfish in the Pacific Northwest have never experienced.

† This work is supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, and Science of Japan.

* To whom correspondence should be addressed: phone: +81-3-5841-4128; fax: +81-3-5841-4512; e-mail kuwajima@phys.s.u-tokyo.ac.jp.

[1] Abbreviations: GFP, green fluorescent protein; PCR, polymerase chain reaction; CD, circular dichroism; GdnHCl, guanidine hydrochloride; EDTA, ethylendiamine-*N*,*N*,*N*′,*N*′-tetraacetic acid; DTT, dithiothreitol; FTIR, Fourier transform infrared.

10.1021/bi000543l CCC: $19.00
Published on Web 09/08/2000

Consequently, it cannot serve as a gene expression marker. This low maturation efficiency has been ascribed to the low efficiency of folding under that temperature condition. Therefore, many GFP mutants in which the maturation efficiency in vivo is improved have been constructed, and this improved maturation efficiency is often ascribed to their improved folding efficiency in vivo. Among these so-called "folding mutants" of GFP, the Cycle3 mutant, created by Crameri et al. (*13*), is the first one to be constructed by random mutations with PCR. The Cycle3 mutant has three mutations, F99S, M153T, and V163A, on or near the surface of the GFP molecule and matures efficiently in *Escherichia coli* at 37 °C. While other mutants are often subjected to changes in the chromophoric structure and the fluorescence spectrum, the Cycle3 has the same chromophore and the same fluorescence spectrum as wild-type GFP.

Here, we have studied the folding and unfolding reactions of both wild-type GFP and the Cycle3 mutant by fluorescence and circular dichroism (CD) spectroscopy and examined the hydrogen-exchange reactions of the proteins by Fourier transform infrared (FTIR) spectroscopy. We show that, rather surprisingly, the folding and hydrogen-exchange behavior of the Cycle3 mutant is essentially identical to that of the wild-type protein and that the efficient maturation of the Cycle3 mutant in vivo is therefore not due to its efficient folding. In addition, aggregation studies in vitro show that the efficient maturation of the Cycle3 mutant in vivo is brought about by its hydrophilic substitutions, which prevent irreversible aggregation during folding in vivo.

MATERIALS AND METHODS

Reagents. D_2O (>99.8%) was obtained from CEA. GdnHCl was a reagent-grade specially prepared for biochemical use from Nakarai Tesque, Inc. (Kyoto, Japan). Other chemicals were guaranteed reagent-grade. Concentration of GdnHCl was determined by an Atago 3T refractometer by use of (*14*)

$$[\mathrm{GdnHCl}] = 57.147(\Delta N) + 38.68(\Delta N^2) - 91.60(\Delta N^3)$$

where ΔN and [GdnHCl] are the reflective index increment at 589 nm at 20 °C and the concentration of GdnHCl, respectively.

Plasmid Construction. The expression plasmid, pGFPT7, was constructed by the following digestion and ligation steps. The coding region of GFP was excised from a commercially available plasmid, pGFP (Clontech), at the *Xba*I and *Eco*RI sites, and the fragment obtained was inserted between the *Nhe*I and *Eco*RI sites of an overexpression vector, pSCREEN-1b(+) (Novagen). By this procedure, GFP was located under the control of a T7 promoter, and a 34-base noncoding region, which was followed by the 5′ end of the GFP coding region, was also included. This noncoding region was deleted by the Kunkel method with a Muta-Gene phagemid in vitro mutagenesis kit (Bio-Rad). For construction of the Cycle3 mutant, a triple mutation (F99S, M153T, and V163A) was introduced to pGFPT7 by the Kunkel method. Sequences of the resulting plasmids were confirmed by an ABI Prism 310 genetic analyzer (Applied Biosystems).

Protein Expression and Purification. The resulting plasmid was transformed into *E. coli* cells of the strain BL21(DE3). Because GFP tends to form inclusion bodies at 37 °C in vivo, high-level expression of the soluble mature form of GFP was achieved by growing the cultures in LB broth at 24 °C to saturation without induction by isopropyl β-D-thiogalactoside (*8*). To prepare soluble extracts, the cells were collected by centrifugation and lysed by French press (Sim-Aminco) or sonication (a Kubota 201M sonicator) in a lysis buffer containing 50 mM Tris buffer, pH 8.0, 1 mM EDTA, and 0.1 M NaCl. Protein purification was carried out by the following steps: (1) differential ammonium sulfate precipitation (20−70%), (2) gel filtration on Sephadex G-75 (Pharmacia), (3) ion-exchange chromatography on DEAE-Sepharose with NaCl gradient elution (0−0.2 M), and (4) gel filtration on Sephadex G-75. All chromatography was performed in 20 mM Tris buffer, pH 7.5, containing 1 mM EDTA and 1 mM DTT, and the fractions containing GFP were determined by the visible green fluorescence. The reducing agent (1 mM DTT) was included for prevention of incorrect disulfide bond formation, because GFP has two cysteine residues that are not bonded to each other in the native structure. We could obtain 400 mg of GFP from an 8-L culture by this procedure. The purity was confirmed by SDS−PAGE. The ratio of the absorbance at 395 nm to that at 280 nm, known to be an index of the purity of GFP, is 1.1−1.2 for pure GFP (*15*). The final sample prepared in this study had a ratio of 1.1 or above.

Comparison of Expression in Vivo between Wild-Type GFP and the Cycle3 Mutant. To investigate the maturation efficiency of each of the two proteins, the *E. coli* cells harboring the expression plasmid were incubated overnight at both 24 and 37 °C. After cell lysis by sonication, the soluble mature fraction was separated from the insoluble fraction by centrifugation; from chromatographic separation of the soluble fraction and the subsequent measurement of the ratio of absorbance at 395 nm to that at 280 nm, the soluble fraction contained exclusively the mature form. An aliquot (5 μL) from each 1 mL fraction of the culture was applied to SDS−PAGE and quantified by Image Master VDS (Pharmacia).

CD and Fluorescence Spectral Measurements. CD and fluorescence spectral measurements were carried out in a Jasco J-720 spectropolarimeter and a Jasco FP-777 spectrofluorometer, respectively. The path length of optional cuvettes was 1.0 mm for the far-UV CD and 10.0 mm for the near-UV CD and fluorescence, respectively. For the fluorescence measurements, the spectral bandwidth was 1.5 nm for both excitation and emission. Samples for spectral measurements were prepared in 20 mM potassium phosphate buffer, pH 7.0, containing 1 mM EDTA and 1 mM DTT for the native state, and an additional 6 M GdnHCl was included for the unfolded state. The protein concentration was 0.5 mg/mL for CD and 0.03 mg/mL for fluorescence. The protein concentration was determined by UV absorption with a molar absorption coefficient of $\epsilon(280) = 2.06 \times 10^4\ \mathrm{M^{-1}\ cm^{-1}}$, which was obtained by the Biuret method. This value is in good agreement with the value of 2.06×10^4, which was predicted from the amino acid composition by the method of Pace et al. (*16*). The CD spectra were represented by the mean residue ellipticity with a mean residue weight of 113.

Samples for the equilibrium unfolding measurements were prepared in 20 mM Tris buffer, pH 7.5, containing 0.1 M NaCl and 1 mM EDTA, and the measurement was performed at two different protein concentrations (0.01 and 0.1 mg/

mL). The path lengths of the optical cuvettes for both far-UV CD and fluorescence measurements were 10.0 and 5.0 mm for the samples with the protein concentration of 0.01 and 0.1 mg/mL, respectively.

Kinetic Fluorescence Measurements. The instrumental conditions were the same as those for the fluorescence spectral measurements, described above. The fluorescence generated by excitation light at 395 nm was monitored at 508 nm. Protein solution was prepared in 20 mM Tris buffer, pH 7.5, containing 0.1 M NaCl, 1 mM EDTA, and 1 mM DTT, and the final protein concentration was between 0.01 and 0.05 mg/mL. For the kinetic measurements, the reactions were initiated by manual mixing of the protein and diluent solutions; the mixing ratio was 10−20 for the refolding and 100 for the unfolding (diluent/protein). Mixing resulted in jumps in the concentration of GdnHCl from 6 M for refolding and 0 M for unfolding to the concentrations indicated. The reaction curve was traced continuously in the spectrofluorometer for the reaction that finished within 1 h, while the reaction mixture was incubated in a water bath and an aliquot of the mixture was taken for the fluorescent measurement at a desired time for the reaction that took over 1 h because there was slow irreversible photolysis of the chromophore. Kinetic data were fitted by the method of nonlinear least-squares to the following equations by KaleidaGraph (Synergy Software). For the unfolding reaction

$$A(t) = \Delta A e^{-kt}$$

where $A(t)$ and ΔA are the fluorescence intensities at time t and the amplitude, respectively, and k represents the first-order rate constant. In this equation, we assume that the fluorescence at infinite time is 0 and the ΔA and k are fitting variables, given that the unfolded state is completely nonfluorescent. For the refolding reaction

$$A(t) = A(0) + \sum \Delta A_i (1 - e^{-k_i t})$$

where ΔA_i and k_i are the amplitude and the first-order rate constant of ith phase, respectively. In this equation, ΔA_i and k_i are the fitting variables. $A(0)$ was not zero because a burst phase occurred in refolding.

Hydrogen-Exchange Experiments by Fourier Transform Infrared Spectroscopy. FTIR spectra were obtained in a Magna-IR E.S.P. System 560 (Nicolet). The protein sample for hydrogen-exchange measurements was first prepared in 20 mM Tris buffer, pH 7.5, and then lyophilized. Because EDTA has strong IR absorbance in the amide band region, we did not use EDTA in the FTIR measurements. CaF_2 cells with a path length of 50 μm were used for both the sample and background measurements. The spectral resolution used was 4 cm^{-1}. Each spectrum was averaged over 32 scans except during the first few minutes, when it was averaged over four scans. The hydrogen-exchange reactions were initiated by dissolving the lyophilized sample in D_2O, and the dead time arising from the mixing was estimated at less than 1 min. The spectra obtained were corrected for H_2O vapor and smoothed by the software provided by the manufacturer. The absorbance of the amide I and II bands was evaluated from the spectra at the wavenumbers of their maxima, 1630 and 1545 cm^{-1}, respectively. The baseline was corrected with the absorbance measured at 1735 cm^{-1}. To evaluate the kinetics of the hydrogen-exchange reactions, the absorbance of the amide II band divided by that of the amide I band was used as an indicator of the fraction of unexchanged amide protons of the peptide bonds and was plotted against time.

Aggregation Studies in Vitro. Protein aggregation during refolding was monitored by light scattering at 640 nm. The light scattering intensity was measured in the FP-777 spectrofluorometer, with the excitation and emission wavelengths both set at 640 nm. The components of the sample solution were the same as those used in the kinetic fluorescence measurements, though the protein concentration during refolding was more than 10-fold higher, at 0.5 mg/mL. The reaction was initiated by a GdnHCl concentration jump from 6 to 0.6 M.

RESULTS

Comparison of Expression and Aggregation in Vivo between Wild-Type GFP and the Cycle3 Mutant. Expression in vivo of the two proteins was carried out as described under Materials and Methods. We were able to discern by their fluorescence whether GFP was well produced in its soluble mature form. SDS−PAGE and the quantification of the protein bands by Image Master VDS (Pharmacia) were carried out for more precise analysis of the efficient maturation of GFP. Both GFP proteins were dominantly produced, relative to production of other proteins of *E. coli*, owing to the control of the T7 promoter. At 37 °C, the Cycle3 mutant became efficiently mature as the soluble form (47% of the expressed Cycle3 mutant), while wild-type GFP was hardly seen in the soluble fraction. Both GFP proteins are able to mature at 24 °C (more than 40%), though even at this temperature about half of the total GFP expressed still forms inclusion bodies.

CD and Fluorescence Spectra. The CD spectra of wild-type GFP and the Cycle3 mutant were measured at protein concentrations from 0.01 to 0.5 mg/mL. The spectra were identical between the two proteins and did not depend on the protein concentration. The far-UV CD spectrum in the native state has a strong negative deflection near 217 nm, indicating a significant spectral contribution from the β-structure. The secondary structure is disrupted by a treatment with 6 M GdnHCl, and the resulting CD spectrum is indicative of a disordered coil. The CD spectrum from the near-UV to the visible region shows that, in addition to the typical peaks arising from the ellipticity of tyrosine and tryptophan residues, the chromophore exhibits a broad CD band centered around 400 nm. The unfolding of the protein diminishes the aromatic peaks though the chromophore CD is still present, with the peak position shifted to around 380 nm.

The fluorescence spectrum excited at 395 nm shows a single peak at 508 nm with a little shoulder on its right side. The protein unfolding is accompanied by the loss of fluorescence, and no emission peaks are observed in the unfolded state. This is because the solvent quenches the fluorescence when the GFP is unfolded. On the basis of these observations, we suggest that the green fluorescence of GFP is a useful indicator of the formation of the native structure.

Comparison of Transition Curves between Wild-Type GFP and the Cycle3 Mutant. To compare the thermodynamic stabilities between the two proteins, we first tried to obtain

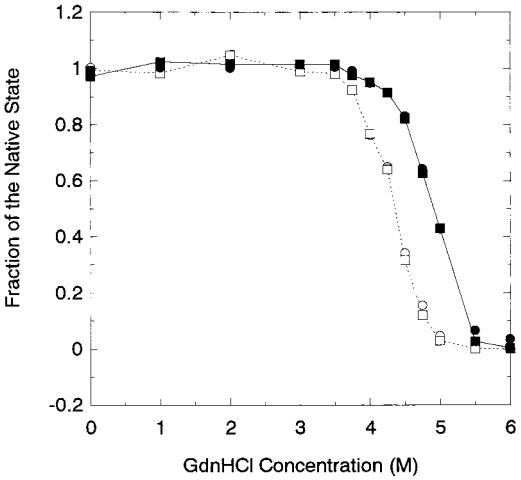

FIGURE 2: GdnHCl-induced unfolding transition curves of wild-type GFP (open symbols) and the Cycle3 mutant (solid symbols) measured by their green fluorescence at 508 nm (squares) and the far-UV CD at 222 nm (circles) at pH 7.0 and 25 °C. Apparent fraction of the native state is plotted against GdnHCl concentration. The dotted and continuous lines connect the fluorescence data of the wild-type GFP and the Cycle3 mutant, respectively. The protein concentration was 0.1 mg/mL.

the transition curves of GdnHCl-induced unfolding (Figure 2). These curves were obtained by measuring the far-UV CD at 222 nm and the green fluorescence at 508 nm, which indicated the presence of the native structure, and these measurements were carried out after overnight incubation of the protein solution at indicated concentrations of GdnHCl at 25 °C. The transition curve measured by the far-UV CD coincides with the transition curves measured by the green fluorescence (Figure 2), indicating that the loss of the green fluorescence is a useful measure of the unfolding transition. The transition curves measured at different protein concentrations (0.01 and 0.1 mg/mL) were also coincident with each other, showing the absence of protein concentration dependence of the transitions between 0.01 and 0.1 mg/mL.

Figure 2 shows that both proteins were very stable against the denaturant, but the Cycle3 mutant is apparently more stable than the wild-type GFP. However, we must mention that the unfolding reactions of the proteins were extremely slow (see kinetic data); they had not reached equilibrium after several days under conditions of medium GdnHCl concentrations. Therefore, the curves in Figure 2 do not represent the equilibrium unfolding. If the mutant unfolds much more slowly, it could be supposed that the Cycle3 mutant is more stable than wild-type GFP. In fact, we observed that the unfolding kinetics of the Cycle3 mutant were slower than those of the wild-type GFP (see below). To further investigate the reversibility of the unfolding transition, we also measured the recovery of the fluorescence of the proteins, which had been previously unfolded once, in concentrated GdnHCl. We observed that the recovery of the fluorescence did not occur in the presence of at least 2 M GdnHCl. Therefore, it is practically impossible to obtain the thermodynamic stabilities of the two proteins because of their unusually slow kinetics and possible irreversibility. This situation was not altered when we changed the pH from 7.0 to 9.0.

Refolding and Unfolding Kinetics. Refolding and unfolding reactions were monitored directly in the fluorescence spectrophotometer (Figure 3), and the logarithm of the first-order rate constant was plotted against the GdnHCl concentration (Figure 4). The far-UV CD spectra and green fluorescence showed full unfolding of the proteins after a 3-h incubation at 6 M GdnHCl, so that we can follow the total global refolding and unfolding reactions by means of the appropriate concentration jumps of GdnHCl. During the folding reactions in the GdnHCl-induced unfolding state, however, both wild-type GFP and the Cycle3 mutant tended to aggregate when the protein concentration was more than 0.1 mg/mL; thus, we used a concentration lower than 0.05 mg/mL. The percentage of the fluorescence recovery also depended on the incubation time during the unfolded state (*12*). Therefore, we initiated the folding reaction with a fixed incubation time; at 3 h after the unfolding reaction had begun. Due to the slow kinetics of the proteins, we carried out measurements below 1.4 M for refolding and above 3.0 M GdnHCl for unfolding.

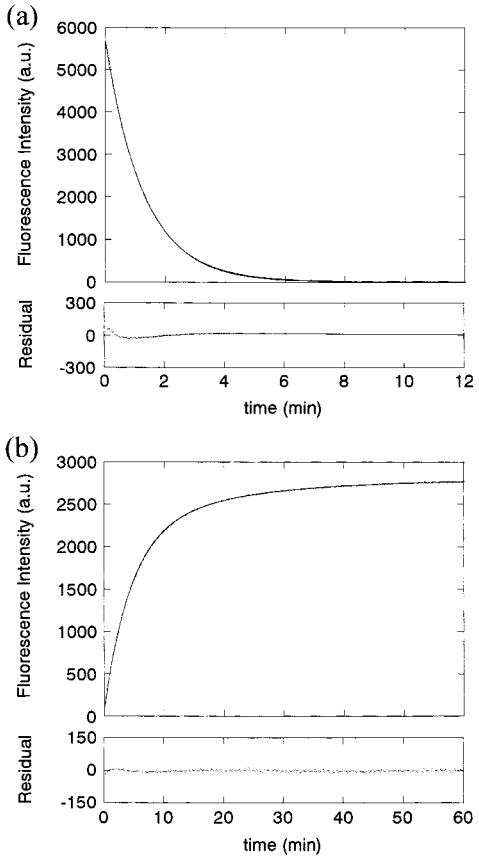

FIGURE 3: (a) Typical kinetic unfolding curve of wild-type GFP caused by a jump of GdnHCl concentration from 0 to 6.5 M at pH 7.5 and 25 °C. The protein concentration was 0.05 mg/mL. The curve (dotted line) was well-fitted to a single exponential (continuous line) ($\Delta A = 5704 \pm 4$, $k = (1.309 \pm 0.001) \times 10^{-2}$ s^{-1}; errors are standard errors of fitting). Residual plot for fitting with the single exponential is shown in the lower panel. (b) Typical kinetic refolding curve of wild-type GFP caused by a jump of GdnHCl concentration from 6 to 0.6 M at pH 7.5 and 25 °C. The protein concentration was 0.05 mg/mL. The curve (dotted line) was well-fitted to double exponentials (continuous line) [$A(0) = 80 \pm 2$, $\Delta A_1 = 2017 \pm 9$, $k_1 = (3.80 \pm 0.01) \times 10^{-3}$ s^{-1}, $\Delta A_2 = 693 \pm 9$, $k_2 = (9.4 \pm 0.1) \times 10^{-4}$ s^{-1}; errors are standard errors of fitting]. Residual plot for fitting with the double exponentials is shown in the lower panel.

The reaction curves of unfolding and refolding of both proteins were able to be fitted with a single exponential,

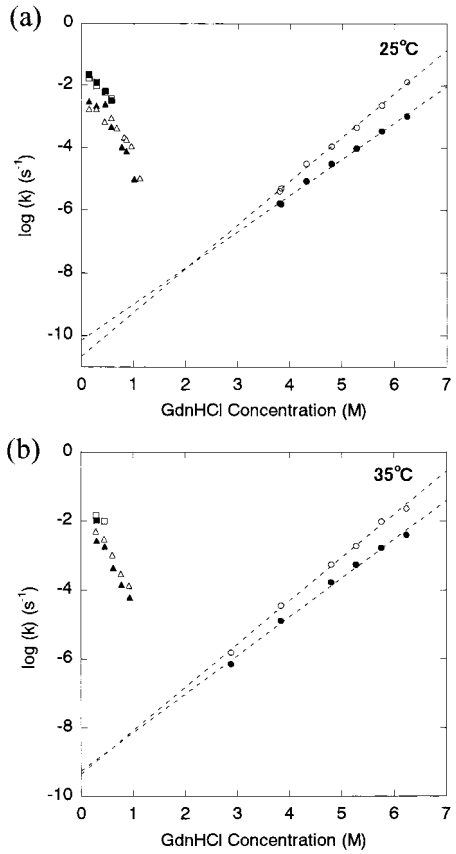

FIGURE 4: GdnHCl concentration dependence of the apparent rate constants of the refolding (fast phase □, slow phase △) and the unfolding reactions (○) of wild-type GFP and the refolding (fast phase ■, slow phase ▲) and the unfolding (●) of the Cycle3 mutant measured by their green fluorescence at pH 7.5, at (a) 25 °C and (b) 35 °C.

except in the case of the refolding reactions at low GdnHCl concentrations (below about 0.6 M at 25 °C and 0.45 M at 35 °C), where the double exponential was more suitable (Figure 3). For purposes of comparing the behavior of the proteins in vivo, and considering that only the Cycle3 mutant could efficiently fold at 37 °C, the experiments were carried out at both 25 and 35 °C. The results at 25 °C were found to be essentially identical to the results at 35 °C.

In the unfolding reaction, the rate constant shows a linear dependence on the GdnHCl concentration like that observed in other globular proteins, although the time constant of the unfolding reaction varies from minutes to weeks, relative to the GdnHCl concentration (Figure 4). In the region of high GdnHCl concentration, the unfolding rate of the Cycle3 mutant is much slower than that of wild-type GFP, and the Cycle3 mutant is apparently more stable than wild-type GFP. However, when extrapolated to zero GdnHCl concentration, the rate constants of the proteins grow close to each other or the rate constant of the Cycle3 mutant becomes faster than that of wild-type GFP because the GdnHCl dependence of the unfolding rate constant is steeper for the Cycle3 mutant.

The refolding kinetics was more complicated than the unfolding kinetics. Above 1.4 M GdnHCl concentration at 25 °C and above 1.0 M GdnHCl at 35 °C, the fluorescence recovery was almost zero, which supported the observation that the transition curves (Figure 2) did not reach their equilibrium. The refolding reaction is biphasic for both

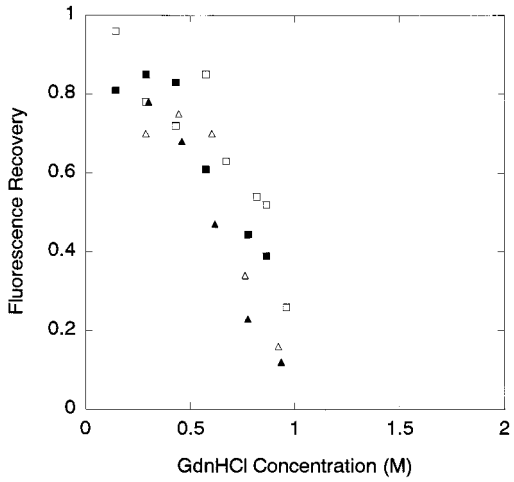

FIGURE 5: GdnHCl dependence of the fluorescence recovery from the unfolded state for wild-type GFP at 25 °C (□) and 35 °C (△), and for the Cycle3 mutant at 25 °C (■) and 35 °C (▲) obtained by the kinetic fitting analysis at each GdnHCl concentration.

proteins at low GdnHCl concentrations, but the slow phase becomes dominant as the concentration is increased, and finally the reaction becomes a single phase (Figure 4). The rate constants of the refolding reactions of the proteins are almost identical under the native condition at both 25 and 35 °C, though wild-type GFP folds slightly faster in the presence of more than 0.6 M GdnHCl. The temperature dependence of the refolding rate is smaller than that of the unfolding rate.

The results of these kinetic measurements show that the proteins have almost identical kinetic stability under the native condition. Our findings contradict the idea that the Cycle3 mutant can fold more rapidly than wild-type GFP at 37 °C. The Cycle3 mutant is not sufficiently stable, compared with wild-type GFP, to explain the difference between their maturation efficiency in vivo. It is remarkable that the refolding and unfolding rates of the proteins showed the same temperature dependence, while only the Cycle3 mutant is able to mature correctly in vivo at 37 °C.

Comparison of Fluorescence Recovery between Wild-Type GFP and the Cycle3 Mutant. We also investigated the GdnHCl dependence of the fluorescence recovery from the unfolded state. We summed up all amplitudes of the kinetic refolding curve obtained by the kinetic fitting analysis at each GdnHCl concentration (Figure 5). The value of 100% fluorescence recovery corresponds to the fluorescence intensity of the same protein concentration at zero GdnHCl concentration. From this perspective, the data shown in Figure 5 can be considered to represent the "refolding" transition curves.

The results show that at a low protein concentration (from 0.01 to 0.05 mg/mL), the renaturation efficiency of both GFP proteins is up to almost 100% when GdnHCl concentration is extrapolated to zero, though it decreases rapidly from about 0.5 M GdnHCl. This transition of the Cycle3 mutant occurred slightly faster than that of wild-type GFP, which is probably due to the slower refolding rate of the Cycle3 mutant at about 1 M GdnHCl concentration or higher (see above). Thus, the advantage that the Cycle3 mutant can become efficiently mature in vivo even at 37 °C was no longer observed.

On the basis of linear extrapolation of the unfolding rate constant to lower GdnHCl concentrations, the unfolding rate

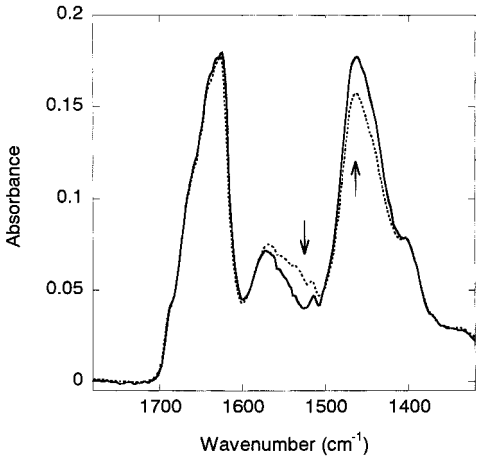

FIGURE 6: Typical FTIR spectra of wild-type GFP at pH 8.0 and 25 °C 1 min (dotted line) and 3 h (continuous line) after the hydrogen-exchange reaction began. The intensity of the amide II band (at approximately 1540 cm^{-1}) is proportional to the number of amide protons that remain unexchanged. Arrows show the direction of changes.

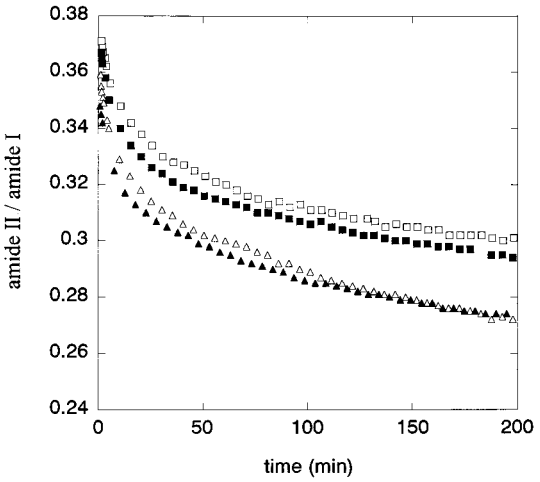

FIGURE 7: Time course of hydrogen-exchange reactions obtained by FTIR spectroscopy for wild-type GFP at 25 °C (□) and 35 °C (△), and for the Cycle3 mutant at 25 °C (■) and 35 °C (▲). The absorbance of amide II divided by that of the amide I is plotted against elapsed time.

of both proteins is expected to be extremely slow. Therefore, it is curious that the recovery of fluorescence does not reach 100%. One explanation for this inability to fold correctly is that some intermediate or misfolded states that do not emit fluorescence may exist and kinetically compete with the mature fluorescent form.

Hydrogen-Exchange Reaction Studied by FTIR Spectroscopy. To investigate the nature of the two proteins further, hydrogen-exchange experiments were carried out in D_2O by FTIR spectroscopy. Figure 6 shows FTIR spectra of wild-type GFP obtained immediately and 3 h after the protein was dissolved in D_2O. Essentially the same spectra were obtained for the Cycle3 mutant. The absorbance bands around 1630 and 1530 cm^{-1} correspond to the amide I and II bands, respectively. The amide I band arising from the peptide CO stretching is insensitive to hydrogen−deuterium exchange and was used as a measure of protein concentration, and the amide II band mainly arising from the peptide NH bending and CN stretching gives the amount of unexchanged amide protons. The amide II band is, however, almost covered by another band arising from COO^-. We can see that the amide II region is diminishing as time elapses; namely, the hydrogen-exchange reaction proceeds.

The time courses of the hydrogen-exchange reactions are shown in Figure 7, in which the absorbance of amide II divided by that of the amide I is plotted against elapsed time. Because a pH dependence of the apparent reaction rate was observed (data not shown), the results were able to be interpreted in terms of the EX2 mechanism; that is, the intrinsic hydrogen exchange rate was much slower than the rate of local fluctuation. Under the assumption of the EX2 mechanism, the hydrogen exchange rate constant is given by

$$k_{obs} = \Gamma k_{ex}$$

where Γ and k_{ex} represent the probability of existence of amide protons that are in the open state and prepared for the exchange reaction and the intrinsic exchange rate constant of amide protons that are not protected, respectively, which variables are given by Hvidt and Nielsen (*17*) as

$$\Gamma = \frac{[O]}{[O] + [C]}$$

$$k_{ex} = (10^{-pH} + 10^{pH-6.0})10^{0.05(T-25)}$$

with k_{ex} in units of reciprocal seconds. [O] and [C] represent the concentration of the amide protons that are in the open state and prepared for exchange and the concentration of the amide protons that are in the closed state and protected from exchange. The value of $1/\Gamma$ is the so-called protection factor that is used as a convenient index to represent how strongly one amide proton is protected from deuterium attack. In this experiment, a portion of the amide protons was not exchanged for deuterium within 24 h at both 25 and 35 °C, and elevated temperature caused protein aggregation. Therefore, we could not observe the whole exchange reaction.

Figure 7 shows the reaction curves of the two proteins, which are almost identical at 25 and at 35 °C, which finding indicates that the local stabilities of the GFP proteins are also identical. This result is not inconsistent with the kinetic fluorescence measurements of the unfolding and refolding shown above.

Aggregation Studies in Vitro. Wild-type GFP and the Cycle3 mutant were almost indistinguishable in terms of the in vitro folding−unfolding and hydrogen-exchange behavior. To further understand the relationship between the in vitro properties and the in vivo maturation efficiencies of the proteins, we investigated the aggregation properties of the two proteins. Figure 8 shows the time courses of the aggregation in vitro during the refolding reaction monitored by light scattering at 640 nm. The protein concentration is more than 10 times higher than those in the refolding kinetic experiments described in the previous section. Both wild-type GFP and the Cycle3 form insoluble aggregates at this protein concentration. The results show that wild-type GFP has a much stronger tendency than the Cycle3 mutant to aggregate and that the rate of aggregation is much faster at 35 than at 25 °C. The rate of aggregation of wild-type GFP is comparable to the rate of folding for wild-type GFP at 35

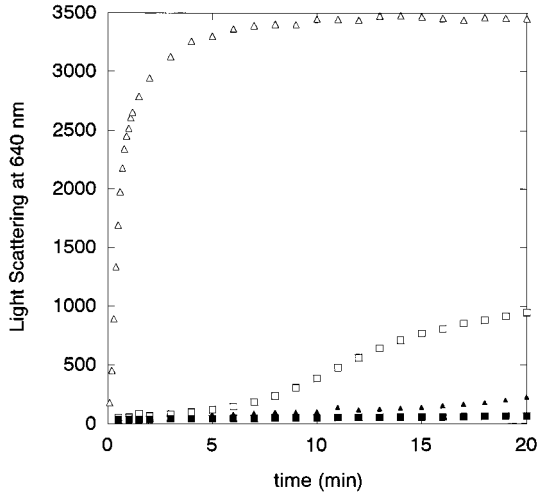

FIGURE 8: Time course of protein aggregation monitored by light scattering at 640 nm for wild-type GFP at 25 °C (□) and 35 °C (△), and for the Cycle3 mutant at 25 °C (■) and 35 °C (▲). The protein concentration was 0.5 mg/mL.

°C. These results are thus consistent with the results of the in vivo experiments showing that wild-type GFP does not efficiently produce the soluble mature form when grown in bacterial culture at 37 °C.

The time course of the aggregation formation measured by the light scattering is sigmoidal due to the presence of a lag phase (Figure 8), which finding suggests that the kinetics of the aggregation are represented by a nucleation−polymerization mechanism characterized by formation of protein aggregates such as inclusion bodies and amyloid fibrils (*18*).

DISCUSSION

Comparison between Wild-Type GFP and the Cycle3 Mutant. As noted above, our results revealed that, at both 25 and 35 °C in vitro, the proteins have identical folding properties such as rate constant and efficiency of folding under the native condition. The local structural fluctuations determined by hydrogen-exchange measurements are also identical between the proteins. The only difference observed is in the tendency toward aggregation during refolding.

Two surface hydrophobic residues (F99 and M153) in wild-type GFP are substituted with hydrophilic ones in the Cycle3 mutant, and one residue (V163) near them is substituted with a less hydrophobic one. Therefore, the Cycle3 mutant is expected to aggregate less, and therefore become more soluble, than wild-type GFP. It is reasonable that these mutations, which are on the surface of the molecule, do not affect the entire stability of protein, because they affect the stability of both the native and the unfolded states to the same extent. Our results indicate that the difference in the maturation efficiency between the two proteins does not arise from the difference in the thermodynamic or kinetic stability but that the difference in hydrophobicity plays a role in the efficient maturation of the Cycle3 mutant in vivo at 37 °C. It has been observed that mutated protein forms insoluble aggregates in vivo when it is destabilized by mutation, but this does not occur in the Cycle3 mutant of GFP.

Other "Folding" Mutants of GFP. Many mutations that improve the in vivo maturation of GFP at 37 °C have been identified (*7*, *19*−*23*), and these mutants are often called "folding" mutants. Cubitt et al. (*24*) remarked in their review that these mutations were divided into the following four classes: (1) mutations close to the chromophore, including five (S65A, -G, -C, -T, and -L) actually within the chromophore, two (F64L and S72A) that are close to it in the central α-helix, and others (Y145F, I167T, T203Y, and S205T) that are partially or completely buried; (2) mutations distant from the chromophore and in a buried location (V163A); (3) surface-located "folding" mutations (F99S, M153T, and S175G); and (4) surface-located "folding" mutations that also lie close to the chromophore (S147P and N149K). Although these mutants show remarkably improved maturation efficiency in vivo and are often called "folding" mutants, little is known about the folding behavior of the mutants in vitro. To our knowledge, this report is the first in which the refolding and the unfolding kinetics of both wild-type GFP and its Cycle3 mutant are measured. Our results show that the Cycle3 mutant is not improved in its folding properties in vitro as compared with those of wild-type GFP. Only the aggregation tendency observed at high concentrations of the protein is improved by the hydrophilic substitutions. Therefore, it is likely that the other "folding" mutants containing such mutations do not really fold efficiently in vitro.

Inclusion-Body Formation and Thermodynamic Stability. Mutational effects on inclusion-body formation have been studied in many proteins, but systematic mutational analyses of inclusion-body formation have been carried out for only a small number of proteins including immunoglobulin light chain V_L domain, P22 bacteriophage tailspike and coat proteins, γ-interferon, and interleukin 1β (*25*, *26*). On the basis of these analyses, it has generally been accepted that there is a significant correlation between inclusion-body formation and the stability of partially folded intermediates. However, whether there is a similar correlation between inclusion-body formation and the stability of the native protein seems to depend on protein species. For example, a good correlation between inclusion-body levels and the midpoint of GdnHcl unfolding curves was observed in recombinant immunoglobulin light chain REI V_L domain (*27*). It is also known that protein destabilization by mutations occasionally leads to the formation of amyloid fibrils in amyloidogenic proteins (*28*−*33*). On the other hand, Chrunyk et al. (*34*) have shown that there are no such correlations between inclusion-body formation and protein stability in interlukin 1β. Specifically, the K97V mutant located near the surface of the molecule produces substantially more interleukin 1β as inclusion bodies than does the wild-type protein, though the mutant protein is thermodynamically more stable than the wild type. The absence of the positive correlation between maturation efficiency in vivo and thermodynamic stability in vitro for the surface mutant of interleukin 1β is thus similar to the present results of the Cycle3 mutant.

Folding of GFP. We have studied the folding of GFP, and two points bear a second mention. First, GFP, compared with typical globular proteins, folds and unfolds very slowly. Figures 2 and 5 show that the unfolding−refolding transitions of GFP are not at equilibrium but refolding is only partially reversible. One explanation for this is that GFP does not reach its equilibrium within a few days because of the slow kinetics at intermediate concentrations of GdnHCl. The

refolding and unfolding rate constants in the presence of 2 M GdnHCl at 25 °C are estimated to be between 10^{-8} and 10^{-9} s^{-1} (Figure 4). However, as remarked earlier, the apparent irreversibility cannot be explained solely by the slow kinetics. At a low GdnHCl concentration, e.g., 0.6 M, the fluorescence recovery was not up to 100% although the refolding is sufficiently fast. Such irreversibility was also observed when the protein was refolded at a moderately high temperature, such as 50 °C; GFP was still stable, but it did not unfold (*10*). Possible explanations for this irreversibility are as follows: (i) GFP has two cysteine residues that are not bonded to each other in the native structure, and an incorrect disulfide bond is formed during the refolding; (ii) the chromophore is destroyed when the protein is incubated in its unfolded state for a long time and is not re-formed during the refolding; and (iii) a misfolded nonfluorescent state exists, and this state kinetically competes with the mature fluorescent state.

Second, we observed double-exponential reaction curves in the refolding of GFP. GFP has nine proline residues. One of them is in the cis conformation in the native state, and the other eight residues are in the trans conformation. The proline isomerization reaction may explain why the double-exponential reaction is observed. In GFP, as GdnHCl concentration increases, the fast phase diminishes and the slow phase becomes dominant. Thus, at a GdnHCl concentration above 0.6 M, the refolding is a single-phase reaction. This is because the rate of the folding reaction of GFP becomes comparable to the rate of the proline isomerization reaction, and these reactions are kinetically coupled.

REFERENCES

1. Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996) *Science 273*, 1392−1395.
2. Yang, F., Moss, L. G., and Phillips, G. M. (1996) *Nat. Biotechnol. 14*, 1246−1252.
3. Cubitt, A. B., Heim, R., Adams, S. R., Boyd, A. E., Gross, L. A., and Tsien, R. Y. (1995) *Nat. Biotechnol. 13*, 151−154.
4. Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G., and Cormier, M. J. (1992) *Gene 111*, 229−233.
5. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. (1994) *Science 263*, 802−805.
6. Yamasaki, M., Hashiguchi, N., Fujiwara, C., Imanaka, T., Tsukamoto, T., and Osumi, T. (1999) *J. Biol. Chem. 274*, 35293−35296.
7. Heim, R., Prasher, D. C., and Tsien, R. Y. (1995) *Nature 373*, 663−664.
8. Heim, R., Prasher, D. C., and Tsien, R. Y. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 12501−12504.
9. Miesenbock, G., De Angelis, D. A., and Rothman, J. E. (1998) *Nature 394*, 192−195.
10. Makino, Y., Amada, K., Taguchi, H., and Yoshida, M. (1997) *J. Biol. Chem. 272*, 12468−12474.
11. Bokman, S. H., and Ward, W. W. (1981) *Biochem. Biophys. Res. Commun. 101*, 1372−1380.
12. Ward, W. W., and Bokman, S. H. (1982) *Biochemistry 19*, 4535−4540.
13. Crameri, A., Whiteborn, E. A., Tate, E., and Stemmer, W. P. C. (1996) *Nat. Biotechnol. 143*, 315−319.
14. Pace, C. N. (1986) *Methods Enzymol. 131*, 266−280.
15. Ward, W. W., Prentice, H. J., Roth, A. F., Cody, C. W., and Reeves, S. C. (1982) *Photochem. Photobiol. 35*, 803−808.
16. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) *Protein Sci. 11*, 2411−2423.
17. Hvidt, A., and Nielsen, S. O. (1966) *Adv. Protein Chem. 21*, 287−386.
18. Harper, J. D., and Lansbury P. T. (1997) *Annu. Rev. Biochem. 66*, 385−407.
19. Delagrave, S., Hawtin, R. E., Silva, C. M., Yang, M. M., and Youvan, D. C. (1995) *Nat. Biotechnol. 13*, 151−154.
20. Cormack, B. P., Valdivia, R. H., and Falkow, S. (1996) *Gene 173*, 33−38.
21. Heim, R., and Tsien, R. Y. (1996) *Curr. Biol. 6*, 178−182.
22. Siemering, K. R., Golbik, R., Sever, R., and Haseloff, J. (1996) *Curr. Biol. 6*, 1653−1663.
23. Kimata, Y., Iwaki, M., Lim, C. R., and Kohno, K. (1997) *Biochem. Biophys. Res. Commun. 232*, 69−73.
24. Cubitt, A. B., Woollenweber, L. A., and Heim, R. (1999) *Methods Cell Biol. 58*, 19−30.
25. Betts, S., Haase-Pettingell, C., and King, J. (1997) *Adv. Protein Chem. 50*, 243−264.
26. Wetzel, R (1997) *Adv. Protein Chem. 50*, 183−242.
27. Chan, W., Helms, L. R., Brooks, I., Lee, G., Ngola, S., McNulty, D., Maleeff, B., Hensley, P., and Wetzel, R. (1996) *Folding Des. 1*, 77−89.
28. Hurle, M. R., Helms, L. R., Li, L., Chan, W., and Wetzel, R. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 5446−5450.
29. Kelly, J. W. (1996) *Curr. Opin. Struct. Biol. 6*, 11−17.
30. Kelly, J. W., Colon, W., Lai, Z., Lashuel, H. A., McCulloch, J., McCutchen, S. L., Miroy, G. J., and Peterson, S. A. (1997) *Adv. Protein Chem. 50*, 161−181.
31. Funahashi, J., Takano, K., Ogasahara, K., Yamagata, Y., and Yutani, K. (1996) *J. Biochem. (Tokyo) 120*, 1216−1223.
32. Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C., and Pepys, M. B. (1997) *Nature 385*, 787−793.
33. Canet, D., Sunde, M., Last, A. M., Miranker, A., Spencer, A., Robinson, C. V., and Dobson, C. M. (1999) *Biochemistry 38*, 6419−6427.
34. Chrunyk, B. A., Evans, J., Lillquist, J., Young, P., and Wetzel, R. (1993) *J. Biol. Chem. 268*, 18053−18061.
35. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

BI000543L