

Modulating Protein Folding Rates *in Vivo* and *in Vitro* by Side-chain Interactions between the Parallel β Strands of Green Fluorescent Protein*

Received for publication, May 31, 2000, and in revised form, June 6, 2000
Published, JBC Papers in Press, June 7, 2000, DOI 10.1074/jbc.M004734200

Jane S. Merkel and Lynne Regan \ddagger §

From the Department of Molecular Biophysics and Biochemistry and \ddagger Chemistry, Yale University, New Haven, Connecticut 06520-8114

We have identified pairs of residues across the two parallel β strands of green fluorescent protein that facilitate native strand register of the surface-exposed β barrel. After constructing a suitable host environment around two guest residues, minimizing interactions of the guest residues with surrounding side-chains yet maintaining the wild-type protein structure and the chromophore environment, we introduced a library of cross-strand pairings by cassette mutagenesis. Colonies of *Escherichia coli* transformed with the library differ in intracellular fluorescence. Most of the fluorescent pairs have predominantly charged and polar guest site residues. The magnitude and the rate of fluorescence acquisition *in vivo* from transformed *E. coli* cells varies among the mutants despite comparable levels of protein expression. Spectroscopic measurements of purified mutants show that the native protein structure is maintained. Kinetic studies using purified protein with fully matured chromophores demonstrate that the mutants span a 10-fold range in folding rates with undetectable differences in unfolding rates. Thus, green fluorescent protein provides an ideal system for monitoring determinants of *in vivo* protein folding. Cross-strand pairings affect both protein stability and folding kinetics by favoring the formation of native strand register preferentially to non-native strand alignments.

demonstrate that residues serine 65, tyrosine 66, and glycine 67 cyclize autocatalytically and oxidize to form the chromophore (3, 4). No other cellular components are required for GFP fluorescence (5).

GFP emits its characteristic fluorescence when it is properly folded. Only soluble protein is fluorescent. GFP isolated from the inclusion body fraction of *Escherichia coli* cell lysates is nonfluorescent but can be refolded *in vitro* to generate the fluorophore (2, 6, 7). It has been suggested that GFP aggregates in inclusion bodies prior to productive folding (7). GFP variants have been selected for increased cellular fluorescence without altering the chromophore environment. These same variants reduce the fraction of protein localizing to inclusion bodies when expressed in *E. coli* relative to wild-type GFP (8–12). Positions of mutations that increase cellular fluorescence map to loops and to both centrally directed and surface-exposed β strand positions on the structure.

Because formation of mature GFP can be directly detected in cells, GFP provides an ideal system in which to study β sheet protein folding in the cellular milieu. The folding of β sheets is not well understood. Here, we focus on cross-strand pairings between the parallel strands on the solvent-exposed β barrel of GFP.

We propose that interactions on the surface of β sheets modulate the stability of β sheets. Statistical analyses of cross-strand residue pairings in proteins of known structure (13–15) distinctively demonstrate non-random pairing; moreover preferences depend on the type of cross-strand site. Antiparallel β sheets have two types of sites, one in which the backbones of both residues hydrogen-bond to one another and one in which backbone hydrogen-bonding groups are directed away from the cross-strand residue. The arrangement is different in parallel β sheets. There is only one type of pairing, but the two residues in the pair are not equivalent. One residue, at the hydrogen-bonding (HB) site, hydrogen-bonds to the backbone groups on both sides of its cross-strand residue, at the non-hydrogen-bonding (nonHB) site. Experimentally, it has been shown that amino acid side-chains interact across β strands in measurably favorable and unfavorable combinations (16–19). We propose cross-strand pairings in which both side chains can adopt their lowest energy side-chain conformation are favorable in β sheets. Energetically disfavored pairings result when one or both residues of the pair adopt higher energy side-chain conformations, a consequence of steric constraints of the particular cross-strand site (18). Thus, the type of site and the side-chain conformational preference influence favorable and unfavorable pairings in β sheets. In sum, all of the above studies demonstrate that cross-strand pairings modulate the stability of proteins and suggest favorable pairings for particular types of cross-strand sites.

The green fluorescent protein (GFP)¹ isolated from the jellyfish *Aequorea victoria* autocatalytically generates a covalently attached fluorophore when correctly folded and therefore fluoresces without external cofactors. This property permits the use of GFP in a wide range of applications (1). The 238-residue protein consists of a long, regular β barrel formed from eleven β strands, the first and sixth of which align in parallel orientation; the others are antiparallel. A single α helix passes through the center of the barrel (Fig. 1A). The fluorescent chromophore is covalently attached to the protein at the center of the helix inside the barrel, shielded from solvent within the hydrophobic core of the β barrel (2). Biochemical experiments

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Dept. of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Ave., New Haven, CT 0652–8114. Tel.: 203-432-9843; Fax: 203-432-5175; E-mail: lynne@csb.yale.edu.

¹ The abbreviations used are: GFP, green fluorescent protein; HB, hydrogen-bonding site of a parallel cross-strand pair; nonHB, non-hydrogen-bonding site of a parallel cross-strand pair; DTT, dithiothreitol.

To identify cross-strand pairings in parallel sheets that facilitate correct strand register and protein folding in a cellular environment, we introduced a library of pairings into GFP at two positions. We investigated a subset of these cross-strand pairings *in vivo* and *in vitro* to demonstrate the structural integrity of these mutants by several spectroscopic methods. By following the intrinsic fluorescence of native protein, we have observed the slow protein folding process within cells and variations present among selected mutants. These findings agree with *in vitro* observations from the purified variants. We propose a model for strand alignment in parallel β sheets, a problem highlighted by the nonconsecutive association of component β strands.

EXPERIMENTAL PROCEDURES

“Wild-type” GFP Construct and Cross-strand Variants—The GFP background chosen for this study incorporated the following mutations relative to the wild-type GFP isolated from jellyfish: three amino acid substitutions (F99S, M153T, V163A) identified through a selection experiment for the phenotype, which improves the yield of soluble fluorescent protein in *E. coli* (9), an historical polymerase chain reaction-induced mutation, Q80R (5), and the K238N mutation at the unstructured C terminus (20). All other positions are identical to wild-type. The variant containing these 5 mutations serves as “wild-type” in this study. Host site substitutions (L15T, D19T, S30A, E32A, E111A, V120T, E124T) were designed to create the environment around the cross-strand guest positions 17 and 122 in GFP (Fig. 1B) and were constructed by site-directed mutagenesis using standard protocols. DNA encoding a polyhistidine tag upstream of the initiation methionine was cloned between the *NheI* site in the gene and an upstream *NdeI* site on the pET11a plasmid (Novagen). Randomized oligonucleotide primers were used to generate the library of cross-strand pairings allowing all nucleotides at the first two positions and cytosine, guanine, and thymine at the third position for codon 17, thus eliminating two of three termination codons while maintaining the potential for all residues to be encoded. The primer for codon 122 had equimolar mixtures of all nucleotides at all three codon positions.

Expression in *E. coli*—Protein expression results from background induction in BL21 (DE3) cells (Novagen) in the absence of induction by isopropyl-1-thio- β -D-galactopyranoside. 6-ml starter cultures of freshly transformed cells in LB medium in culture tubes were grown for 4–5 h at 25–28 °C. All 6 ml of this inoculum was used per liter of growth medium. Growth continued for 60 h total at 25–28 °C with 200 rpm shaking speed.

Protein Purification—Cell pellets from growth cultures of 1–3 liters were lysed by sonication in 40–50 ml 20 mM Tris (pH 8.0), 100 mM NaCl. Lysate was centrifuged at 13 krpm, 4 °C for at least 30 min, and the supernatant was decanted for further purification. Purification with TALON metal affinity resin (CLONTECH) was performed according to the manufacturer’s instructions in 20 mM Tris, pH 8.0, and 100 mM NaCl at room temperature. Protein was eluted in two sequential 4-ml volumes of buffer with 250–500 mM imidazole. Nucleic acids (and imidazole) were removed in a final purification step by gel filtration chromatography, using a Superdex 75 column (Amersham Pharmacia Biotech) equilibrated in 20 mM Tris, pH 8.0, and 100 mM NaCl.

Spectroscopic Methods—Protein concentration was determined by the aromatic residue UV-vis absorption of acid-denatured protein at 280 nm (21). Protein concentrations were determined from the mean of triplicate measurements. The estimated error is $\pm 5\%$. All purified protein samples were measured in 20 mM Tris (pH 8.0), 100 mM NaCl. UV-vis measurements were performed on a Hewlett Packard 845X UV-visible Chemstation on 10 μ M protein samples. CD spectra were recorded on an Aviv 62DS circular dichroism spectrometer (Aviv Instruments). Far-UV CD spectra were acquired from 190–260 nm with 0.5 nm sampling. Spectra were recorded on 30 μ M protein samples in a quartz cuvette with a 0.1-cm path length. Fluorescence measurements were recorded on a Hitachi F-4500 fluorescence spectrophotometer with the photomultiplier tube set at 700 V. For *in vivo* measurements, excitation (5.0 nm slit width) was set at 397 nm, and emission (2.5 nm slit width) was monitored from 410–700 (value reported for 509 nm). *In vitro* measurements were performed on 100 μ M protein with both excitation and emission slit widths set to 5 nm and photomultiplier tube voltage set at 700 V. Excitation and emission of the chromophore was measured at 397 and 550 nm (so that the minor excitation peak at 475 nm could be resolved), respectively.

In Vivo Folding Measurements—Freshly transformed *E. coli* cells

were used to inoculate 3-ml cultures of LB medium supplemented with 50 μ g/ml ampicillin. Culture tubes were incubated at 25–28 °C with 200 rpm agitation for 60 h because of the slow folding/maturation time of some of the mutants. A 100- μ l aliquot was taken from the culture and centrifuged in an Eppendorf tube at 13 krpm for 2 min. LB medium was pipetted off. This step removes any extracellular GFP in the medium resulting from cell death or leaking from cells as observed in some *E. coli* strains. At the same time, removing the medium eliminates background fluorescence from LB so that fluorescence measurements report only intracellular GFP levels. 1 ml of phosphate-buffered saline with final concentrations of 137 mM NaCl, 2.7 mM KCl, 4.3 mM $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$, and 1.4 mM KH_2PO_4 (~pH 7.3) (22) was added to the tube, and cells were resuspended and frozen at –20 °C until measurement. For experiments in which samples were taken every 2 h during a time course, 100 μ l of starter culture (A_{600} at 2) was used to inoculate 6 ml of LB medium (with ampicillin) in a culture tube. 100- μ l aliquots were taken and prepared as described above.

In Vitro Folding Measurements—100 nM protein was equilibrated in 100 mM citrate buffer over a pH range of 1.8 to 8.2, 100 mM NaCl, and 1 mM DTT at room temperature for 16 h. Both chromophore fluorescence and aromatic residue fluorescence were measured by emission at 509 nm from excitation at 397 nm and 280 nm, respectively. Prior to refolding, protein was fully unfolded in 100 mM citrate (pH 1.93) and 100 mM NaCl with DTT present. This mixture was then diluted 107.5-fold to a final buffer of 20 mM Tris (pH 8.0), 100 mM NaCl, and 1 mM DTT. At this pH, all GFP variants are fully folded (Fig. 4A). Because there are 11 prolines in GFP, one of which is a *cis*-proline in the native structure (P89), unfolding was carried out immediately before refolding experiments. Refolding was monitored by chromophore fluorescence at 509 nm. All measurements were performed in triplicate. Refolding spectra fit single exponential growth curves with an *R* of 0.99 or better. The quality of the fit was lower when the data were fit to higher exponentials. Protein unfolding was monitored by loss of chromophore fluorescence at 509 nm. Protein was diluted 250-fold from pH 8.0 into 1 ml of 100 mM citrate buffer (pH 4.35), 100 mM NaCl, 1 mM DTT to a final protein concentration of 100 nM. At this pH, all GFP variants are fully unfolded (Fig. 4A). All measurements were performed in triplicate. 70–80% of this transition occurs within the dead time of manual mixing techniques (within 12 s). Unfolding curves were fit by a single exponential rate with an *R* of 0.98 or better. Higher exponentials decreased the quality of the fit.

RESULTS

Host-Guest Environment between Parallel β Strands on Surface Preserves Tertiary Structure of GFP—We created a double guest site across the two parallel β strands on the solvent-exposed face of the β barrel of GFP and minimized possible interactions between the guest residues and the surrounding side chains on the protein surface (Fig. 1B), in accordance with previous β sheet host-guest studies (16–19, 23, 24). None of the residues on the two parallel β strands of GFP has been identified by other studies to alter the folding properties or the chromophore environment of the protein. Positions in the center of the first and sixth strands sample a homogeneous β sheet environment and minimize turn effects. The $17_{\text{HB}}122_{\text{nonHB}}$ double-guest site was chosen because it best represents the solvent accessibilities of all types of β sheets (both barrel and sheet arrangements). The host site with alanine residues at both guest site positions (E17A, R122A) constitutes the reference state for this study, Ala¹⁷–Ala¹²² (Fig. 1B). Ala¹⁷–Ala¹²² was expressed in *E. coli* cells, and the fluorescence was compared with wild-type. Both types of colonies fluoresce after shifting the cells to room temperature for approximately 1 day, consistent with previous observations regarding the thermosensitivity of GFP (25). Cellular fluorescence of cells expressing Ala¹⁷–Ala¹²² has the same absorption and emission spectra as wild-type, indicating that substitutions on the protein surface creating the host-guest site maintains the tertiary environment of the buried chromophore in GFP. Having created an environment on the surface of GFP without altering the structure of folded protein, we incorporated a library of amino acid pairings at the two guest positions in GFP and monitored the fluorescence of intact *E. coli* cells.

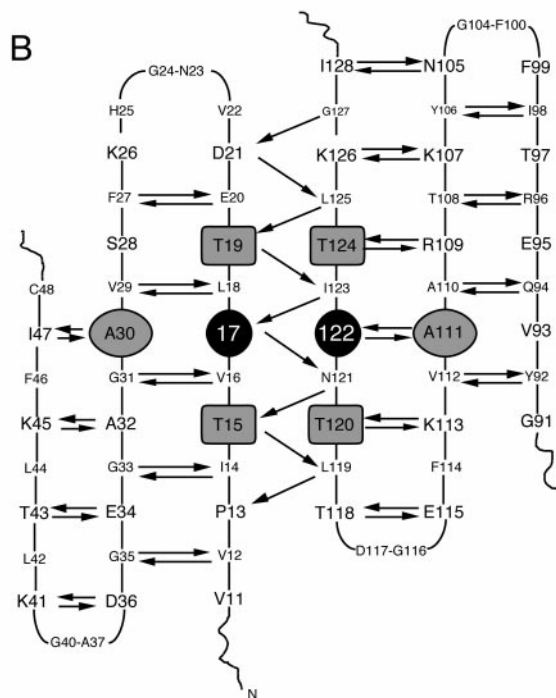
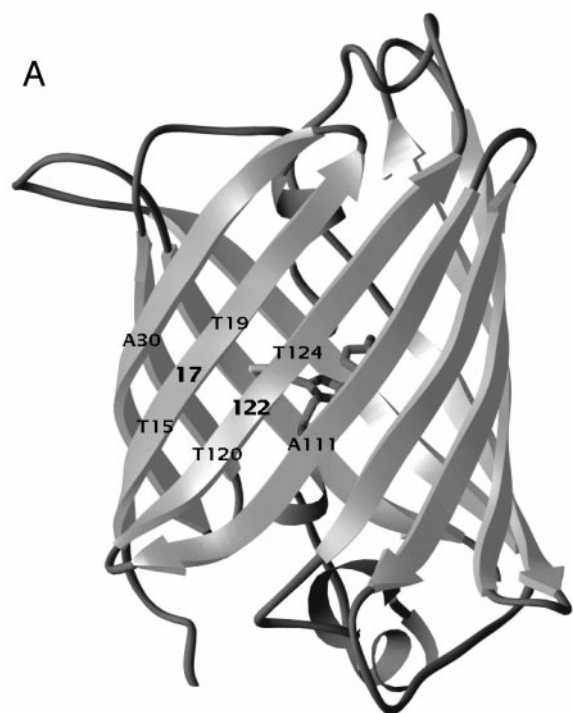


FIG. 1. GFP with the host-guest site for cross-strand selection at positions 17 and 122 between parallel β strands. A, ribbon representation with the fluorophore represented by a ball and stick model at the center of the helix. Figure was made in Ribbons (36) from the Palm *et al.* (20) GFP crystal structure (deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank, code 1emk). B, schematic representation of the hydrogen-bonding pattern of β strands 1 through 6.

Cross-strand Pairings on Surface of GFP Alter Level of Cellular Fluorescence—Substitution of all amino acids at both guest positions (17 and 122) was achieved by ligating inserts encoding residues 2–138 (including both guest site positions) from cassette mutagenesis using degenerate oligonucleotide primers into a pET11a vector containing the remaining resi-

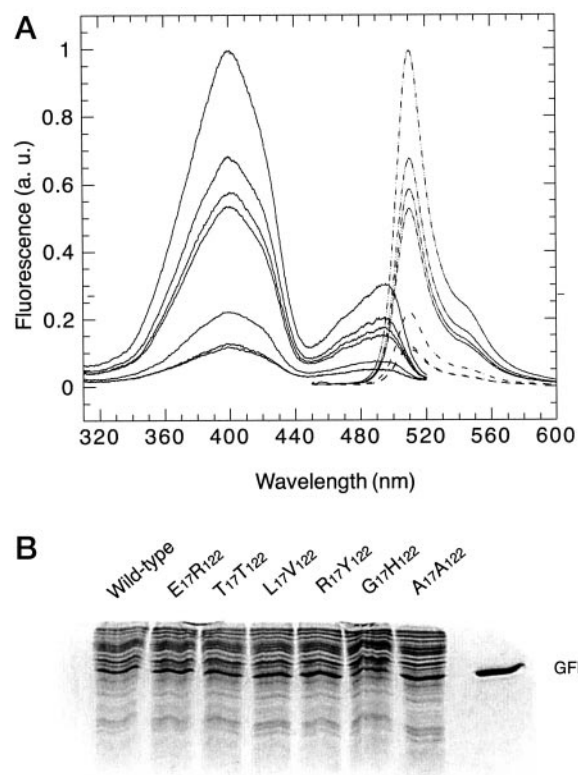


FIG. 2. Fluorescence properties of GFP variants in *E. coli* cells. Cells were grown at 25 °C for 60 h. A, fluorescence properties of GFP variants in *E. coli*, in order of decreasing excitation (solid lines) and emission (dashed lines); wild-type, Glu¹⁷-Arg¹²², Thr¹⁷-Thr¹²², Leu¹⁷-Val¹²², Arg¹⁷-Tyr¹²², Gly¹⁷-His¹²², Ala¹⁷-Ala¹²². B, total protein expression of GFP variants in *E. coli*. Purified wild-type GFP standard (far right lane) marks the location of GFP in the lysates.

dues of the GFP gene. Transformed *E. coli* differed in the intensity of green fluorescence on a Petri dish when grown at 37 °C and transferred to room temperature for at least 24 h. Approximately half of the 200 colonies were nonfluorescent. Qualitatively, 10% of the colonies were less fluorescent than Ala¹⁷-Ala¹²², 20% were equivalent to Ala¹⁷-Ala¹²², and a final 20% were more fluorescent than Ala¹⁷-Ala¹²², displaying nearly wild-type fluorescence. No colonies fluoresced as brightly as wild-type GFP, suggesting that seven of nine substitutions introduced to create the host-guest site result in decreased fluorescence. The genes of GFP variants were sequenced to determine the identity of guest site residues and to eliminate clones from the study that have a mutation at any site on the gene. Observed differences in fluorescence may result from protein expression, protein solubility, or protein stability.

The fluorescence properties of intact cells expressing different GFP variants were compared. All cells gave the characteristic GFP excitation and emission profiles confirming the maintenance of chromophore environment (Fig. 2A). One explanation for fluorescence differences among GFP variants is the amount of soluble, folded protein in cells, despite comparable levels of total protein expression (Fig. 2B). GFP was prominent in all of the lysates, and the level of GFP expression was comparable. Hence, fluorescence differences do not result from large differences in protein expression.

Based on whole cell fluorescence, the GFP variants displayed between 10 and 60% of wild-type fluorescence (Table I). The highest fluorescing colonies have predominantly polar and charged residues at the guest sites. Hydrophobic, aromatic, and β -branched residues, which have high β sheet propensities, are

TABLE I
Relative whole cell fluorescence of cross-strand pairs in GFP

Residue pair (17 _{HB} 122 _{nonHB})	Relative whole cell fluorescence (% wild-type GFP)
Wild-type (ER)	100
RT	57
AT	49
ER	49
NV	43
AR	42
SR	42
MT	41
QH	40
AI	40
NR	40
TT	40
LV	38
AC	36
AN	35
NI	35
RI	33
RN	32
MR	32
RV	31
RY	28
SH	28
QL	26
QS	26
EH	26
ED	25
WH	25
KH	25
RH	23
MS	23
MQ	22
WR	12
AA	12
GH	11

not observed frequently, except for threonine, which is found at both guest positions and in a number of different pairings. The observed amino acid pairings at the solvent-exposed site indicate a preference for polar and charged residues, which may enhance the solubility of GFP. Hydrophobic residues are also observed among the best pairings (alanine, isoleucine, valine) including the hydrophobic pair, Leu¹⁷-Val¹²², suggesting that solubility is not the sole determinant of favored pairings. GFP is a much larger protein than other peptides or proteins in which cross-strand pairings have been studied (16–19). Therefore, residues unfavorable for β sheet formation substituted into a protein may destabilize GFP to a lesser extent than a smaller protein, which may further account for the diversity of amino acids seen at the guest sites.

Cross-strand pairings capable of a number of possible types of side-chain interactions include Thr¹⁷-Thr¹²², in which threonine residues may hydrogen-bond to one another or to nearby host-site residues, possible cation- π (26) interacting pairs (Lys¹⁷-His¹²², Arg¹⁷-His¹²², Arg¹⁷-Tyr¹²², and Trp¹⁷-Arg¹²²), a π - π pair (Trp¹⁷-His¹²²), and a salt-bridge, Glu¹⁷-Arg¹²². Pairs of residues with complementary polarity (Glu¹⁷-His¹²², Met¹⁷-Arg¹²², Arg¹⁷-Ser¹²², Arg¹⁷-Thr¹²²) were selected. Rare codon usage at positions 17 or 122 did not significantly influence overall protein expression as several of the most fluorescent clones contained rare codon sequences. Cross-strand pairings at two positions on GFP dramatically alter the magnitude of intracellular fluorescence.

Cross-strand Pairings Affect the Rate of GFP Maturation in Cells—A few representative clones were chosen for further characterization based on the diverse chemical natures of their cross-strand pairings. Glu¹⁷-Arg¹²² was chosen to assess the effect of host-site substitutions in GFP (wild-type contains Glu¹⁷-Arg¹²² but is surrounded by the naturally occurring

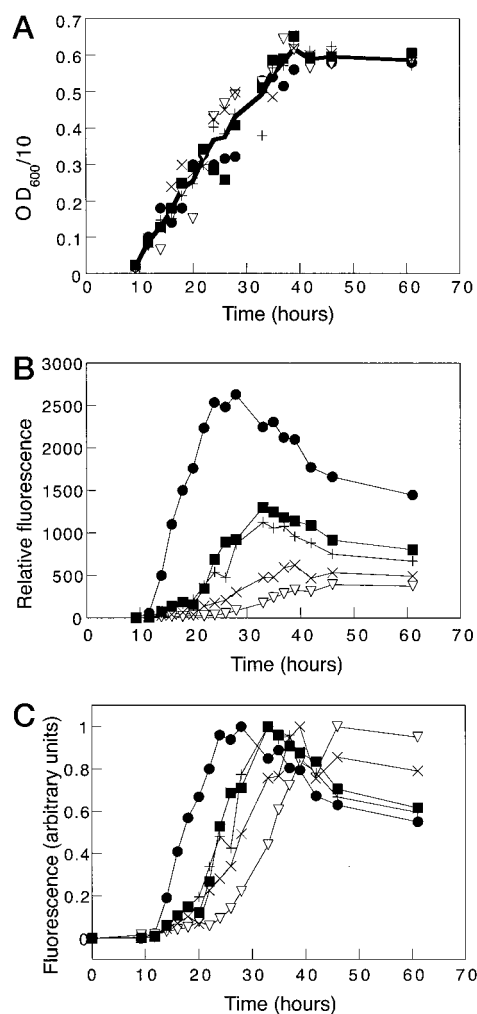


FIG. 3. Fluorescence acquisition of GFP variants *in vivo*. A, growth curves for cultures expressing wild-type (●), Glu¹⁷-Arg¹²² (■), Thr¹⁷-Thr¹²² (+), Arg¹⁷-Tyr¹²² (×), and Ala¹⁷-Ala¹²² (∇). B, total fluorescence emission at 509 nm for 1:10 dilution of *E. coli* cells expressing different GFP variants. C, relative fluorescence emission for 1:10 dilution of *E. coli* cells expressing different GFP variants.

residues) because it contains a salt bridge. Other highly fluorescing colonies chosen include the polar pair, Thr¹⁷-Thr¹²², and the hydrophobic pair, Leu¹⁷-Val¹²². In the middle of the whole cell fluorescence scale, Arg¹⁷-Tyr¹²², a possible cation- π pair was chosen. The glycine-containing pair, Gly¹⁷-His¹²², was chosen to represent a poorly fluorescing clone. These variants were compared with wild-type and Ala¹⁷-Ala¹²². Cultures were inoculated with equal amounts of cells, and there were no differences in *E. coli* growth rates for cultures expressing different GFP variants (Fig. 3A). Fluorescence measurements showed pronounced variation in final fluorescence levels depending on the cross-strand pair (Fig. 3B). Scaled relatively (Fig. 3C), the variation in the rate at which fluorescence appears among mutants is clear. GFP matures slowly, and fluorescence from cultures grown at room temperature peaks at 24 h for wild-type and between 30 and 48 h for the variants. The variants with the highest final fluorescence (Glu¹⁷-Arg¹²², Thr¹⁷-Thr¹²²) mature more quickly than that with intermediate fluorescence (Arg¹⁷-Tyr¹²²) and the poorly fluorescing variant, Ala¹⁷-Ala¹²². The fluorescence for all variants, which does not appear to be caused by cell death, decreases at the longest time points as the optical density of the samples plateaus at 40 h and does not further decrease even at the longest time points (Fig. 3A). The fluorescence decrease may arise from

protein degradation at long culture times. Both the final fluorescence level and the rate of fluorescence vary in cells depending on the residues paired across parallel β strands on the surface of GFP.

Spectral Properties of Purified GFP Variant Proteins Are Identical—The fluorescence properties of soluble protein purified from *E. coli* cells expressing GFP cross-strand variants were determined as a sensitive probe of chromophore environment. The excitation and emission profiles of the variants are identical. The major emission peak at 509 nm and shoulder at 540 nm confirm that the wild-type environment around the chromophore is maintained in all of the cross-strand variants. The fluorescence of the single intrinsic tryptophan at position 57, investigated by excitation at 294 nm, results in a large emission peak at 509 nm with a shoulder at 540 nm, the signature of the chromophore fluorescence. The structure of GFP shows the tryptophan to lie 11 Å above the chromophore within the β barrel, and tryptophan emission is absorbed by the nearby chromophore, consistent with previous findings (27). The fluorescences of GFP variants are indistinguishable from one another in response to tryptophan absorption, providing further evidence for structural conservation within the protein core.

The absorption spectra of these variants in the ultraviolet and visible region are superimposable. The peak at 280 nm corresponds to aromatic residues in the protein. The chromophore absorption peaks at 397 and 475 nm are also prominent features in the spectrum. The CD spectrum in the far-UV region is indicative of a β -sheet protein with a minimum at 217 nm. Fluorescence, absorption, and CD spectroscopies indicate that the wild-type chromophore site, aromatic residue environments, and overall backbone structure are preserved in the cross-strand variants. The variant proteins are spectroscopically indistinguishable from wild-type.

Cross-strand Pairings Affect Folding Rates in Vitro—Differences in the rate of fluorescence acquisition among selected GFP variants within cells may reflect differences in the folding rates of GFP. Therefore, we examined the kinetics of refolding and unfolding of the purified proteins. The chromophore is covalently bound to the native protein. By performing refolding studies, *in vitro* measurements isolate the folding event from chromophore maturation. Furthermore, the possible effects of cellular components that affect folding (*e.g.* chaperones) are eliminated. GFP has been shown to reversibly refold after either acid- or alkali-induced denaturation (28, 29). We confirmed that the selected GFP variants display reversible folding behavior under these conditions. Fig. 4A shows the equilibrium denaturation of the acid-induced unfolding transition for wild-type and selected GFP variants. The denaturation is cooperative. Loss of fluorescence followed by chromophore or tryptophan absorption has the same pH profile. We compared the rates of refolding and unfolding of purified GFP mutants at conditions where all variants can be completely refolded or unfolded, respectively. Representative traces for each variant of refolding (Fig. 4B) and unfolding (Fig. 4C) followed by chromophore fluorescence fit to a single exponential. The refolding rate constants of purified variants (Table II) correlate with *in vivo* measurements and vary nearly an order of magnitude among the mutants. The unfolding spectra display small differences among wild-type GFP and selected variants and therefore have similar unfolding rates (Table II).

Cross-strand Pairings Modulate GFP Stability—Refolding and unfolding rate constants were used to calculate relative stability differences ($\Delta\Delta G_u$) between the double-alanine reference mutant, Ala¹⁷-Ala¹²², and selected variants for the acid-induced unfolding transition (Table III). For this calculation,

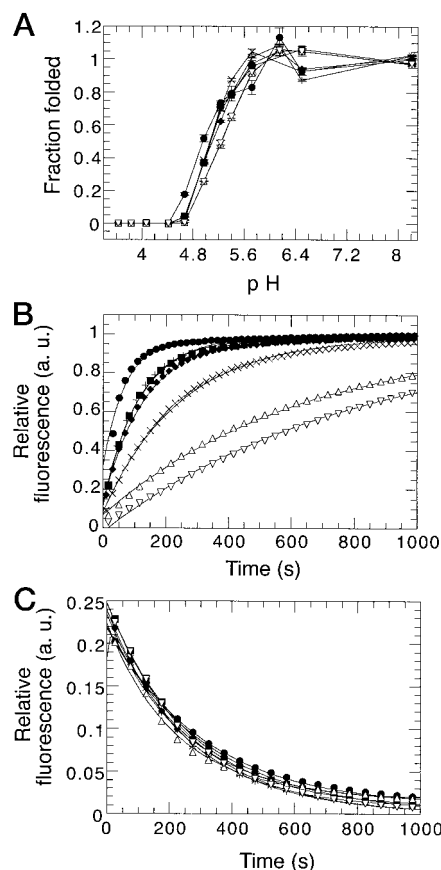


FIG. 4. Folding and unfolding of GFP variants *in vitro* followed by fluorescence. A, equilibrium acid denaturation of wild-type (●), Glu¹⁷-Arg¹²² (■), Thr¹⁷-Thr¹²² (+), Leu¹⁷-Val¹²² (◆), Arg¹⁷-Tyr¹²² (×), Gly¹⁷-His¹²² (△), and Ala¹⁷-Ala¹²² (∇). Protein samples were equilibrated in 100 mM citrate buffer, 100 mM NaCl, and 1 mM DTT at room temperature. B, refolding of wild-type GFP and selected variants; C, unfolding of wild-type GFP and selected variants. Refolding spectra were measured in 20 mM Tris (pH 8.0), 100 mM NaCl, 1 mM DTT at room temperature. Unfolding spectra were measured in 100 mM citrate buffer (pH 4.35), 100 mM NaCl, 1 mM DTT at room temperature. The lines represent a single-exponential fit for each spectrum.

TABLE II
Refolding and unfolding rate constants of selected GFP variants

Variant	Refolding rate constant ^a	Unfolding rate constant ^b
	s^{-1}	s^{-1}
Glu ¹⁷ -Arg ¹²²	0.0083 ± 0.0006	0.0038 ± 0.0003
Thr ¹⁷ -Thr ¹²²	0.0075 ± 0.0005	0.0037 ± 0.0001
Leu ¹⁷ -Val ¹²²	0.0070 ± 0.0002	0.0041 ± 0.0001
Arg ¹⁷ -Tyr ¹²²	0.0036 ± 0.0002	0.0040 ± 0.0001
Gly ¹⁷ -His ¹²²	0.00152 ± 0.00005	0.0047 ± 0.0007
Ala ¹⁷ -Ala ¹²²	0.0014 ± 0.0003	0.0038 ± 0.0003

^a Refolding was performed on protein denatured at pH 1.93, which was diluted 110-fold into 20 mM Tris (pH 8.0), 100 mM NaCl, 1 mM DTT.

^b Unfolding was performed on protein by 250-fold dilution into 100 mM citrate (pH 4.35), 100 mM NaCl, 1 mM DTT. Wild-type GFP refolds with a rate constant of 0.017 ± 0.001 (s^{-1}) and unfolds with a rate constant of 0.0031 ± 0.0003 (s^{-1}).

we assume that the large variation in refolding rate constants accounts for the majority of the stability differences, and the small range of the unfolding rate constants is conserved throughout the pH range. Cross-strand pairings modulate the stabilities of the variants within 1 kcal mol⁻¹, and the charged pair, Glu¹⁷-Arg¹²² is the most stabilizing, consistent with previous studies investigating cross-strand pairings in β sheet proteins (16–19). The relative stabilities of mutants result from both the intrinsic β -sheet propensities of residues at the two guest sites and the effects of cross-strand interactions between

TABLE III
 $\Delta\Delta G_u$ of the unfolding transition between $A_{17}A_{122}$ and selected GFP variants

$\Delta\Delta G_u^{25^\circ\text{C}}$ was calculated using the equation $\Delta\Delta G_u = -RT\ln(kf_{AA}/ku_{AA} \times ku_{XY}/kf_{XY})$. kf is the folding rate constant, and ku is the unfolding rate constant (Table II). AA denotes the Ala¹⁷-Ala¹²² reference, and XY denotes a cross-strand variant in GFP, calculated at 25 °C ($T = 298$ K). The $\Delta\Delta G_u^{25^\circ\text{C}}$ of wild-type GFP relative to Ala¹⁷-Ala¹²² is 1.6 kcal mol⁻¹.

Variant	$\Delta\Delta G_u^{25^\circ\text{C}}$ kcal mol ⁻¹
Glu ¹⁷ -Arg ¹²²	1.05 ± 0.06
Thr ¹⁷ -Thr ¹²²	1.01 ± 0.06
Leu ¹⁷ -Val ¹²²	0.91 ± 0.06
Arg ¹⁷ -Tyr ¹²²	0.5 ± 0.1
Gly ¹⁷ -His ¹²²	-0.1 ± 0.2
Ala ¹⁷ -Ala ¹²²	0 ± 0.2

the two guest residues. Between parallel β strands on the solvent-exposed barrel of GFP, cross-strand pairings modulate both the protein stability and the folding rate, without strongly affecting the unfolding rate.

DISCUSSION

Cross-strand Pairings Specify Folding or Aggregation in Slow Folding Proteins—Chromophore maturation is proposed to be the last stage in the protein folding pathway of GFP (7), occurring subsequent to protein folding. By studying a library of GFP mutants with different residue pairings across the parallel β strands, we observe differences in total cellular fluorescence. Despite differences in the magnitude of fluorescence among GFP variants *in vivo*, similarity by fluorescence, absorption, and CD spectroscopies *in vitro* suggests that the amount of mature protein differs in cells and not the structure of variant proteins. Hence, intracellular fluorescence monitors the soluble fraction of GFP expressed within cells. Total protein expression is comparable (Fig. 2B); thus mutants have different ratios of protein localizing to soluble and inclusion-body fractions. *In vitro* refolding experiments measure dramatic differences in the folding rate among GFP variants and suggest that faster folding rates within cells expressing GFP allow for greater total fluorescence. The fraction of the protein molecules that fold early are protected from aggregation and are capable of forming the fluorophore. Furthermore, we demonstrate that substitutions at two cross-strand positions can have a dramatic effect on the folding rate of a protein *in vitro*, which correlates with protein maturation within cells.

Model of the Folding Pathway Modulated by Cross-strand Interactions in Vivo—Combining thermodynamic and kinetic data, we analyze how cross-strand pairings affect the formation of fluorescent native protein from the “unfolded” state. However, we cannot distinguish between an unfolded state of the protein and an intermediate from *in vivo* or *in vitro* results. The differences in folding rates suggest that favorable pairings destabilize the unfolded state. In the simplest possible model, a destabilization of the unfolded state without changes in the transition state or folded state accounts for both the increase in stability and the faster folding rates of favorable cross-strand pairs. Disulfide bonds present in protein variants of T4 lysozyme have been shown to destabilize the unfolded state by reducing the entropy of the unfolded state (30–32). This effect is magnified in examples where disulfides restrict longer loops. Another example of a decrease in entropy destabilizing the unfolded state is observed in protein variants with glycine replacements to any other residue or replacements of any residue by the more constrained proline (33, 34). Hence, a destabilization in the unfolded state due to loss of entropy can account both for the stability and the folding rate differences in GFP observed for favorable cross-strand pairings between par-

allel β strands relative to Ala¹⁷-Ala¹²².

A more complex model includes the contributions of favorable pairings in stabilizing the folded state. Both destabilizing the unfolded state and stabilizing the folded state were seen in favorable single substitutions in the B1 domain of IgG-binding protein G (35). Because unfolding rates are relatively conserved among variants, it is possible that cross-strand interactions, which stabilize the folded state, are also present in the transition state, reducing its energy by a similar magnitude in both states. These stabilizations conserve the height of the unfolding barrier yet lower the folding barrier. Stability increases in favorable pairings are influenced by destabilization in the unfolded state and favorable interactions present in the folded state. Alternatively, interactions that stabilize both the transition state and the folded state without affecting the stability of the unfolded state could also account for the observed results.

We have selected variants at two cross-strand sites between the parallel β strands on the solvent-exposed β barrel of GFP. Neither mutations at these two positions nor mutations at the closest surrounding residues alter the wild-type spectral features of the covalently bound fluorophore in the core of the protein. Residue pairings at the cross-strand site alter the total fluorescence inside *E. coli* cells, which correlates with the rate of fluorescence acquisition from mature GFP. Proteins with favorable cross-strand pairings fold more quickly than those with less favorable pairings to achieve a higher final fluorescence. Unfavorable pairings result in unfolded or partially folded protein aggregates before becoming fluorescent. The *in vivo* folding rates correlate with those observed *in vitro* from studies with purified proteins. We propose that correct β -strand register between nonlocal sequences in β sheet proteins results from energetically preferred pairings, which fold more quickly than less preferred pairings. Conversely, arrangements with incorrect β strand alignments are less stable and fold slowly. Thus, the native strand alignment, having the kinetic and thermodynamic advantage, forms in preference to non-native strand alignments to specify correct register between β strands.

REFERENCES

- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. (1994) *Science* **263**, 802–805
- Ormö, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996) *Science* **273**, 1392–1395
- Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G., and Cormier, M. J. (1992) *Gene (Amst.)* **111**, 229–233
- Cody, C. W., Prasher, D. C., Westler, W. M., Prendergast, F. G., and Ward, W. W. (1993) *Biochemistry* **32**, 1212–1218
- Tsien, R. Y. (1998) *Annu. Rev. Biochem.* **67**, 509–544
- Heim, R., Prasher, D. C., and Tsien, R. Y. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 12501–12504
- Reid, B. G., and Flynn, G. C. (1997) *Biochemistry* **36**, 6786–6791
- Cormack, B. P., Valdivia, R. H., and Falkow, S. (1996) *Gene (Amst.)* **173**, 33–38
- Cramer, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. C. (1996) *Nature Biotechnol.* **14**, 315–319
- Siemering, K. R., Golbik, R., Sever, R., and Haseloff, J. (1996) *Curr. Biol.* **6**, 1653–1663
- Patterson, G. H., Knobel, S. M., Sharif, W. D., Kain, S. R., and Piston, D. W. (1997) *Biophys. J.* **73**, 2782–2790
- Stauber, R. H., Horie, K., Carney, P., Hudson, E. A., Tarasova, N. I., Gaitanaris, G. A., and Pavlakis, G. N. (1998) *BioTechniques* **24**, 462–471
- Wouters, M. A., and Curmi, P. M. G. (1995) *Proteins* **22**, 119–131
- Hutchinson, E. G., Sessions, R. B., Thornton, J. M., and Woolfson, D. N. (1998) *Protein Sci.* **7**, 2287–2300
- Cootes, A. P., Curmi, P. M. G., Cunningham, R., Donnelly, C., and Torda, A. E. (1998) *Proteins* **32**, 175–189
- Smith, C., and Regan, L. (1995) *Science* **270**, 980–982
- Merkel, J. S., and Regan, L. (1998) *Folding & Design* **3**, 449–455
- Merkel, J. S., Sturtevant, J. M., and Regan, L. (1999) *Structure* **7**, 1333–1343
- Blasie, C. A., and Berg, J. M. (1997) *Biochemistry* **36**, 6218–6222
- Palm, G. J., Zdanov, A., Gaitanaris, G. A., Stauber, R., Pavlakis, G. N., and Wlodawer, A. (1997) *Nat. Struct. Biol.* **4**, 361–365
- Gill, S. C., and von Hippel, P. H. (1989) *Anal. Biochem.* **182**, 319–326
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., (eds) (1992) *Short Protocols in Molecular Biology*, 2nd Ed., Greene Publishing Associates and John Wiley & Sons, New York

23. Minor, D. L. J., and Kim, P. S. (1994) *Nature* **367**, 660–663
24. Minor, D. L. J., and Kim, P. S. (1994) *Nature* **371**, 264–267
25. Lim, C. R., Kimata, Y., Oka, M., Nomaguchi, K., and Kohno, K. (1995) *J. Biochem. (Tokyo)* **118**, 13–17
26. Ma, J. C., and Dougherty, D. A. (1997) *Chem. Rev.* **97**, 1303–1324
27. van Thor, J. J., Pierik, A. J., Nugteren-Roodzant, I., Xie, A., and Hellingwerf, K. J. (1998) *Biochemistry* **37**, 16915–16921
28. Bokman, S. H., and Ward, W. W. (1981) *Biochem. Biophys. Res. Commun.* **101**, 1372–1380
29. Ward, W. W., and Bokman, S. H. (1982) *Biochemistry* **21**, 4535–4540
30. Matsumura, M., Becktel, W. J., Levitt, M., and Matthews, B. W. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 6562–6566
31. Matsumura, M., Signor, G., and Matthews, B. W. (1989) *Nature* **342**, 291–293
32. Matsumura, M., and Matthews, B. W. (1991) *Methods Enzymol.* **202**, 336–356
33. Matthews, B. W., Nicholson, H., and Becktel, W. J. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 6663–6667
34. Nicholson, H., Tronrud, D. E., Becktel, W. J., and Matthews, B. W. (1992) *Biopolymers* **32**, 1431–1441
35. Smith, C. K., Bu, Z., Anderson, K. S., Sturtevant, J. M., Engelman, D. M., and Regan, L. (1996) *Protein Sci.* **5**, 2009–2019
36. Carson, M., and Bugg, C. E. (1986) *J. Mol. Graphics* **4**, 121–122