

Engineering and characterization of a superfolder green fluorescent protein

Jean-Denis Pédelacq¹, Stéphanie Cabantous¹, Timothy Tran², Thomas C Terwilliger¹ & Geoffrey S Waldo¹

Existing variants of green fluorescent protein (GFP) often misfold when expressed as fusions with other proteins. We have generated a robustly folded version of GFP, called 'superfolder' GFP, that folds well even when fused to poorly folded polypeptides. Compared to 'folding reporter' GFP, a folding-enhanced GFP containing the 'cycle-3' mutations and the 'enhanced GFP' mutations F64L and S65T, superfolder GFP shows improved tolerance of circular permutation, greater resistance to chemical denaturants and improved folding kinetics. The fluorescence of *Escherichia coli* cells expressing each of eighteen proteins from *Pyrobaculum aerophilum* as fusions with superfolder GFP was proportional to total protein expression. In contrast, fluorescence of folding reporter GFP fusion proteins was strongly correlated with the productive folding yield of the passenger protein. X-ray crystallographic structural analyses helped explain the enhanced folding of superfolder GFP relative to folding reporter GFP.

Wild-type green fluorescent protein (GFP) misfolds when expressed in *E. coli*¹. Better-folded variants^{1–3} of GFP are widely employed as protein fusion tags^{1,4–8}, but the fused proteins can reduce the folding yield and fluorescence of these GFPs^{9–13}. Circularly permuted variants of GFP, as well as biosensors using fused receptor domains and engineered GFPs, tend to aggregate and misfold when expressed in *E. coli*^{14–16}. A more robustly folded version of GFP would be very useful.

Despite considerable effort aimed at improving the folding of GFP expressed alone^{1–3}, available GFP variants fold well and are brightly fluorescent only when expressed alone or when fused to well-folded proteins^{9–12}. The resistance of folded GFP to chemical and thermal denaturation¹ makes the selection of even more robustly folded GFP variants challenging. We reasoned that even better-folded versions of GFP could be identified by expressing libraries of GFP variants as fusions with an N-terminal, poorly folded, 'bait' polypeptide that interferes with the productive folding of the fused GFP moiety⁹, and then selecting brighter clones corresponding to GFP variants that can still fold. We used this approach to generate a very robustly folded version of GFP, superfolder GFP. The fluorescence of superfolder GFP fusion is unaffected by fusion partner misfolding and is directly proportional to total expression regardless of the solubility of the fusion, making superfolder GFP fluorescence a robust reporter of fusion protein expression. X-ray crystallographic studies of both superfolder GFP and folding reporter GFP reveal a five-membered ion-pair network in the superfolder GFP structure, mediated by the mutation S30R. This helps explain why the mutation S30R substantially improves the folding robustness of superfolder GFP.

RESULTS

Generating fluorescent proteins with enhanced folding robustness

Starting with the folding reporter GFP⁹, a well-folded GFP variant bearing the 'cycle-3' mutations¹⁷ F99S, M153T, V163A, and the 'enhanced GFP' mutations¹⁸ F64L and S65T, we screened libraries of folding reporter GFP variants as C-terminal fusions to poorly folded, bullfrog red-cell, H-subunit ferritin, an insoluble protein when expressed alone in *E. coli* at 37 °C⁹. Colonies expressing the ferritin–folding reporter GFP fusion at 37 °C showed very faint fluorescence⁹. After four rounds of DNA shuffling during which we picked ever brighter fluorescent clones expressed at 37 °C, we obtained the highly fluorescent, ferritin–superfolder GFP fusion. Superfolder GFP contains the folding reporter mutations and six new mutations: S30R, Y39N, N105T, Y145F, I171V and A206V (Fig. 1).

Liquid cultures of *E. coli* BL21(DE3) cells expressing ferritin–superfolder GFP fusions were ~50-fold more fluorescent than cells expressing ferritin–folding reporter GFP at 37 °C (Table 1). *E. coli* cells expressing superfolder GFP alone were approximately twofold more fluorescent than cells expressing folding reporter GFP alone (see 'controls' in Table 2). We also expressed six variants of folding reporter GFP at 27 °C or 37 °C, each bearing one of the single-point mutations derived from superfolder GFP as C-terminal fusions with ferritin (Table 1). Each of the mutations increased the amount of folded fluorescent GFP fused to poorly folded ferritin relative to the folding reporter GFP fusion (Table 1). The fluorescence emission spectra of superfolder GFP and folding reporter GFP were superimposable (Supplementary Fig. 1a online). Fluorescence excitation and ultraviolet-visible spectra of folding reporter GFP each peaked at 490 nm, whereas the corresponding superfolder GFP spectra peaked

¹Bioscience Division, MS-M888, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. ²Cornell University, Department of Chemistry & Chemical Biology, P.O. Box 305, Ithaca, New York 14851-0305, USA. Correspondence should be addressed to G.S.W. (waldo@lanl.gov).

Received 20 June; accepted 25 August; published online 20 December 2005; doi:10.1038/nbt1172

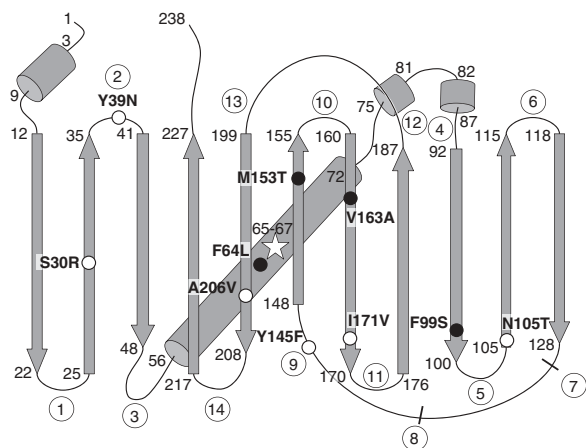


Figure 1 Schematic representation of the GFP scaffolding. Folding-enhancing mutations of folding reporter GFP include the enhanced GFP mutations¹⁸ F64L and S65T, and the cycle-3 GFP mutations¹⁷ F99S, M153T, and V163A (black circles). Superfolder GFP mutations include S30R, Y39N, N105T, Y145F, I171V, and A206V (white circles). Positions of new start sites for fourteen circularly permuted variants (circled numbers; see **Table 2**, and **Supplementary Table 1** online).

near 485 nm (**Supplementary Fig. 1a,b** online). The relative quantum efficiencies of folding reporter GFP and superfolder GFP were 0.72 and 0.65, using the experimentally determined molar extinction coefficients $8.24 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1}$ and $8.33 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1}$. Superfolder GFP photobleached slightly faster than folding reporter GFP, losing 20–25% of the initial fluorescence after 40 min continuous illumination in a plate reader (**Supplementary Fig. 1c** online).

We tested whether the superfolder mutations would improve the folding of color variants of folding reporter GFP. Blue, cyan and yellow fluorescent proteins (BFP, CFP, YFP) were produced from folding reporter GFP and superfolder GFP by incorporating the additional mutations Y66H¹⁹ (BFP), Y66W¹⁹ (CFP) and T203Y²⁰ (YFP), respectively. *E. coli* colonies expressing the color variants of superfolder GFP as C-terminal fusions with poorly folded ferritin were brightly fluorescent, whereas colonies expressing the corresponding folding reporter GFP color variant fusions were faintly fluorescent (see **Supplementary Fig. 2** online).

To test the generality of our approach, we engineered a folding-enhanced variant of the tetrameric *Discosoma* red fluorescent protein

DsRed.T4 (ref. 21) fused to poorly folded ferritin. The new folding enhanced DsRed contains twelve additional mutations A2E, R17H, T43N, A44V, V105A, Q114E, F118L, E160D, E176D, F177I, S203N and A217T. The engineered DsRed migrated as a tetramer by native polyacrylamide gel electrophoresis (PAGE) and analytical gel filtration (data not shown) and was not characterized further. Folding reporter GFP, superfolder GFP and the GFP color variants migrated as monomers by analytical gel filtration loaded at 10 mg/ml (data not shown).

Refolding kinetics and stability of GFP variants

We analyzed the effects of the superfolder GFP mutations on GFP folding kinetics and stability. Fluorescent folding reporter GFP, superfolder GFP and six folding reporter GFP constructs each bearing one of the superfolder GFP mutations were denatured in urea²², and the recovery of fluorescence was monitored upon dilution in fresh buffer. Both superfolder GFP and folding reporter GFP recovered more than 95% of the starting fluorescence within 4 min, and superfolder GFP displayed a ~ 3.5 -fold faster initial rate for fluorescence recovery relative to folding reporter GFP (**Fig. 2a**). Both long- and short-term progress curves were simultaneously fitted with three exponentials using nonlinear least squares (see **Table 1**, and **Supplementary Figs. 3–6** online). The greatest effect on the initial rapid recovery of nonfusion GFP fluorescence was shown by S30R and Y39N, consistent with the enhanced fluorescence of these mutants fused to poorly folded ferritin (**Table 1**). N105T, I171V and A206V did not significantly improve the refolding rate of folding reporter GFP. Y145F decreased the initial folding rate of folding reporter GFP, even though it nearly doubled the fluorescence of folding reporter GFP fused to ferritin at 37 °C (**Table 1**).

We tested the tolerance of the various GFP constructs to urea during refolding. Aliquots of denatured GFP variants were refolded in buffers containing varying concentrations of urea at 15 °C, and the fraction of recovered fluorescence was determined at various times (**Supplementary Fig. 7** online). Fluorescence recovery was rapid at low concentrations of urea, but took substantially longer above the transition concentration as previously noted during the refolding of fluorescent proteins²³. The fraction of fluorescence recovered at 52 h was plotted as a function of urea concentration in the refolding buffer (**Fig. 2b**), and the mid-point concentration at which 50% of the fluorescence was recovered was determined by sigmoidal fits to the equilibrium renaturation plots (**Fig. 2b**) using nonlinear least squares (**Table 1**). The stabilities of the GFP variants were estimated from the dependence of the free energy on urea concentration as previously

Table 1 GFP fusion fluorescence, nonfusion refolding kinetic parameters, and equilibrium refolding

GFP variant	^a F _{rel} 27 °C	^a F _{rel} 37 °C	^b k ₁ (10 ⁻¹ s ⁻¹)	^b k ₂ (10 ⁻² s ⁻¹)	^b k ₃ (10 ⁻² s ⁻¹)	^c a ₁	^c a ₂	^c a ₃	^d C _{1/2} (M)	^e ΔG(H ₂ O) (kcal mol ⁻¹)	^f m (kcal mol ⁻¹ M ⁻¹)
Folding reporter	81.0	81.0	3.8 ± 0.1	5.1 ± 0.3	1.43 ± 0.03	0.48 ± 0.03	0.22 ± 0.03	0.29 ± 0.05	3.39 ± 0.03	7.30 ± 0.12	2.15 ± 0.06
Superfolder	20.5	51.0	13.5 ± 0.8	21.5 ± 0.9	1.85 ± 0.06	0.33 ± 0.01	0.35 ± 0.06	0.32 ± 0.02	4.91 ± 0.02	9.57 ± 0.21	1.97 ± 0.06
^h S30R	5.9	4.1	5.9 ± 0.4	10.0 ± 0.3	1.60 ± 0.20	0.43 ± 0.02	0.24 ± 0.06	0.33 ± 0.05	3.77 ± 0.03	8.55 ± 0.11	2.27 ± 0.03
^h Y39N	11.2	3.7	5.8 ± 0.1	8.6 ± 1.0	1.60 ± 0.20	0.49 ± 0.04	0.20 ± 0.04	0.31 ± 0.03	3.47 ± 0.04	7.52 ± 0.13	2.17 ± .03
^h N105T	2.9	3.6	3.7 ± 0.2	7.9 ± 0.7	1.48 ± .03	0.43 ± 0.02	0.24 ± 0.05	0.33 ± 0.06	3.70 ± 0.02	7.33 ± 0.17	1.98 ± 0.04
^h Y145F	1.6	1.9	3.4 ± 0.1	4.1 ± 0.2	1.07 ± 0.02	0.38 ± 0.06	0.23 ± 0.05	0.39 ± 0.02	3.55 ± 0.04	7.27 ± 0.28	2.07 ± 0.13
^h I171V	2.1	1.9	4.1 ± 0.1	6.6 ± 0.8	1.50 ± 0.10	0.48 ± 0.04	0.18 ± 0.03	0.35 ± 0.04	3.37 ± 0.03	6.68 ± 0.27	1.98 ± 0.11
^h A206V	3.9	3.1	3.9 ± 0.2	5.6 ± 0.5	1.28 ± 0.02	0.47 ± 0.03	0.20 ± 0.03	0.34 ± 0.04	3.68 ± 0.04	7.24 ± 0.10	1.97 ± 0.04

^aLiquid culture whole-cell fluorescence for indicated ferritin–GFP variant fusion, divided by fluorescence of ferritin–folding reporter GFP fusion at indicated temperature (F_{rel}). Liquid culture fluorescence normalized by optical density at 600 nm. Average of triplicate readings, relative uncertainty $\sim 5\%$. ^bRate constants for three-exponential fits of refolding progress curves of nonfusion GFP variant. Errors reported are one standard deviation, four replicates (k₁–k₃). ^cRelative magnitude of each exponential process (a₁–a₃). ^dTransition concentration of urea at which 50% of the initial fluorescence is recovered during refolding of urea-unfolded nonfusion GFP protein (C_{1/2}), determined by fitting the renaturation profiles of **Figure 2b**. ^eFree energy of denaturation ΔG° = ΔG(H₂O) – m[urea]. ^fMeasure of dependence of ΔG as a function of denaturant concentration, that is, slope of plots in **Figure 2c**. ^gFluorescence of ferritin–folding reporter GFP fusion is 3,400 at 27 °C, 1,400 at 37 °C (arbitrary units). ^hFolding reporter GFP bearing the indicated superfolder GFP mutation.

Table 2 Comparison of solubility and whole-cell fluorescence of fourteen circularly permuted variants of folding reporter GFP and superfolder GFP

cNumber	dStarting aa	aFraction soluble		bWhole-cell fluorescence	
		Folding reporter GFP	Superfolder GFP	Folding reporter GFP	Superfolder GFP
1	23	0.00	0.35	140	5,600
2	39	0.00	0.40	10	8,950
3	51	0.05	0.70	300	22,300
4	91	0.05	0.25	170	4,750
5	102	0.00	0.20	180	5,450
6	117	0.15	0.35	1,800	8,650
7	129	0.00	0.05	90	200
8	140	0.05	0.60	80	11,000
9	145	0.00	0.75	80	8,750
10	157	0.15	0.75	910	20,000
11	173	0.10	0.80	1,700	25,000
e12	189	0.00	0.95	90	1,450
13	195	0.05	0.75	40	11,500
14	214	0.05	0.25	90	3,450
fControl	1	0.85	0.90	11,000	25,000

aFraction soluble expressed in liquid cultures of *E. coli* BL21(DE3) at 37 °C as determined by SDS-PAGE densitometry. Relative uncertainty is ~5%, average of three replicates. bWhole-cell fluorescence of liquid cultures of *E. coli* BL21(DE3) at 37 °C, normalized by dividing by optical density of culture at 600 nm. Relative uncertainty is ~5%, average of three replicates. cCircularly permuted variant number corresponding to Fig. 1. dAmino acid (aa) position of start of variant. Stop of variant previous amino acid in GFP. eLow fluorescence is due to poor expression level of circularly permuted variant no. 12, less than 5% the expression level of the nonpermutated construct. Expression of other circularly permuted variants is similar to the nonpermutated construct as determined by SDS-PAGE densitometry. fNonpermutated constructs (control). Relative expression levels of folding reporter GFP and superfolder GFP are equal as determined by SDS-PAGE densitometry.

described²³ (Table 1 and Fig. 2c). Superfolder GFP is substantially more stable than folding reporter GFP (Table 1 and Fig. 2b). Notably, although S30R clearly improved both folding kinetics and urea tolerance during refolding, other mutations exerted differential effects. For example, N105T increased urea tolerance during refolding more than did Y39N, whereas Y39N improved refolding kinetics more than did N105T (Table 1).

Tolerance of circular permutation

Circular permutation of proteins can drastically decrease protein folding yields^{14,15}. To further compare the folding robustness of folding

reporter GFP and superfolder GFP, we engineered fourteen circularly permuted variants of each, ligating the GFP N and C termini using a short GGGS linker, targeting loops between β -strands as new translation initiation sites (Fig. 1, and Supplementary Table 1 online). Only the folding reporter GFP variants beginning at amino acid positions 117, 157 and 172 were appreciably fluorescent, in agreement with published observations^{14,15} (Table 2). In contrast, all superfolder GFP variants, with the exception of the poorly expressed variant beginning at amino acid 129, were permissive of circular permutation, and the fluorescence and solubility of each superfolder GFP variant was greater than the corresponding folding reporter GFP variant (Table 2).

Tolerance of random mutations

Random mutagenesis of proteins often produces aggregation-prone variants. We investigated whether the improved folding of superfolder GFP would make it more tolerant of random mutagenesis compared to folding reporter GFP. We processed folding reporter GFP and superfolder GFP in parallel, and used gene shuffling to create variant libraries with a ~0.7% mutation rate. We used flow cytometry to measure the fluorescence histogram of *E. coli* cells induced in liquid culture at 37 °C for each GFP library and starting GFP variant (Fig. 3). The folding reporter GFP mutant library contained a greater fraction of weakly fluorescent cells (Fig. 3a) compared with the superfolder GFP mutant library (Fig. 3b). The liquid cultures of the libraries were fractionated and analyzed by SDS-PAGE. The pooled variants of the folding reporter GFP library (Fig. 3a, inset) were less soluble than the pool of superfolder GFP variants (Fig. 3b, inset). Apparently the superfolder mutations make the folding of GFP tolerant of mutations that would otherwise reduce the folding yield of GFP.

Superfolder GFP as a fusion tag and protein expression reporter

We investigated whether superfolder GFP would fluoresce brightly when fused to a variety of proteins with different expression levels, folding yields and solubilities. We expressed superfolder GFP and folding reporter GFP as C-terminal fusions with eighteen *P. aerophilum* proteins^{9,24} (Supplementary Table 2 online), as pET vector constructs in *E. coli* BL21(DE3) at 37 °C. For each test protein, the expression level of the folding reporter GFP fusion and corresponding superfolder GFP fusion were similar as determined by SDS-PAGE (Fig. 4a). *E. coli* colonies expressing folding reporter GFP fusions were bright only for protein no. 2, sulfite reductase, a well-expressed (Fig. 4a), soluble protein expressed alone (Supplementary Table 2). Folding reporter GFP fusion fluorescence was weak for well-expressed

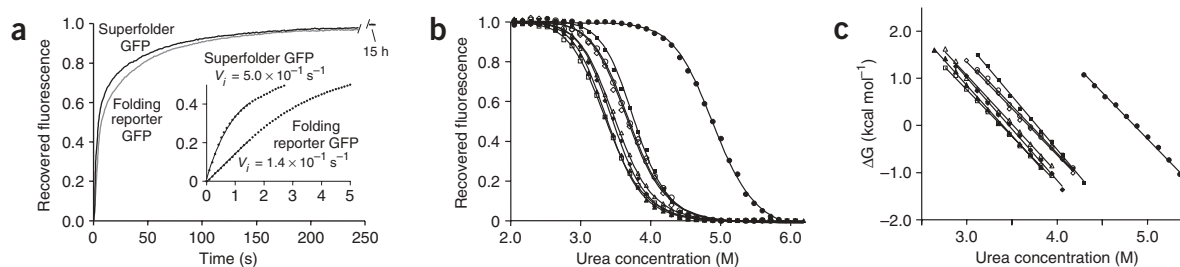
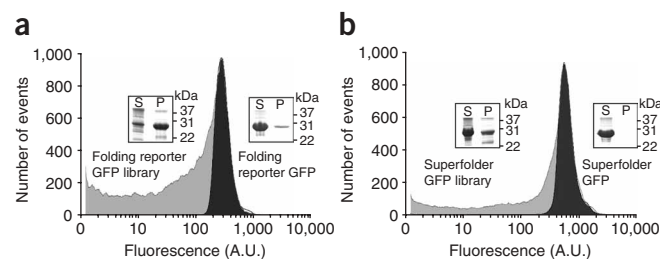


Figure 2 GFP refolding kinetics and equilibrium renaturation plots. (a) Long-term (250 s) progress curves for recovery of fluorescence during refolding of urea-denatured superfolder GFP (black) and folding reporter GFP (gray) upon 20-fold dilution of denatured samples in fresh buffer containing 1 mM DTT at 25 °C. Short-term progress curves for superfolder GFP and folding reporter GFP refolding (inset). Initial rates V_i were obtained from slope at $t = 0$ s of 2nd-order polynomials fitted to the first 1.5 s of short-term progress curves. Fluorescence normalized by dividing by final fluorescence value at 15 h. (b) Equilibrium refolding plots (fraction of recovered fluorescence) for urea-denatured GFP variants as a function of the final concentration of urea in refolding buffer. GFP variants are (left to right): I171V (□), FR (▲), Y145F (◆), Y39N (△), A206V (◇), N105T (○), S30R (■), and SF (●). Recovered fluorescence normalized by dividing by fluorescence of corresponding nondenatured samples diluted in parallel. (c) Dependence of the standard free energy of denaturation on urea concentration assuming a two-state folding model for GFPs (Table 1).

Figure 3 Tolerance of folding reporter GFP and superfolder GFP to random mutation. Fluorescence histograms (number of cells having indicated fluorescence) obtained by flow cytometry of liquid cultures of *E. coli* BL21(DE3) expressing indicated parental GFP variant (black), or libraries of random point mutants derived from parental variant (gray). **(a)** Fluorescence histograms for *E. coli* expressing parental folding reporter GFP (black), or a pooled random mutation library derived from folding reporter GFP (gray). SDS-PAGE of soluble (S) and pellet (P) samples of pooled random mutation library of indicated starting parental variant (insets). **(b)** Fluorescence histograms for superfolder GFP (black), or a pooled random mutation library derived from superfolder GFP (gray). SDS-PAGE of soluble (S) and pellet (P) samples of pooled random mutation library of indicated starting parental variant (insets).

partially soluble or insoluble test proteins (Fig. 4a), reflecting the nonfusion protein-folding yield and solubility (Supplementary Table 2), as previously observed⁹. Colonies expressing superfolder GFP fusions were considerably brighter than the corresponding folding reporter GFP fusions, especially for proteins that were partially soluble or insoluble expressed alone (Fig. 4a and Supplementary Table 2).

Superfolder GFP whole-cell colony fluorescence appeared to depend on total fusion protein expression (Fig. 4a). In the absence of GFP misfolding, the total cell fluorescence should be directly proportional to the number of molecules of expressed fusion protein. To more accurately determine the correlation between fusion fluorescence and total moles of protein expressed, we first estimated the amount of each expressed fusion protein by SDS-PAGE densitometry, and then normalized it by dividing by the theoretical molecular weight of each fusion (Supplementary Table 2) to obtain the 'fusion expression.' Superfolder GFP whole-cell liquid culture fluorescence was



linearly proportional to fusion expression (Fig. 4b). In contrast, folding reporter GFP whole-cell liquid culture fluorescence was poorly correlated with fusion expression (Fig. 4b), even when the outlier sulfite reductase was omitted from the fit (Fig. 4b, inset). Data points corresponding to partially soluble or insoluble proteins clustered near the origin of the plots for folding reporter GFP fusions. The productive folding of the fused folding reporter GFP is reduced by misfolding of the upstream fused-protein domain^{9–12}.

To ascertain whether the increased superfolder GFP fusion fluorescence was associated with a corresponding increase in the amount of soluble superfolder GFP fusion protein, we expressed the eighteen *P. aerophilum* test proteins with N-terminal 6HIS tags, alone or fused to C-terminal folding reporter GFP or superfolder GFP. Aliquots of the soluble and pellet fractions were analyzed using a fluorescence plate reader, and also resolved on SDS-PAGE gels (Fig. 4c). Well-expressed proteins that were insoluble when expressed without the

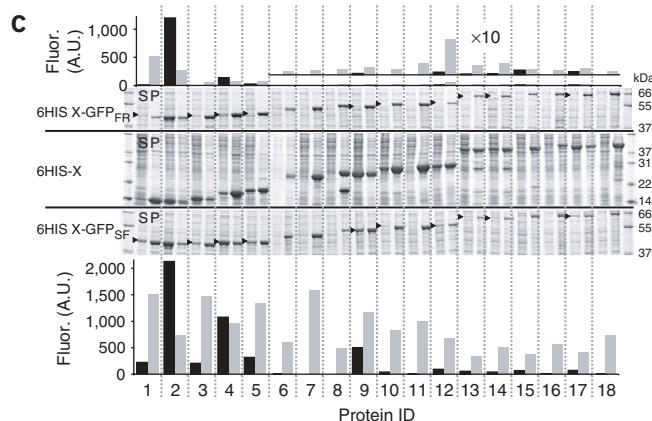
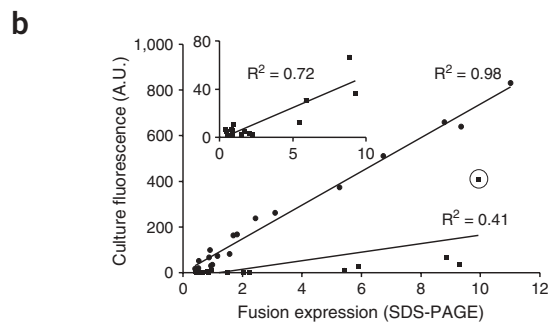
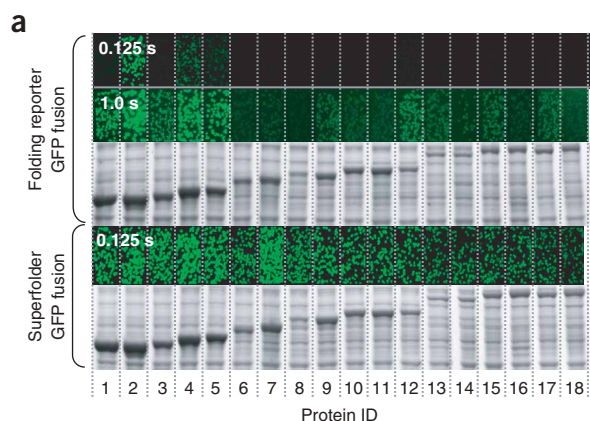


Figure 4 Fluorescence, expression level, and solubility of GFP fusions to eighteen *P. aerophilum* control proteins. **(a)** *In vivo* fluorescence of *E. coli* colonies resting on nitrocellulose membranes, expressing indicated *P. aerophilum* test protein at 37 °C with C-terminal folding reporter GFP (top) or superfolder GFP (bottom), at indicated exposure times. SDS-PAGE of aliquots of cells expressing the indicated GFP fusion constructs in liquid culture. **(b)** Whole-cell liquid culture fluorescence of same constructs as in **a**, as a function of the amount of expressed fusion protein (fusion expression) proportional to the number of moles of fusion protein (see Methods). N-terminal fusions with superfolder GFP (●) or folding reporter GFP (■). Circled datum corresponds to the most fluorescent fusion construct (sulfite reductase). Fit to folding reporter data, omitting the data point for sulfite reductase (inset). **(c)** SDS-PAGE of soluble (S) and pellet (P) fractions of liquid cultures of cells expressing the indicated *P. aerophilum* control protein as folding reporter GFP fusion (top), nonfusion (middle) or superfolder GFP fusion (bottom). Bar graphs of fluorescence of soluble (S) and pellet (P) fractions of folding reporter GFP fusions (upper) or superfolder GFP fusions (lower). Tenfold scaled bar graph for folding reporter GFP data (inset, upper). Arrows indicate expected position of soluble fusion protein.

fused GFP (for example proteins 6, 7, 8, 11, 16 and 18) were also insoluble when fused to either folding reporter GFP or superfolder GFP (Fig. 4c). In the case of superfolder GFP, the insoluble fractions were brightly fluorescent (Fig. 4c). Apparently, the superfolder GFP mutations somehow uncoupled the folding trajectory of the GFP from the misfolded, insoluble fused upstream protein domain without increasing the solubility of the fusion.

To more accurately measure the effect of the fused GFP domains on the solubility of each construct, we analyzed SDS-PAGE gels (Fig. 4c) by densitometry (Supplementary Table 2). Bright superfolder GFP fluorescence was associated exclusively with the pellet fraction for insoluble fusions, indicating that GFP fusion solubility is not a prerequisite for the production of superfolder GFP fluorescence *per se* (Fig. 4c). Most of the partially soluble proteins were insoluble when expressed as fusions with C-terminal folding reporter GFP. Only sulfite reductase (protein no. 2) and translation initiation

factor (protein no. 4) were detectably soluble as folding reporter GFP fusions (Fig. 4c and Supplementary Table 2). Proteins that were partially soluble when expressed without the fused GFP (proteins nos. 1, 3, 4, 5, 9, 12 and 17, see Supplementary Table 2) were totally insoluble when fused to folding reporter GFP (Fig. 4c), but were detectably soluble as superfolder GFP fusions (Fig. 4c and Supplementary Table 2).

X-ray structures of superfolder GFP and folding reporter GFP

We determined the structures of folding reporter GFP at a resolution of 2.3 Å and of superfolder GFP at 1.4 Å (Supplementary Table 3 online). The chromophore formed from residues T65-Y66-G67 is buried in the middle of the central helix. The C α backbone trace of folding reporter GFP and superfolder GFP can be superimposed with a root mean square deviation (r.m.s.d.) of 0.7 Å from residues 2 to 232 (Fig. 5a). Subtle differences between superfolder GFP and folding

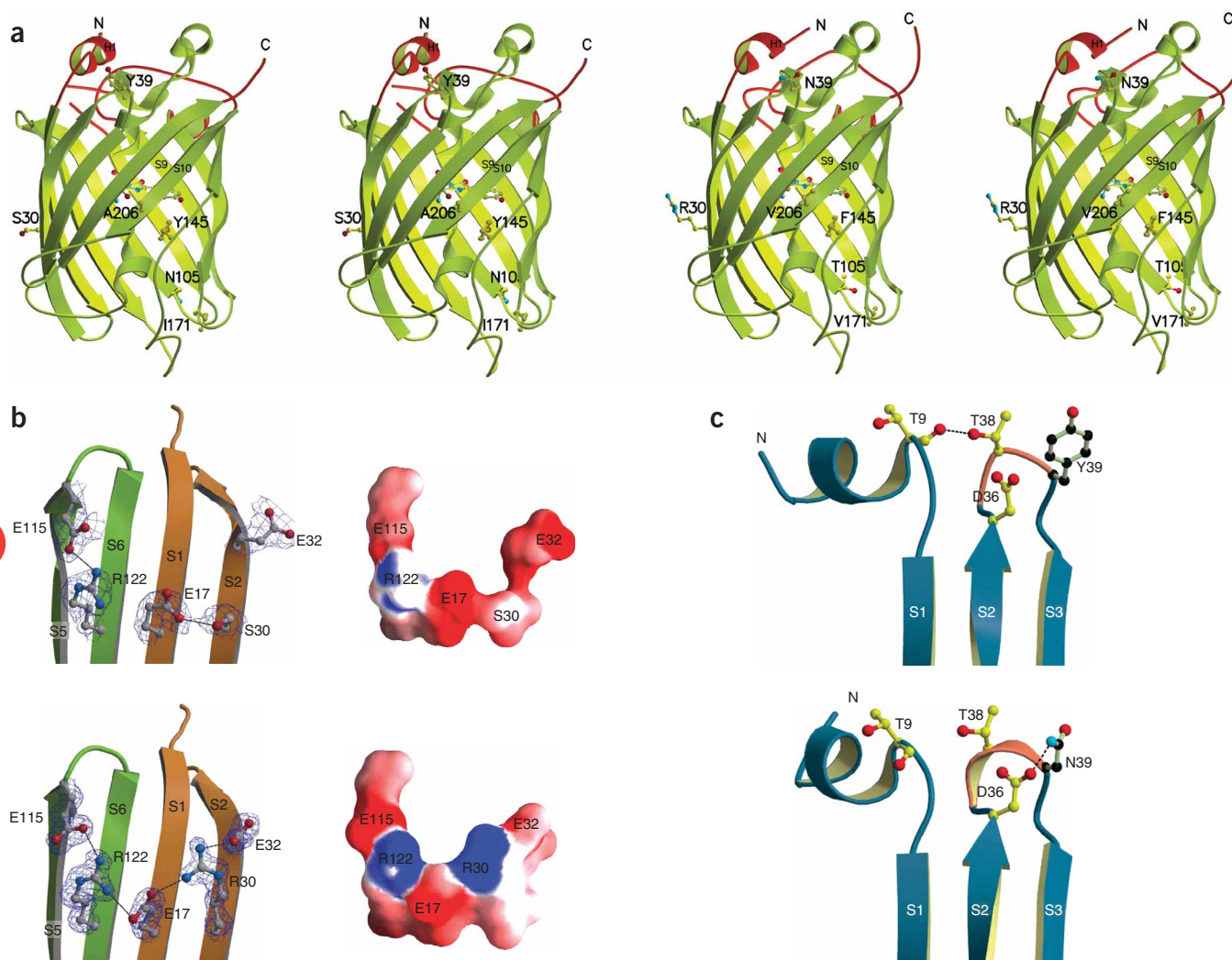


Figure 5 Three-dimensional structure of folding reporter GFP and superfolder GFP. (a) Stereo views of folding reporter GFP (left) and superfolder GFP (right) showing disposition of superfolder GFP mutations and corresponding amino acids in the folding reporter GFP (labeled ball-and-stick models). (b) Ribbon diagram (left) and charge surface (right) of region near amino acid position 30 for folding reporter GFP (top) and superfolder GFP (bottom). The superfolder GFP mutation S30R (bottom) mediates the formation of an extended electrostatic network involving amino-acid residues E115 (S5), R122 (S6), E17 (S1), R30 (+) and E32 (S2). (c) Region of GFP scaffolding near amino acid 39 for folding reporter GFP (Y39) (top) and superfolder GFP (N39) (bottom). In superfolder GFP, a hydrogen bond between a side-chain oxygen of D36 and the side-chain NH group of N39 tightens the loop, likely enhancing the formation of the 3_{10} α -helix structure at the turn between β -strands S2 and S3.

reporter GFP, likely resulting from the different crystal-packing forms, were observed at the top end of the β -barrel (see **Supplementary Fig. 8** online). This region is $>8 \text{ \AA}$ from the closest mutated residue (Y39N).

GFP dimerizes only weakly in dilute solutions²⁵, but folding reporter GFP crystallizes as an antiparallel dimer, similar to other reported GFP crystal structures. Superfolder GFP crystallizes as a monomer. A206K substantially reduces the tendency of GFP to dimerize even in concentrated solutions²⁶, ostensibly by charge repulsion and introduction of the bulky side chain. Based on the structure of the dimeric folding reporter GFP, we generated a model of the dimer of superfolder GFP (data not shown). A206V comes into close contact (1.97 \AA) with the phenyl group of F223 from the symmetry-related monomer, potentially hindering dimerization. Alternatively, the monomeric superfolder structure may result from unique crystal packing (see **Supplementary Fig. 8**).

Structural clues to superfolder GFP enhanced-folding robustness

Among the superfolder GFP mutations, S30R contributes most to GFP folding fluorescence as a ferritin fusion at 37 °C, and has the fastest refolding kinetics and greatest tolerance of urea during refolding (**Table 1**). A comparison of the folding reporter GFP and superfolder GFP structures offers a structural basis for these observations. The amino acid residue located at position 30 is in the middle of β -strand S2, with the backbone amide protons and carbonyl oxygen hydrogen-bonded to residue 47 from the anti-parallel β -strand S3 (**Fig. 5a,b**). In the folding reporter GFP structure, the side chain of S30 is hydrogen bonded to one of the side-chain oxygens of the conserved residue E17 (**Fig. 5b**). In the superfolder GFP structure, S30R causes local conformational changes, which affect the side-chain positions of E17 and E32 (**Fig. 5b**). S30R mediates the formation of an electrostatic network involving E17, E32 and the polar side chains of E115 from β -strand S5 and R122 from β -strand S6. In addition, the positively charged planar guanidinium group of R122 is also hydrogen bonded to the side chain of E32 via the R122 N- ζ . The chain of alternating acidic and basic charged residues is composed of E32–R30–E17–R122–E115 from the four adjacent β -strands S1, S2, S5 and S6 (**Fig. 5b**). This five-membered intramolecular ionic network might be expected to increase the global stability of the β -can structure, because the organization of ion pairs in networks offers an energetic advantage over single ion-pairs due to the combined reduced entropy cost of both binding a third residue^{27,28} and desolvation²⁹.

The residue at position 39 is located in the region that connects strand S2 to strand S3 (**Fig. 5c**). Two consecutive type I β -turns are observed in folding reporter GFP, at residues 36 to 39, and 37 to 40 (**Fig. 5c**). In superfolder GFP, these are converted into a single 3_{10} helix with the mutation Y39N at the C terminus. An additional hydrogen bond is present in this turn in the superfolder GFP, attributable to the slight change in the backbone dihedral angles Φ (from 89° to 84°) and Ψ (from -135° to -123°) associated with the Y39N mutation. This H-bond links the C = O group of N39 to the N-H group of K41, a small movement that may help to stabilize the 3_{10} α -helix. In superfolder GFP, the side chain of D36 undergoes a substantial conformational change relative to its orientation in the folding reporter GFP, forming a hydrogen bond between a side-chain oxygen of D36 and a nitrogen atom of N39 (**Fig. 5c**). This hydrogen bond tightens the loop (the C α distance between these two residues decreases from 5.78 \AA in the folding reporter GFP to 5.58 \AA in superfolder GFP), possibly enhancing the formation of the 3_{10} α -helix structure at the turn between β -strands S2 and S3. In contrast, in the folding reporter GFP structure, stabilization

of Y39 only involves intramolecular Van der Waals contacts with the side chains of D36 and T38 or intermolecular Van der Waals contacts with the side chains of D210, P211 and N212 from the adjacent antiparallel monomer.

All superfolder mutations are solvent exposed with the exception of Y145F. The presence of the N105T side chain has no effect on the orientation of the nearest residue side chains. In folding reporter GFP, the side-chain CO group of N105 is 4.09 \AA away from the C γ atom of the solvent exposed side chain of K107. In both folding reporter GFP and superfolder GFP, side-chain orientations of six residues (P58–V61–I167–H148–H169–L207) and seven ordered water molecules surrounding residue 145 are conserved. The C α distance to the closest mutated residue, A206V, is 6.51 \AA in superfolder GFP, and 6.08 \AA in folding reporter GFP. In superfolder GFP, the I171V mutation causes the side chain of L141 to move towards V171, stabilizing both side chains via nonpolar interactions.

DISCUSSION

In this work, we demonstrate a simple folding interference method for improving the folding robustness of fluorescent proteins³⁰. We used poorly folding bait proteins (or ‘folding interference’ domains) as N-terminal fusions with both monomeric GFP and tetrameric DsRed.T4 (ref. 21) to interfere with proper folding of the fluorescent proteins; then fluorescent protein variants that can still fold were identified. Folding of fluorescent protein domains can also be reduced by circular permutation of the GFP scaffolding¹⁴ or by insertion of poorly folded peptides into the GFP scaffolding¹³, but these approaches are more complicated. Although small physical proteins with improved stability have been selected using phage display³¹, the extreme conditions required to unfold GFP, that is, by heating to 95 °C in 9 M urea, would be expected to also unfold phage coat proteins. Our folding interference method is simple to apply and should be considered whenever fluorescent proteins are to be used as general fusion tags.

Superfolder GFP refolds with an initial rate of $5.0 \times 10^{-1} \text{ s}^{-1}$ (**Fig. 2a** and **Table 1**), making it one of the fastest folding GFPs yet reported. Refolding progress curves of the urea-denatured S65T variant of cycle-3 GFP can be modeled with two exponential processes²² with rate constants of $k_1 = 2.54 \times 10^{-2} \text{ s}^{-1}$ and $k_2 = 2.44 \times 10^{-3} \text{ s}^{-1}$. In contrast, the refolding progress curves for urea-denatured folding reporter GFP or superfolder GFP must be modeled with three relaxations, suggesting more complex kinetics. All six superfolder GFP mutations decrease folding interference, but by different mechanisms. S30R and Y39N provide the greatest improvement in folding robustness, refolding kinetics and resistance to urea denaturation (**Table 1**).

A detailed crystallographic analysis revealed an unusual extended charge network mediated by R30. Interestingly, arginine is the only amino acid residue that is capable of participating in the formation of double salt bridges with aspartate or glutamate amino acid side chains. Y145F and I171V reduce GFP folding interference, but do not substantially improve folding kinetics (**Table 1**). It is likely that these mutations eliminate aggregation-prone, off-pathway intermediates from the GFP folding trajectory, rather than improving stability of the native state. N105T and A206V are solvent-exposed mutations located within strands and bordered on both sides by another β -strand. Both mutations result in substitutions by amino acids with increased β -strand propensities. Among the 20 naturally occurring amino acids, threonine is the residue with the highest β -sheet propensity, whereas valine, alanine and asparagine are in fifth, fifteenth and seventeenth positions, respectively³². Such mutations

can sometimes increase the expression yield of soluble protein and improve thermodynamic stability³³, perhaps explaining the selection of N105T and A206V.

Superfolder GFP mutations make the folding of GFP resistant to random mutation, a process that normally reduces the GFP folding yield (Fig. 3). Similarly, 'global suppressor mutations' render tail spike proteins tolerant of mutations that interfere with folding and assembly^{34–37}. Second site mutations also compensate for drug resistance mutations that normally interfere with β -lactamase folding^{38,39}. Stability of the native state does not necessarily confer immunity to folding interference. For example, natively folded DsRed is at least as stable as superfolder GFP²³, but DsRed is prone to folding interference (see Supplementary Fig. 3 online). We have not determined whether superfolder GFP exhibits an increased rate of chromophore cyclization relative to folding reporter GFP. However, correct folding of the GFP β -barrel appears to be a prerequisite for formation of the fluorescent chromophore¹.

Superfolder GFP has numerous applications. Protein fusions with existing GFP variants are often weakly fluorescent and poorly soluble^{9–13}, and partially soluble proteins can become insoluble when expressed as GFP fusions. Superfolder GFP folds well regardless of the folding status or solubility of the fusion partner, and superfolder GFP fusions are more soluble than conventional GFP fusions, expanding the range of proteins amenable to cell trafficking experiments⁴⁰ using fluorescent protein tagging. Currently, alternative approaches for detecting and quantifying soluble or insoluble proteins *in vivo* often involve the time-consuming and complex steps of cell fixing, permeabilization and probing with specific, labeled antibodies⁴⁰. Proteins can be labeled with fluorogenic biarsenical FAsH or ReAsH⁴¹, but staining of endogenous cysteine-rich cellular proteins results in a 16-fold lower signal-to-noise ratio compared with GFP fusions for mammalian cell applications⁴². Our new circularly permuted variants of superfolder GFP (Table 2) should prove useful as alternative fluorescence-resonance energy transfer (FRET) partners¹⁴, since the FRET efficiency is strongly dependent on the angle between donor and acceptor dipoles⁴³. Superfolder fluorescent proteins could be further engineered to produce novel GFP-based fusion protein adducts, biosensors and split proteins²⁴, as well as folding reporters with alternative topologies, perhaps detecting misfolding events that might be missed by the C-terminal GFP reporter⁹.

METHODS

Construction of plasmids. The fluorescent protein fusion plasmid was constructed by inserting the *BglII/XhoI* fragment of pET21(+)(Novagen) into the corresponding site of pET28(+)(Novagen), and the *BamHI*(GGATCC)/*EcoRI*(GAATTC) site was replaced with the DNA fragment GGATCCGCTGGCTCCGCTGCTGTTCTGGCGAATTC coding for the amino acid linker GSAGSAAGS GEF. We avoided large bulky hydrophobic residues in designing the linker. The *BglII/NdeI* fragment was subcloned from pET28(+)(Novagen) into the fusion plasmid. The *NdeI/BamHI* cloning site was replaced by a frameshift stuffer with three translational stops CATATGTGTTAACTGAGTAGGATCC, and the resulting vector was digested with *NdeI* and *BamHI* to receive inserts. GFP variants (this work) or DsRed.T4 (ref. 21) were amplified by PCR, and cloned into the *EcoRI/KpnI* site. Test proteins from *P. aerophilum*⁹ amplified by PCR were cloned into the *NdeI/BamHI* site as 6HIS-(thrombin cleavage site)-(test protein)-(GSAGSAAGS GEF linker)-fluorescent protein. An N-6HIS pET vector was engineered by replacing the *BamHI*-linker-*EcoRI*-GFP-*KpnI/XhoI* module in the GFP fusion vector with a module *BamHI*-TAA (stop codon)-*XhoI*. A C-terminal pET expression vector was engineered by replacing the *BglII/NdeI* fragment of the GFP fusion vector with the *BglII/NdeI* fragment of pET 21(+)(Novagen), and replacing the *BamHI*-linker-*EcoRI*-GFP-*KpnI/XhoI* module with the module *BamHI*-CACCATCACCACCATCATTA-*XhoI*.

Engineering superfolder GFP, SF DsRed, GFP color variants and GFP circularly permuted variants. Folding reporter GFP or DsRed.T4 were cloned as C-terminal fusions with ferritin in the modified pET fusion expression plasmid (above) and amplified by PCR using primers including ~50 bp of flanking vector sequence on each end of the amplicon. PCR amplicons were subjected to DNA fragmentation and shuffling using published protocols⁴⁴. The cDNA library plasmid pool was transformed into an *E. coli* BL21 (DE3) expression strain (Novagen). The expression library, containing $\sim 1 \times 10^5$ variants, was plated on nitrocellulose membranes using two successive 400-fold dilutions of a 1.0 OD₆₀₀ nm frozen 20% glycerol/Luria-Bertani (LB) stock, yielding $\sim 3 \times 10^3$ colonies per plate. After overnight growth at 37 °C, the membrane was transferred to an LB/agar plate containing 50 μ g kanamycin/ml media, plus 1 mM isopropyl- β -D-thiogalactopyranoside (IPTG) for 3 h at 37 °C. Clones exhibiting the brightest fluorescence (488 nm excitation/530 nm emission for GFP, 580 nm excitation/620 nm emission for DsRed) were picked and frozen as -80 °C 20% glycerol freezer stocks, pooled and subjected to another round of evolution. Approximately 20,000 clones were screened per round of evolution. After four rounds of mutagenesis, selection and recombination, the 20 brightest clones were backcrossed by DNA shuffling against the starting parental sequence to remove nonessential mutations⁴⁴. The ten brightest clones were sequenced for GFP or DsRed optima using fluorescent dideoxy terminator DNA sequencing, and each contained the superfolder GFP or SF DsRed consensus mutations (above). Blue (Y66H), cyan (Y66W), and yellow (T203Y) green fluorescent protein variants of folding reporter GFP and superfolder GFP were engineered by primer-directed mutagenesis (see Supplementary Methods online for primer sequences). GFP circularly permuted variants were engineered by a first PCR linking the native N and C termini using primers coding for the amino acid linker GGGs flanked by homology either to the first or last 18 bp of the scaffolding, designed to also remove the initiator methionine and terminator codons. The gel-purified amplicon was used as the template for a second PCR with primers specifying new starting and stopping amino acid positions (see Supplementary Table 1 online). Single-point superfolder GFP mutants were constructed using the folding reporter GFP template and indicated primers (see Supplementary Table 1 online). Sequences of constructs were confirmed by fluorescent dideoxy terminator DNA sequencing.

Quantum efficiency, molar extinction coefficients, fluorescence excitation and emission spectra, and photobleaching. For molar extinction coefficient determinations and quantum efficiency calculations, purified superfolder GFP and folding reporter GFP were diluted in 100 mM TRIS, pH 7.5, 150 mM NaCl, 10% glycerol (TNG) buffer such that the optical densities were 0.50 at 488 nm. Protein concentrations were determined using the bicinchoninic acid (BCA) reagent assay kit (Pierce). Using the predicted molecular weights for the 6HIS-tagged fluorescent proteins, the corresponding molar concentrations for superfolder GFP (27.747 kDa/mole) and folding reporter GFP (27.742 kDa/mole) were 6.10×10^6 M and 6.00×10^6 M, respectively. The molar extinction coefficients were determined by dividing the optical density by the molar concentration. Sodium fluorescein was diluted in 0.1 M NaOH to a final optical density of 0.5 at 488 nm. Areas under fluorescence emission spectra traces were integrated over 500 to 650 nm, (488 nm excitation), and the areas of the GFP traces were divided by the area for the fluorescein trace, then multiplied by the expected quantum efficiency of fluorescein (0.92) as previously described⁴² to yield the relative fluorescence quantum yield for the fluorescent protein. For excitation spectra, fluorescence emission was monitored at 510 nm. For emission spectra, fluorescence excitation was 490 nm, 5 nm bandpass throughout (Supplementary Fig. 1a). Data were collected at 20 °C using a Perkin-Elmer LS 50 B spectrofluorimeter with 1-cm path length. Ultraviolet-visible absorption spectra were collected at 20 °C using a Cary 300 ultraviolet-visible spectrophotometer with 1-cm path length (Supplementary Fig. 1b). For photobleaching experiments, the protein and sodium fluorescein samples were diluted a further 20-fold in TNG buffer and fluorescence was monitored for 2400 s at 10 s intervals using a FL600 Microplate Fluorescence Reader (Bio-Tek) (488 nm excitation, 530 nm emission, 10 nm band pass) during continuous illumination (Supplementary Fig. 1c).

In vivo whole-cell colony plate fluorescence imaging. *E. coli* BL21 (DE3) cells expressing test proteins with an N-terminal 6HIS and C-terminal GFP or

DsRed variant were grown to saturation in LB containing 35 $\mu\text{g/ml}$ kanamycin diluted in 20% glycerol at $\text{OD}_{600\text{ nm}} = 1.0$ for -80°C freezer stocks. Cells were diluted successively with two 400-fold dilutions in LB and plated on nitrocellulose membranes. After overnight growth at 32°C , the fusion protein expression was induced by transferring the membranes to LB/agar plates with 1 mM IPTG at 37°C for 3 h. The induced colonies on the plates were illuminated using an Illumatool Lighting System (LightTools Research) and photographed with a DC290 digital camera (Kodak) with the following excitation/emission filter sets: BFP (380 nm/440 nm), CFP (440 nm/470 nm long-pass), GFP (488 nm /520 nm long-pass), YFP (512 nm/535 nm long-pass), DsRed (580 nm /610 nm long-pass).

Liquid culture whole-cell fluorescence measurements. We grew 200- μl cultures of *E. coli* BL21(DE3) bearing fusion or nonfusion constructs with specified GFP or DsRed variants, or GFP circularly permuted variants in LB medium containing kanamycin (35 $\mu\text{g/ml}$) overnight. We grew 100-fold dilutions in 1-ml cultures to $\sim 0.6 \text{ OD}_{600\text{ nm}}$ in a 96-well deepwell plate and induced them at 37°C with 1 mM IPTG for 3 h. Cultures were centrifuged at 3,000g for 15 min at 5°C , washed and resuspended with 110 μl of TNG buffer. Aliquots of the 110- μl cell suspensions were diluted 400-fold (two successive 20-fold dilutions) in a 96-well microplate (Nunc-Immuno plate, Nunc) and fluorescence was measured (488-nm excitation, 530-nm emission, 10-nm band pass for GFP or 580-nm excitation, 620-nm emission, 10-nm bandpass for DsRed) with a FL600 Microplate Fluorescence Reader. Fluorescence was normalized by dividing by the $\text{OD}_{600\text{ nm}}$ measured for tenfold dilutions of the 110- μl cell suspensions.

SDS-PAGE, solubility determinations and fluorescence of soluble and pellet fractions. Cultures were grown as for liquid culture whole-cell fluorescence measurements (see above). After induction, the culture pellets were resuspended with 110 μl of TNG buffer, and sonicated in a 96-well PCR plate using a plate sonicator. The lysate was fractionated by centrifugation to yield the soluble and the pellet fractions. We mixes 15 μl of the soluble and pellet fractions with 15 μl of 2 \times SDS denaturing buffer containing 100 mM TRIS, 200 mM dithiothreitol (DTT), 4% SDS, 0.2% bromophenol blue and 20% glycerol, and were heated for 15 min at 100°C . The denatured samples were resolved on a 4–20% gradient Criterion SDS-PAGE (Biorad). The protein samples were stained using Gel Code Blue stain reagent (Pierce) and imaged using a GS-800 Calibrated Densitometer (Biorad). Aliquots of the same soluble and pellet fractions were diluted up to 400-fold (two successive 20-fold dilutions) in a 96-well microplate (Nunc-Immuno plate) and fluorescence was measured (488-nm excitation, 530-nm emission, 10-nm band pass for GFP) with a FL600 Microplate Fluorescence Reader. Protein SDS-PAGE spot optical densities, proportional to the mass of expressed fusion protein, were measured by quantitative densitometry of SDS-PAGE of whole-cell aliquots, then divided by the theoretical molecular weight of the GFP fusion (obtained from **Supplementary Table 2**) to obtain the ‘fusion expression,’ proportional to the number of moles of fusion protein.

Flow cytometry analysis and solubility determination of GFP variants and libraries. *E. coli* BL21 (DE3) cells expressing folding reporter GFP, superfolder GFP or pooled libraries of $\sim 1 \times 10^6$ variants of each obtained by DNA shuffling⁴⁴ were grown as 3-ml LB cultures containing kanamycin (35 $\mu\text{g/ml}$). Cultures were diluted 100-fold in fresh 3-ml LB kanamycin shake cultures and grown 2 h at 37°C to $\text{OD}_{600\text{ nm}} 0.6$, then induced 3 h with 1 mM IPTG at 37°C . After two successive 400-fold dilutions, 5×10^4 cells from each pool were analyzed using a Becton-Dickinson FACS Caliber flow cytometer, and fluorescence (488-nm excitation, 520-nm emission) was scaled by scattering to compensate for differences in cell morphology and size. The maximum of each fluorescence histogram (number of events as a function of fluorescence) was scaled to $\sim 1,000$ to facilitate comparison of the histograms of starting variants with each library pool.

Kinetic and equilibrium refolding measurements. Purified GFP variants bearing a C-terminal 6HIS tag (see above) at 5 mg/ml in 20 mM HEPES-NaOH, pH 7.5, were diluted 400-fold in 9M urea, 1 mM DTT and unfolded at 95°C for 5 min. To measure refolding kinetics, 10- μl aliquots of the diluted, unfolded samples in a 96 well microplate (Nunc-Immuno plate, Nunc) were

rapidly diluted with 200 μl of fresh TNG buffer and fluorescence was measured using a FL600 Microplate Fluorescence Reader (488-nm excitation, 530-nm emission, 10-nm band pass). A 1-mm diameter pinhole aperture was fitted to the excitation filter to reduce the intensity by approximately tenfold to minimize photobleaching¹ of the GFP. In separate experiments, data were collected for 25 s at 0.1 s intervals (short-term progress curve), or 250 s at 1.0 s intervals (long-term progress curve). Data collection was initiated a few seconds before the injection of diluent to establish a baseline trace. Up to three exponential functions were fit simultaneously to both short- and long-term progress curves using the SOLVER function in EXCEL (Microsoft). Long-term fluorescence data F_{Lj} for $j = 250$ data points can be modeled using the equation $F_{Lj} = A_L + \sum_{i=1}^3 a_i e^{-b_i(t_j+C_L)}$ for $i = 1$ to 3, where the exponential factors a_i and b_i , the time offset C_L and fluorescence offset A_L are adjustable parameters. Short-term fluorescence data F_{Sj} for $j = 250$ data points were simultaneously modeled using the same exponential terms in the equation $F_{Sj} = A_S + V_S \sum_{i=1}^3 a_i e^{-b_i(t_j+C_S)}$ for $i = 1$ to 3. An additional adjustable scaling factor V_S compensated for small differences between sample volumes between long- and short-term experiments, and ranged from 0.97 to 1.04. The objective function minimized the square root of the sum of the squares of the difference between the measured fluorescence data (short-term and long-term progress curves) and the models, evaluated over the 250 data points. Two independent long-term experiments and four independent short-term experiments were collected for each GFP variant, and up to eight pair-wise combinations of data (long- and short-term) were fitted simultaneously ($250 \times 2 = 500$ data points). Taken over both the short- and long-term progress curves together, three-exponential fits improved the goodness-of-fit (r.m.s.d.) fourfold relative to two-exponential fits, whereas four-exponential fits failed to improve the goodness-of-fit substantially (see **Supplementary Figs. 3–6** online). The EXCEL worksheet is available upon request (waldo@lanl.gov). Equivalent results were obtained whether the total fluorescence at infinite time was allowed to float, or was fixed at the fluorescence value measured at 15 h, indicating that the kinetic measurements were sampled for a sufficient time to constrain the kinetic parameters in the fits. Initial rates of the initial rapid fluorescence increase were also estimated by fitting a 2nd order polynomial to the first 1.5 s of short-term progress curves using the SOLVER function of EXCEL, and evaluating the first derivative at the injection time ($t = 0$) (**Fig. 2a**). Equilibrium fluorescence values were measured by diluting urea denatured GFP variants (see above) into TNG containing 5 mM DTT to various final urea concentrations between 0.63 and 6.2 M in increments of 0.1 M urea, and allowing refolding to proceed up to 52 at 15°C . Fluorescence values were measured using a FL600 Microplate Fluorescence Reader (488-nm excitation, 530-nm emission, 10-nm band pass) and scaled by dividing by the fluorescence levels of corresponding nondenatured samples diluted in parallel as a reference. Midpoint recovery concentrations of urea C_m (recovery of 50% of the initial fluorescence) were determined from sigmoidal fits using SOLVER in EXCEL, to the scaled fluorescence value F using the equation $F_j = a + b/(1 + (C_j/C_m)^h)$, where a , b , C_m and h are adjustable parameters, and C_j is the molarity of the urea in the refolding experiment j . The data were used to calculate the dependence of the standard free energy of denaturation, $\Delta G^\circ = -RT \ln K$, on urea concentration, where R is the gas constant, T is the absolute temperature and K is the equilibrium constant, which can be calculated from the experimental data by using the standard equation⁴⁵ $K = [(y)N - (y)]/[(y) - (y)D]$, where (y) is the observed value of the parameter used to follow unfolding, and $(y)N$ and $(y)D$ are the (y) values for the native state and the denatured state, respectively, under the same conditions under which (y) was measured. The conformational stability of the various GFP variants was subsequently estimated using the equation $\Delta G^\circ = \Delta G(\text{H}_2\text{O}) - m[\text{urea}]$ where $\Delta G(\text{H}_2\text{O})$ is an estimate of the conformational stability of a protein that assumes that the linear dependence continues to infinite dilution of denaturant, and m is a measure of the dependence of ΔG on urea concentration⁴⁶.

Expression and purification of GFP variants. Single-colony transformants of either folding reporter GFP, superfolder GFP or the single-point mutants S30R, Y39N, N105T, Y145E, I171V or A206V of folding reporter GFP each bearing a C-terminal 6HIS tag in *E. coli* BL21(DE3) were grown in LB medium containing kanamycin (35 $\mu\text{g/ml}$). One colony was picked, inoculated in 3 ml of the same medium, shaken overnight at 30°C , and used to inoculate

a 1-L culture that was grown to mid-log phase at 37 °C and subsequently induced with 1 mM IPTG for 6 h. The cell pellets were harvested by centrifugation at 5,500g for 15 min at 5 °C, and stored at -80°C. Cells were disrupted by sonication and the cell-free extract was centrifuged (100,000g, 30 min at 5 °C) and the supernatant loaded onto a 10-ml volume metal affinity resin (Talon resin, Clontech) equilibrated in buffer A (150 mM NaCl, 100 mM HEPES-NaOH, pH 7.5). Unbound proteins were washed off with buffer A containing 10 mM imidazole. The bound protein was then eluted with buffer B (200 mM imidazole, 150 mM NaCl, 100 mM HEPES-NaOH, pH 7.5) to a final volume of 15 ml. Ammonium sulfate was added to 80% saturation (~0.48 mg added to 1 ml of protein solution) at 20 °C. The solution was stirred for 15 min at the same temperature until dissolved, then incubated on ice for an additional 30 min. The mixture containing the precipitated protein was centrifuged and the supernatant discarded. The precipitate was progressively dissolved in 3 ml buffer C (20 mM HEPES-NaOH, pH 7.5), and the protein solution was dialyzed overnight against a 500-fold (vol/vol excess of the same buffer.

Crystallization. The C-terminally His-tagged proteins were subjected to trypsin proteolysis (Sigma) following a protocol similar to one described²⁰. Equal volumes of purified GFP solution at 15 mg/ml (using an extinction coefficient at 490 nm of $39.2 \times 10^3 \text{ M}^{-1}\text{cm}^{-1}$) and trypsin at 0.25 mg/ml were incubated for 30 min at 37 °C. The degradation reaction was stopped by the addition of 0.5 mM PMSF (phenylmethylsulfonyl fluoride) and the mixture was reapplied onto the metal affinity resin to remove any undigested protein containing the intact 6HIS tag. Digested protein was concentrated to ~15 mg/ml using an Amicon Microcon-10 concentrator (Millipore), and characterized by SDS-PAGE. Wizard II (Hampton Research), crystallization conditions no. 3 (20% PEG 8000, 0.1 M Tris, pH 8.5, 0.2 M MgCl₂), and no. 18 (20% PEG 3000, 0.1 M Tris, pH 7, 0.2 M calcium acetate) were first identified for the folding reporter GFP at 22°C in hanging drops containing 2 µl of the concentrated protein and 1 µl of a well solution. Refinement of condition no. 18 resulted in plates of good quality which grew to their final size (~0.5-mm long and 0.05-mm thick) after a period of 2 d equilibration against a reservoir solution containing 20% PEG 3000, 0.1 M Tris (pH 7 to 9.5) and 0.1–0.3 M calcium acetate. The best superfolder GFP crystals were obtained by mixing 2 µl of the ~15 mg/ml protein eightfold diluted in 20 mM HEPES-NaOH, pH 7.5 of the concentrated protein with 1 µl of a reservoir solution containing 0.1 M HEPES-NaOH, pH 7.5, 1 M sodium acetate and 0.05 M cadmium sulfate. Crystals appeared after 2 d of equilibration and reached their maximum size of 0.5 mm × 0.5 mm × 0.3 mm within 2 weeks. Crystals of folding reporter and superfolder variants of GFP were rapidly transferred into paratone oil using a crystal transfer loop (Hampton Research), and flash frozen in liquid nitrogen before placing them on a goniometer head centered in a cold nitrogen-gas stream.

Data collection, molecular replacement and refinement. Using 1.07 Å synchrotron radiation, monochromatic data sets were collected for the folding reporter and superfolder GFP variants on beamline X8C at the National Synchrotron Light Source (NSLS). Crystal data collection statistics are summarized in **Supplementary Table 3** online. Folding reporter GFP crystallized in space group P2₁2₁2₁ with unit cell dimensions a = 85.4 Å, b = 87.2 Å, and c = 145.6 Å and four molecules per asymmetric unit. The crystal system of superfolder GFP was assigned as P3₁21, with the unit-cell dimensions a = b = 88.4 Å, c = 69.4 Å and $\gamma = 120^\circ$. Reflection intensities were processed with the program *MOSFLM* v6.0⁴⁷. The program *SCALA* of the CCP4 suite of programs was used to merge and scale these intensities and compute the structure-factor amplitudes⁴⁷. In addition, self-rotation functions were calculated using the program *POLARRFN*⁴⁷ to confirm the symmetry and the number of molecules present in the asymmetric unit.

Both crystal structures were determined by molecular replacement on the basis of the F99S/M153T/V163A model (Protein Data Bank ID code 1B9C). Molecular replacement and crystallographic refinement were carried out with the CNS package⁴⁸, and model building was performed by using the program Turbo-Frodo (<http://www.afmb.univ-mrs.fr/~TURBO->). The final models of the folding reporter and superfolder GFP have R values of 20.6%. Crystallographic data are given in **Supplementary Table 3** online. Clear density starts at residue 3 and extends throughout the entire molecule, with the exception of the side chains of residues 12, 16, 45 and 68.

PDB accession numbers. FR-GFP and SF-GFP structure factors and coordinates have been deposited in the RCSB PDB under accession numbers 2B3Q and 2B3P, respectively.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

The authors wish to acknowledge Hong Cai for help in collecting flow cytometry data, Brian Mark for helpful comments, and the NIH and LDRD-DR for generous support.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Biotechnology* website for details).

Published online at <http://www.nature.com/naturebiotechnology/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

1. Tsien, R.Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998).
2. Cormack, B.P., Valdivia, R.H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).
3. Cramer, A., Whitehorn, E.A., Tate, E. & Stemmer, W.P. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* **14**, 315–319 (1996).
4. Tsukamoto, T. *et al.* Visualization of gene activity in living cells. *Nat. Cell. Biol.* **2**, 871–878 (2000).
5. Ayoub, J.C., Shaner, N.C., Sanger, J.W. & Sanger, J.M. Expression of green or red fluorescent protein (GFP or DsRed) linked proteins in non-muscle and muscle cells. *Mol. Biotechnol.* **17**, 65–71 (2001).
6. Babiychuk, E., Van Montagu, M. & Kushnir, S. N-terminal domains of plant poly (ADP-ribose) polymerases define their association with mitotic chromosomes. *Plant J.* **28**, 245–255 (2001).
7. Bachi, A. *et al.* The C-terminal domain of TAP interacts with the nuclear pore complex and promotes export of specific CTE-bearing RNA substrates. *RNA* **6**, 136–158 (2000).
8. Brumwell, C., Antolik, C., Carson, J.H. & Barabrese, E. Intracellular trafficking of hnRNP A2 in oligodendrocytes. *Exp. Cell Res.* **279**, 310–320 (2002).
9. Waldo, G.S., Standish, B.M., Berendzen, J. & Terwilliger, T.C. Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* **17**, 691–695 (1999).
10. Waldo, G.S. Genetic screens and directed evolution for protein solubility. *Curr Opin. Chem. Biol.* **7**, 33–38 (2003).
11. Nakayama, M. & Ohara, O. A system using convertible vectors for screening soluble recombinant proteins produced in *Escherichia coli* from randomly fragmented cDNAs. *Biochem. Biophys. Res. Commun.* **312**, 825–830 (2003).
12. Pedelacq, J.D. *et al.* Engineering soluble proteins for structural genomics. *Nat. Biotechnol.* **20**, 927–932 (2002).
13. Peelle, B. *et al.* Intracellular protein scaffold-mediated display of random peptide libraries for phenotypic screens in mammalian cells. *Chem. Biol.* **8**, 521–534 (2001).
14. Baird, G.S., Zacharias, D.A. & Tsien, R.Y. Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl. Acad. Sci. USA* **96**, 11241–11246 (1999).
15. Topell, S., Hennecke, J. & Glockscher, R. Circularly permuted variants of the green fluorescent protein. *FEBS Lett.* **457**, 283–289 (1999).
16. Topell, S. & Glockscher, R. Circular permutation of the green fluorescent protein. *Methods Mol. Biol.* **183**, 31–48 (2002).
17. Cramer, A., Whitehorn, E.A., Tate, E. & Stemmer, W.P. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* **14**, 315–319 (1996).
18. Patterson, G.H., Knobel, S.M., Sharif, W.D., Kain, S.R. & Piston, D.W. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys. J.* **73**, 2782–2790 (1997).
19. Heim, R., Prasher, D.C. & Tsien, R.Y. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. USA* **91**, 12501–12504 (1994).
20. Orm6, M. *et al.* Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**, 1392–1395 (1996).
21. Bevis, B.J. & Glick, B.S. Rapidly maturing variants of the *Discosoma* red fluorescent protein (DsRed). *Nat. Biotechnol.* **20**, 83–87 (2002).
22. Battistutta, R., Negro, A. & Zanotti, G. Crystal structure and refolding properties of the mutant F99S/M153T/V163A of the green fluorescent protein. *Proteins* **41**, 429–437 (2000).
23. Stepanenko, O.V. *et al.* Comparative studies on the structure and stability of fluorescent proteins EGFP, zFP506, mRFP1, “dimer2”, and DsRed1. *Biochemistry* **43**, 14913–14923 (2004).
24. Cabantous, S., Terwilliger, T.C. & Waldo, G.S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol.* **23**, 102–107 (2005).
25. Phillips, G.N., Jr. Structure and dynamics of green fluorescent protein. *Curr Opin Struct. Biol.* **7**, 821–827 (1997).

26. Zacharias, D.A., Violin, J.D., Newton, A.C. & Tsien, R.Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913–916 (2002).
27. Dao-pin, S. *et al.* Structural and genetic analysis of electrostatic and other interactions in bacteriophage T4 lysozyme. *Ciba Found. Symp.* **161**, 52–62 (1991).
28. Yip, K.S. *et al.* The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure* **3**, 1147–1158 (1995).
29. Goldman, A. How to make my blood boil. *Structure* **3**, 1277–1279 (1995).
30. Shagin, D.A. *et al.* GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. *Mol. Biol. Evol.* **21**, 841–850 (2004).
31. Magliery, T.J. & Regan, L. Combinatorial approaches to protein stability and structure. *Eur. J. Biochem.* **271**, 1595–1608 (2004).
32. Minor, D.L., Jr. & Kim, P.S. Measurement of the beta-sheet-forming propensities of amino acids. *Nature* **367**, 660–663 (1994).
33. Ewert, S., Honegger, A. & Pluckthun, A. Structure-based improvement of the biophysical properties of immunoglobulin VH domains with a generalizable approach. *Biochemistry* **42**, 1517–1528 (2003).
34. Fane, B. & King, J. Intragenic suppressors of folding defects in the P22 tailspike protein. *Genetics* **127**, 263–277 (1991).
35. Fane, B., Villafane, R., Mitraki, A. & King, J. Identification of global suppressors for temperature-sensitive folding mutations of the P22 tailspike protein. *J. Biol. Chem.* **266**, 11640–11648 (1991).
36. Mitraki, A., Danner, M., King, J. & Seckler, R. Temperature-sensitive mutations and second-site suppressor substitutions affect folding of the P22 tailspike protein in vitro. *J. Biol. Chem.* **268**, 20071–20075 (1993).
37. Mitraki, A., Fane, B., Haase-Pettingell, C., Sturtevant, J. & King, J. Global suppression of protein folding defects and inclusion body formation. *Science* **253**, 54–58 (1991).
38. Sideraki, V., Huang, W., Palzkill, T. & Gilbert, H.F. A secondary drug resistance mutation of TEM-1 beta-lactamase that suppresses misfolding and aggregation. *Proc. Natl. Acad. Sci. USA* **98**, 283–288 (2001).
39. Huang, W. & Palzkill, T. A natural polymorphism in beta-lactamase is a global suppressor. *Proc. Natl. Acad. Sci. USA* **94**, 8801–8806 (1997).
40. Davis, T.N. Protein localization in proteomics. *Curr. Opin. Chem. Biol.* **8**, 49–53 (2004).
41. Adams, S.R. *et al.* New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. *J. Am. Chem. Soc.* **124**, 6063–6076 (2002).
42. Martin, B.R., Giepmans, B.N., Adams, S.R. & Tsien, R.Y. Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. *Nat. Biotechnol.* **23**, 1308–1314 (2005).
43. Gautier, I. *et al.* Homo-FRET microscopy in living cells to measure monomer-dimer transition of GFT-tagged proteins. *Biophys. J.* **80**, 3000–3008 (2001).
44. Stemmer, W.P. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**, 10747–10751 (1994).
45. Tanford, C. Protein denaturation. *Adv. Protein Chem.* **23**, 121–282 (1968).
46. Pace, C.N. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266–280 (1986).
47. Collaborative Computational Project, N. The CCP4 Suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
48. Brunger, A.T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921 (1998).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.