nature biotechnology

Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein

Stéphanie Cabantous, Thomas C Terwilliger & Geoffrey S Waldo

Existing protein tagging and detection methods are powerful but have drawbacks. Split protein tags can perturb protein solubility[1–4] or may not work in living cells[5–7]. Green fluorescent protein (GFP) fusions can misfold[8] or exhibit altered processing[9]. Fluorogenic biarsenical FLaSH or ReASH[10] substrates overcome many of these limitations but require a polycysteine tag motif, a reducing environment and cell transfection or permeabilization[10]. An ideal protein tag would be genetically encoded, would work both *in vivo* and *in vitro*, would provide a sensitive analytical signal and would not require external chemical reagents or substrates. One way to accomplish this might be with a split GFP[11], but the GFP fragments reported thus far are large and fold poorly[11,12], require chemical ligation[13] or fused interacting partners to force their association[11–14], or require coexpression or co-refolding to produce detectable folded and fluorescent GFP[11,12]. We have engineered soluble, self-associating fragments of GFP that can be used to tag and detect either soluble or insoluble proteins in living cells or cell lysates. The split GFP system is simple and does not change fusion protein solubility.

To achieve the split GFP protein tagging and detection scheme outlined in **Figure 1a**, we first tested several pairs of fragments from either the folding reporter GFP, which contains the mutations F99S, M153T, V163A[15], F64L and S65T[16], or the exceptionally stable 'superfolder' GFP, containing the folding reporter GFP mutations and S30R, Y39N, N105T, Y145F, I171V and A206V (G.S.W., data not shown). Coexpression of the superfolder GFP fragments containing amino acids 1–214 (GFP 1–10) and 214–230 (GFP 11) gave fluorescent *Escherichia coli* colonies. No detectable complementation occurred with the corresponding folding reporter GFP fragments (**Fig. 1b**). Superfolder GFP 1–10 was insoluble, but incubation of refolded inclusion bodies with soluble sulfite reductase from *Pyrobaculum aerophilum*[17], C-terminally tagged with wild-type GFP 11 (sulfite reductase–GFP 11) resulted in a time-dependent increase in fluorescence (**Fig. 1b**).

We evolved superfolder GFP 1–10 by DNA shuffling[18] to improve its solubility and increase its complementation with sulfite reductase–GFP 11. After three rounds of shuffling and selection of the brightest clones, *in vitro* complementation of the soluble lysate of the best variant, termed GFP 1–10 OPT, improved 80-fold (**Fig. 1b**) relative to the same amount of refolded superfolder GFP 1–10. In addition to the folding reporter GFP mutations, GFP 1–10 OPT contains S30R, Y145F, I171V and A206V substitutions from superfolder GFP and seven new mutations: N39I, T105K, E111V, I128T, K166T, I167V and S205T. This protein is ~50% soluble when expressed in *E. coli* at 37 °C (data not shown). Ultraviolet-visible spectra of 10 mg/ml solutions of the nonfluorescent GFP 1–10 OPT lacked the 480 nm absorption band of the red-shifted GFP[19] (data not shown), suggesting that the addition of GFP 11 triggers a folding step required to generate the cyclized chromophore[19]. Purified GFP 1–10 OPT, superfolder GFP and folding reporter GFP were each studied by analytical gel filtration loaded at 10 mg/ml. GFP 1–10 OPT eluted as 60% dimer, 35% monomer and 5% higher-order aggregates, whereas the full-length folding reporter GFP and superfolder GFP both eluted as >95% monomer, with a trace of dimer and higher-order aggregates (data not shown).

The C-terminal wild-type GFP 11 fusion tag dramatically reduced the solubility of several *P. aerophilum*[17] test proteins (**Table 1**). 3-hexulose 6-phosphate synthase (HPS) alone was 60% soluble but insoluble when fused to wild-type GFP 11 (**Fig. 1c**, **Table 1**). We used HPS as 'bait' in a directed evolution schema in *E. coli* to discover mutants of GFP 11 for which the HPS-GFP 11 fusion solubility matched that of HPS alone. Libraries of HPS-GFP 11 variants and GFP 1–10 OPT were expressed in succession from independently inducible compatible plasmids (Methods and **Supplementary Fig. 1** online) to avoid false positives caused by cotranslational folding and complementation of insoluble variants of HPS-GFP 11 with GFP 1–10 OPT (**Supplementary Fig. 2** online). The best GFP 11 construct (L221H, F223Y, T225N) (termed GFP 11 M3, **Supplementary Table 1** online) balanced reduced perturbation of fusion protein solubility (**Fig. 1c**, **Table 1**) with good complementation (**Fig. 1d**).

We measured fluorescence progress curves for complementation of purified sulfite reductase–GFP 11 M3 in 200 µl reactions in a microtiter plate (**Fig. 2**). We avoided potential higher-order kinetic effects by initiating the complementation using a high concentration and large molar-excess of GFP 1–10 OPT (800 pmol). Complementation fluorescence was a linear function of analyte concentration (**Fig. 2a**), and progress curves over a wide concentration range could be superimposed by linear scaling (**Fig. 2b**). 0.1 – 200 pmol sulfite reductase–GFP 11 M3 could be accurately quantified within 1 h after the addition of GFP 1–10 OPT (**Fig. 2a**), and greater than

Bioscience Division, MS-M888, Los Alamos National Laboratory, PO Box 1663, Los Alamos, New Mexico 87545, USA. Correspondence should be addressed to G.S.W. (waldo@lanl.gov).

Published online 5 December 2004; doi:10.1038/nbt1044

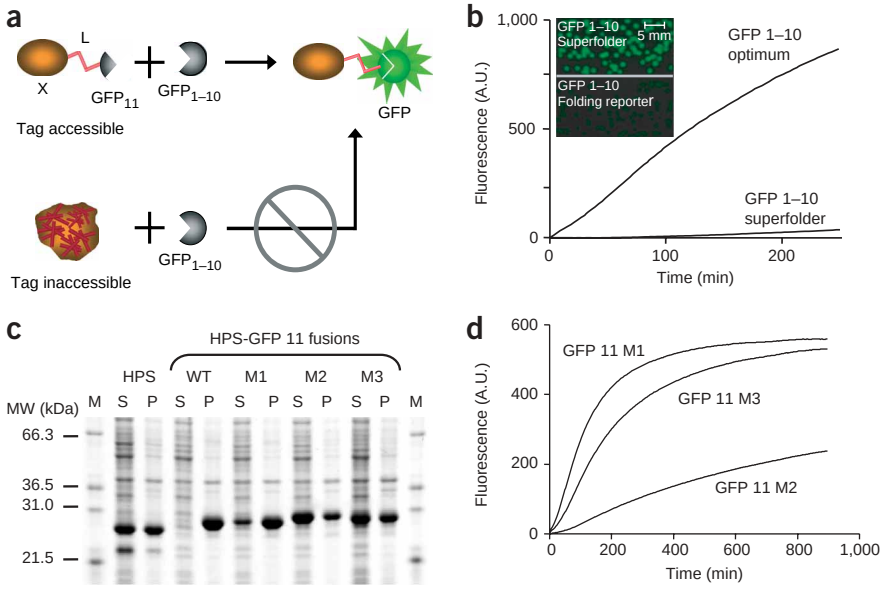

Figure 1 Principle of split GFP complementation and engineering fragments for improved solubility and complementation. (**a**) Split GFP fragment complementation. A protein of interest (X) is fused to a small GFP fragment (β-strand 11, residues 215–230) via a flexible linker (L). The complementary GFP fragment (β-strands 1–10, residues 1–214) is expressed separately. Neither fragment alone is fluorescent. When mixed, the small and large GFP fragments spontaneously associate, resulting in GFP folding and formation of the fluorophore. Processes that make the small GFP tag inaccessible, such as misfolding or aggregation, can prevent complementation. (**b**) *In vitro* complementation efficiency of GFP 1–10 variants. Fluorescence progress curves for complementation of 20 μl of 1 mg/ml refolded superfolder GFP 1–10 (lower trace) or an equal amount of soluble optimized GFP 1–10 OPT fragment (upper trace) after addition of 180 μl buffer containing 1 mg/ml soluble sulfite reductase fused to wild-type GFP 11. Arbitrary fluorescence units (A.U.). *In vivo* complementation of GFP 1-10 variants (inset). Fluorescent images of *E. coli* BL21(DE3) colonies on nitrocellulose membranes coexpressing GFP 1–10 from superfolder GFP (top), or folding reporter GFP (bottom), along with sulfite reductase fused with wild-type GFP 11. (**c**) Engineering the split GFP fragments for improved performance. SDS-PAGE of soluble (S) and pellet fractions (P) of *E. coli* BL21(DE3) cells expressing the protein hexulose phosphate synthase (HPS) alone, HPS fused to GFP 11 (WT) or HPS fused to the three GFP 11 optima (M1, M2, M3). (**d**) Fluorescence complementation kinetic traces for the three GFP 11 mutants fused to sulfite reductase (50 pmol) after the addition of excess GFP 1–10 OPT (800 pmol) *in vitro*. The final volume of each assay was 200 μl. Arbitrary fluorescence units (A.U.).

10 pmol protein required only 15 min before quantification (data not shown). Fluorescence assays showed that folding reporter GFP bearing a C-terminal GFP 11 M3 tag rapidly bound to 6-His–GFP 1–10 OPT immobilized on Talon resin (Novagen). The washed beads gained additional fluorescence at a rate comparable to that observed in solution (**Fig. 2c**), indicating that the kinetics of fluorescence formation was not limited by the rate of association of the GFP fragments.

Next, we tested the effect of chemical adjuvants. *In vitro* complementation was improved ~30% by 5 mM dithiothreitol but quenched by 0.1% wt/vol SDS or >2 M urea (**Supplementary Fig. 3** online). Complementation was inefficient below pH 6.5, with an apparent pK_a of 7.3 at 25 °C. (**Supplementary Fig. 4** online). After complementation, the fluorescent GFP moiety displayed a slow time-dependent decrease in fluorescence in >5 M urea ($t_{1/2}$ ≈ 20 h). Complemented split GFP had an apparent pK_a of 5.5 at 25 °C, similar to that of 'enhanced' GFP[16] (data not shown).

To test whether the split GFP system could accurately quantify different proteins *in vitro*, we expressed 18 *P. aerophilum* proteins as pET vector constructs with C-terminal GFP 11 M3 tags in liquid culture and then analyzed the soluble and pellet fractions using SDS-PAGE (**Fig. 3a**) and the split GFP complementation system. Aliquots of the soluble fraction were assayed by adding excess GFP 1–10 OPT to

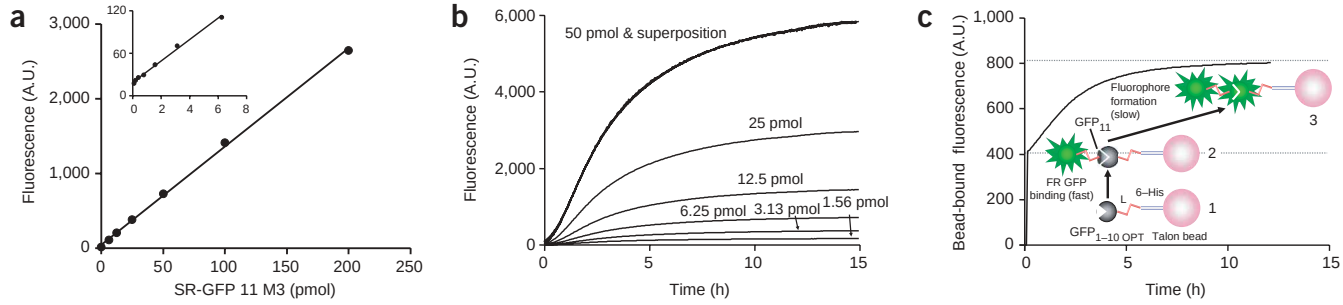

Figure 2 *In vitro* characterization of split GFP complementation reaction. (**a**) Sensitivity of split GFP complementation. 20 μl aliquots containing 0.1 to 200 pmol of GFP 11 M3-tagged sulfite reductase were mixed with 180 μl aliquots containing 800 pmol GFP 1–10 OPT to start complementation. Fluorescence measured for each solution 1 h after addition of GFP 1–10 OPT. Arbitrary fluorescence units (A.U.). (**b**) Superimposition of progress curves for complementation of 50, 25, 12.5, 6.25, 3.13 and 1.56 pmol samples. The data were fit to the 50 pmol progress curve by subtracting a small constant and applying a scaling factor, calculated by nonlinear least-squares using the Excel data analysis tool Solver. Scaling factors are: 1.00, (50 pmol); 1.96, (25 pmol); 4.02, (12.5 pmol); 8.15, (6.25 pmol); 15.64, (3.13 pmol); 36.52, (1.56 pmol). The excellent superposition indicates that the shape of the progress curve does not depend on the concentration of the tagged protein, or depletion of the pool of unbound GFP 1–10 OPT fragment. Arbitrary fluorescence units (A.U.). (**c**) Complementation of Talon resin-bound 6-His GFP 1–10 OPT by folding reporter GFP tagged with C-terminal GFP 11 M3. (1) Talon resin with bound 6-His GFP 1–10 OPT, (2) rapid increase in bead-bound fluorescence by binding of folding reporter GFP via fused C-terminal GFP 11 M3, (3) slow fluorescence formation due to complementation. Arbitrary fluorescence units (A.U.).

initiate complementation. Even though Coomassie dye exhibits protein-dependent variations in staining efficiency[20], after the completion of complementation and folding (~6 h), there was a strong correlation between the measured fluorescence values and the amount of protein as visualized by SDS-PAGE (**Fig. 3a**). Insoluble proteins dissolved in 9 M urea and diluted 20-fold with buffer containing excess GFP 1–10 OPT gave fluorescence that correlated well with the amount of insoluble protein visualized by SDS-PAGE (**Fig. 3a**). In contrast, when solubilized pellets were diluted with fresh buffer before the addition of an aliquot of concentrated GFP 1–10 OPT, several of the well-expressed insoluble proteins (that is, polysulfide reductase and nucleotide diphosphate kinase, **Table 1** and **Fig. 3a**) gave no detectable complementation (data not shown). These proteins had probably misfolded and aggregated upon dilution, making the GFP 11 M3 tag inaccessible before the addition of the GFP 1–10 OPT moiety (**Fig. 1a**).

A practical split protein tagging system could be used *in vivo* to label and detect either soluble or insoluble proteins. We theorized that soluble protein could be assayed in living *E. coli* cells by first expressing the tagged protein for a limited time and then shutting off the expression to allow the tagged protein to develop its intrinsic solubility phenotype before expression of the complementary GFP fragment in the same cellular compartment. From the results of our co-refolding *in vitro* pellet assays, we expected that coexpressing the GFP 11 M3 tagged protein and GFP 1–10 OPT would lead to structural complementation and commitment to the development of GFP fluorescence before the aggregation of the test protein *in vivo*, enabling an estimate of the total expressed protein.

We expressed each of the 18 *P. aerophilum* test proteins with N-terminal 6-His and C-terminal GFP 11 M3 tags using the modified pTET plasmid (**Supplementary Table 2** and **Supplementary Fig. 1** online) in an *E. coli* BL21(DE3) strain bearing GFP 1–10 OPT on a pET plasmid. The fluorescent colonies were imaged after coexpression or after sequential expression, and soluble and pellet fractions of the same constructs were analyzed by SDS-PAGE (**Fig. 3b**) after sequential induction in liquid culture. We assessed the amount of useful, nonaggregated 6-His-tagged protein by binding soluble fractions to an excess of Talon resin (Novagen) before the SDS-PAGE analyses. The *in vivo* colony fluorescence assay after coinduction reported total protein in agreement with SDS-PAGE, whereas colony fluorescence of cells subjected to sequential induction agreed with SDS-PAGE of Talon-bound soluble protein (**Fig. 3b**). Colonies expressing highly soluble proteins were bright both when GFP 1–10 was coinduced and sequentially induced (proteins 2, 4 and 5; **Fig. 3b**). Colonies expressing insoluble proteins were much brighter when GFP 1–10 was coinduced (proteins 8, 11, 15, 16 and 18; **Fig. 3b**). Proteins 1, 4, 5, 7, 9, 12 and 14 were each less soluble when expressed from the very strong *T7* promoter[21] of the pET system (**Table 1** and **Fig. 3a**) than from the weaker *tet* promoter[22] of the pTET plasmid (**Fig. 3b**). The influence of promoter strength on protein expression levels and solubility has been noted previously[23,24,26,27].

Until this work, no split protein fragment complementation systems had been specifically engineered to minimize perturbation of reporter tags on fusion protein folding or solubility. Our methods could be used to improve the solubility and complementation of other protein fragments derived from GFP[11,12] or from enzymes such as β-galactosidase (*LacZ*) or dihydrofolate reductase (DHFR). Proteins can be tagged and detected *in vivo* using the split lacZ system[1,3,4], but the lacZα tag can decrease protein solubility. Maltose binding protein-lacZα fusions were reported as ~60% soluble[4], even though maltose binding protein expressed alone is fully soluble[28]. Our method could be used to engineer versions of the lacZα tag that do not reduce fusion protein solubility. Fragments of split DHFR used to detect protein-protein interactions are poorly folded and mostly insoluble (USPTO 6270964). Our approach could be used to engineer DHFR fragments with improved solubility and reduced auto-association[29,30], making them more suitable for detecting protein-protein interactions.

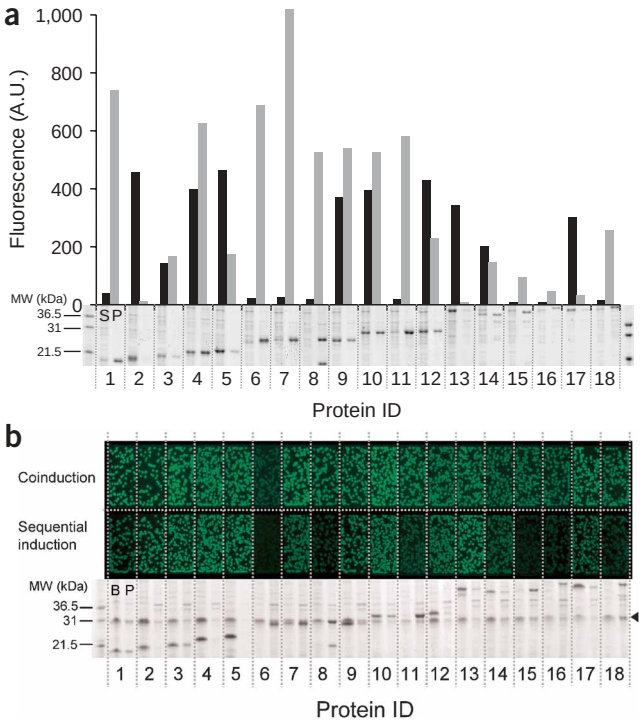

Figure 3 *In vitro* protein quantification and *in vivo* protein expression and solubility screens. (**a**) Protein quantification of eighteen *P. aerophilum* test proteins (see Legend, **Table 1**) expressed as N-terminal fusions with GFP 11 M3. The GFP fragment complementation assay fluorescence of soluble (black bars) and unfolded pellet fractions (gray bars) using GFP 1–10 OPT (top). SDS-PAGE of the corresponding soluble (S), and pellet fractions (P) (bottom). Note that protein #8, tartrate dehydratase β-subunit, shows a second lower band at ~13 kDa. Arbitrary fluorescence units (A.U.). (**b**) *In vivo* solubility and expression screen using split GFP. Eighteen *P. aerophilum* test proteins (see **Table 1**) expressed with an N-terminal 6-His tag and a C-terminal GFP 11 M3 tag from a *tet* promoter plasmid, were cloned into an *E. coli* BL21 (DE3) strain containing a pET plasmid expressing GFP 1–10 OPT. Fluorescence images of colonies on plates after coinduction of the tagged constructs and GFP 1–10 OPT (top), or transient expression of the tagged constructs followed by expression of the GFP 1–10 OPT (middle). Scale of colony images as in **Fig. 1b**. SDS-PAGE of Talon resin bead-bound soluble (B) and pellet fractions (P) from cells sequentially induced in liquid culture (bottom). Adventitiously-bound GFP 1–10 OPT (apparent molecular weight ~29 kDa) is indicated by arrow. Note that nucleoside diphosphate kinase (protein #7) is partially soluble (see band slightly below band corresponding to GFP 1–10 OPT in Talon resin-bound fraction). Polysulfide reductase-GFP 11 M3 fusions (**Table 1**) produced intensely red-colored colonies, absorbing the 488 nm excitation light and reducing whole-cell fluorescence during coexpression despite the good expression of the protein.

The split GFP has many practical applications. Obtaining soluble, well-folded recombinant proteins for downstream applications[23–27,31,32] requires screening large numbers of protein variants (mutants, fragments, fusion tags, folding partners) and testing many expression or refolding conditions[24,33]. Recently, several methods for screening soluble protein expression have been described[34], including antibody-based detection of polyhistidine-tagged proteins using dot

Table 1 Effect of GFP 11 tags on the solubility of eighteen proteins from *Pyrobaculum aerophilum*

#	[b]Protein	[a]Fraction soluble					
		[c]MW	[d]NF	[e]WT	[f]M1	[f]M2	[f]M3
1	DNA-directed RNA polymerase	12.5	0.05	0.00	0.00	0.35	0.10
2	Sulfite reductase (dissimilatory subunit)	12.7	1.00	1.00	1.00	1.00	1.00
3	c-type cytochrome biogenesis factor	14.4	0.77	0.28	0.59	0.86	0.65
4	Translation initiation factor	15.4	0.40	0.30	0.80	0.70	0.45
5	Ribosomal protein S9p	16.4	0.70	0.50	0.75	0.80	0.75
6	Polysulfide reductase subunit	21.0	0.00	0.00	0.00	0.00	0.00
7	Nucleoside diphosphate kinase	21.6	0.00	0.00	0.00	0.15	0.10
8	Tartrate dehydratase β-subunit	23.8	0.00	0.00	0.00	0.00	0.00
9	3-hexulose 6-phosphate synthase	23.1	0.65	0.00	0.30	0.85	0.60
10	Hydrogenase formation protein hypE	26.8	0.35	0.05	0.40	0.70	0.55
11	Methyltransferase	29.3	0.00	0.00	0.00	0.05	0.05
12	Chorismate mutase	29.3	0.70	0.00	0.35	0.65	0.70
13	Tyrosine t-RNA synthetase	36.0	0.95	0.70	0.90	0.90	0.95
14	nirD protein	36.7	0.70	0.15	0.40	0.65	0.45
15	Soluble hydrogenase	37.3	0.00	0.00	0.00	0.05	0.00
16	Aspartate-semialdehyde dehydrogenase	37.4	0.00	0.00	0.00	0.05	0.00
17	Phosphate cyclase	37.4	0.80	0.30	0.85	0.95	0.90
18	Purine-nucleoside phosphorylase	41.7	0.05	0.00	0.00	0.10	0.00

[a]The fraction soluble is given, as determined by SDS-PAGE densitometry (Fraction soluble). Relative uncertainty is ~5%, average of three replicates. [b]Eighteen proteins from the hyperthermophilic archaeon *P. aerophilum*[17] were expressed in *E. coli* BL21(DE3) at 37 °C. [c]Theoretical molecular weight in kDa was calculated from amino acid sequence (MW). [d]Nonfusion (NF) solubility. [e]Solubility of C-terminal fusions with wild-type GFP 11 (WT). [f]Solubility of C-terminal fusions with GFP 11 optima (M1, M2, M3).

blots[35] and a commercial system for detecting proteins labeled with the S-peptide (FretWorks, Novagen), but these assays require multiple steps and do not work in living cells.

Split GFP is a simple, genetically encoded, nonperturbing reporter of protein solubility and aggregation *in vivo* and *in vitro*, making it well suited for high-throughput solubility screens of libraries of protein domains, fragments and mutants. About 10^3–10^5 clones can be readily screened by colony fluorescence. Larger libraries of up to 10^6 variants can be screened by flow cytometry. GFP 11 M3 balances good complementation with minimal perturbation of fused protein solubility, whereas the more perturbing GFP 11 wild-type or GFP 11 M1 tags could be used during *in vivo* solubility screens to favor the selection of exceptionally soluble proteins. The GFP 11 tag could be randomly inserted into protein scaffoldings, and then complemented with GFP 1–10 to search for internal permissive sites and to create new fluorescent protein adducts. Protein trafficking into mitochondria has been monitored using a split intein-split GFP fusion tag[36], using structural complementation of the intein fragments to covalently splice the fused GFP fragments. Our split GFP requires no accessory proteins or splicing, and the GFP 11 tag is less likely to perturb protein folding and solubility compared to large split intein-split GFP fusion tags[36]. Recently split GFP was used to monitor combinatorial promoter activity of several genes in specialized cells of *Caenorhabditis elegans*[37]. The split GFP fragments used required fused interacting coiled-coils for association[11] and appeared unstable in *C. elegans*. Complementation was efficient only when the fragments were coexpressed[37]. Our self-associating, soluble, stable engineered split GFP domains are simpler to use and would likely provide increased sensitivity, especially when monitoring weak promoter activity.

The split GFP system is not perfect. Apparently soluble proteins might nonetheless be aggregated[23,26,33], or the tag might be buried in correctly folded protein, making the GFP 11 M3 tag inaccessible. Proteins soluble *in vivo* might become insoluble *in vitro*. Split GFP is not a dynamic reporter of solubility. Once formed, the fluorescence is stable under physiological conditions. Split GFP complementation would be inefficient in acidic cellular compartments, and the system lacks the signal amplification characteristic of enzymes such as lacZ[3,4] or RNase A[6,7]. Tagged gene products that are cytotoxic or inhibitors of translation might interfere with the subsequent expression of GFP 1–10 OPT in living cells, but this limitation could likely be overcome in mammalian cells by using protein transfection reagents[38] to import the GFP 1–10 moiety. Despite these potential caveats, the split GFP offers a simple, nonperturbing alternative to fusions with full-length GFP, greatly expanding the spatiotemporal options for monitoring promoter activity and *in situ* protein tagging and localization experiments.

METHODS

Finding a feasible split GFP fragment pair and construction reporter plasmids. Test genes from *P. aerophilum* were cloned in pET28a alone and in fusion with C-terminal GFP 11 tags using a linker sequence encoding $(GGGS)_2$. For evolution and *in vivo* solubility screening of test proteins, the GFP 11 cassette was amplified from the pET vector using vector-specific primers and moved to the modified pTET-Spec^R vector (see **Supplementary Methods** online) via the *Nco*I/*Xba*I restriction sites. The 1–10 cassette was moved to a non-polyhistidine–tagged pET vector. We separately coexpressed several pairs of GFP fragments on compatible plasmids in *E. coli*, including amino acids 1–145 and 145–238, 1–155 and 156–238, 1–171 and 171–238, 1–195 and 196–238, 1–214 and 214–238. The junction points corresponded to loops or turns between β-strands[19,39]. Fragment pairs from superfolder GFP consistently produced much brighter colonies than the same pairs from folding reporter GFP. Our objective was to minimize the size of one of the fragments for use as a protein tag, so we focused on the feasible pair with the smallest fragment (1–214 and 214–238). To further reduce the size of the tagging domain, we also tested 1–214 (GFP 1–10) for complementation with 214–230 (GFP 11), eliminating the residues that were disordered in the crystal structure (231–238)[19] from the small fragment (**Supplementary Table 1** online).

Engineering GFP 1–10. Superfolder GFP 1–10 PCR amplicons were subjected to DNA fragmentation and shuffling using published protocols[18]. The 1–10 cDNA library plasmid was transformed into an *E. coli* BL21 (DE3)

PRO expression strain (Clontech) containing the sulfite reductase–GFP 11 wild-type tagged protein on a pPROTET vector (Clontech). The expression library was plated on nitrocellulose membrane using two successive 400-fold dilutions of the stock at 1.0 $OD_{600\,nm}$ frozen in 20% glycerol/Luria-Bertani (LB) medium. After overnight growth at 37 °C, the membrane was transferred to an LB/Agar plate containing 50 μg/ml kanamycin, 35 μg/ml chloramphenicol and 50 μg/ml spectinomycin, plus 1 mM IPTG for 3 h at 37 °C and then moved onto a new plate containing the above antibiotics plus 600 ng/ml anhydrotetracycline (AnTET). Clones exhibiting the most rapid development of fluorescence were picked and stored as freezer stocks at –80 °C in 20% glycerol/LB medium. The clones were grown and induced with 1mM IPTG, and the soluble lysates were screened for complementation efficiency in an *in vitro* assay with an excess of purified sulfite reductase–GFP 11. The best candidates were pooled and subjected to another round of evolution. Mutations were confirmed by fluorescent dye terminator DNA sequencing.

Engineering GFP 11. Hexulose phosphate synthase-GFP 11 (HPS-GFP 11) fusions were amplified by PCR and shuffled using published protocols[18]. The GFP 11 mutant library was expressed as a C-terminal fusion with the bait protein HPS bearing an N-terminal 6-His tag, from a modified pTET plasmid with an AnTET-inducible *tet* promoter[22] (**Supplementary Table 2** and **Supplementary Fig. 1** online) and transformed into a BL21(DE3) strain expressing GFP 1–10 OPT from a modified pET vector containing a p15 replicon. Optima were screened using a sequential induction protocol as follows. After overnight growth at 37 °C, the nitrocellulose membrane bearing colonies was moved onto a selective LB/Agar Bauer plate containing 300 ng/ml AnTet for 3 h at 37 °C to express the HPS-GFP 11 library, transferred to a selective resting plate for 1 h to allow the AnTet to diffuse out of the colonies to shut off expression of the HPS-GFP 11, and finally moved to a selective LB/agar plate containing 1 mM IPTG for 2 h to induce expression of the complementary GFP 1–10 OPT from the pET plasmid. Since the HPS-GFP 11 wild-type construct was entirely insoluble, colonies expressing the HPS-GFP 11 wild type and GFP 1–10 OPT according to the sequential expression protocol were only faintly fluorescent. Brighter clones, ostensibly associated with more soluble HPS-GFP 11 optima, were picked into selective liquid culture 96-well tissue culture plates and saved at −80 °C as 20% glycerol stocks. The clones were grown in 1 ml liquid cultures and induced with 300 ng/ml AnTET. The soluble fractions were screened for complementation efficiency in an *in vitro* assay with an excess of purified GFP 1–10 OPT. Clones with the fastest complementation rates were selected and pooled for an additional round of evolution and screening. Two rounds of evolution yielded two separate GFP 11 mutants, L221H and T225N. We initially focused on the L221H variant, termed GFP 11 M1. This mutation did not entirely eliminate the deleterious effect of GFP 11 on fusion protein solubility. GFP 11 M2 was engineered by combining F223S, a mutation that substantially increased the solubility of a different split GFP fragment (G.S.W., unpublished data) with T225N (**Supplementary Table 1** online). HPS-GFP 11 M2 solubility was greatly improved relative to either HPS-GFP 11 M1 or HPS-GFP 11 wild type. The complementation rate of HPS-GFP 11 M2 with GFP 1–10 OPT had decreased about fivefold relative to HPS-GFP 11 M1 for comparable amounts of soluble fusion protein. To increase the complementation rate, we removed K214, a duplicate of the C-terminal residue of GFP 1–10 OPT and screened a 64-fold degeneracy library at the hotspot position 223 using a degenerate primer set and cloned the resulting variants of GFP 11 M2 as C-terminal fusions with HPS to search for more conservative mutations. The soluble fractions of ~200 clones were screened in an *in vitro* assay with GFP 1–10 OPT.

Expression and refolding of the large GFP 1–10 fragment. Pellets containing superfolder GFP 1–10 and GFP 1–10 OPT from 500 ml cell culture were separately resuspended in 15 ml 50 mM Tris pH 7.4, 0.1 M NaCl, 10% glycerol (TNG buffer) and washed three times with 5 ml of B-PER reagent (Pierce), then twice with 5 ml of TNG. 70 mg of the washed pellet was unfolded with 1 ml of 9 M urea and refolded by adding 10 volumes of TNG buffer. The soluble solutions were filtered through a 0.2 μm syringe filter. Protein quantification was performed using the Bio-Rad Protein Assay reagent (Bio-Rad). To make the GFP 1–10 OPT assay reagent solution, a 1 ml aliquot of the refolded GFP 1–10 OPT was diluted ~20-fold with TNG. Even though the engineered GFP 1–10 OPT was ~50% soluble, the pellet fraction was processed to take advantage of the enrichment and partial purification afforded by using inclusion bodies.

Metal affinity resin purification of soluble sulfite reductase–GFP 11 fusion proteins. The soluble *P. aerophilum*[17] protein sulfite reductase was cloned as an N-terminal fusion with GFP 11, GFP 11 M1, M2 or M3 fragments, in an N-terminal 6-His pET vector. A 200 ml culture of BL21(DE3) cells expressing each fusion was grown to $OD_{600\,nm}$ ~0.5, induced with 1 mM IPTG for 4 h at 37 °C, pelleted by centrifugation, resuspended in 2 ml TNG and sonicated. The soluble fractions were loaded onto Talon metal affinity resin (Clontech). After two washes with TNG loading buffer supplemented with 10 mM imidazole, 6-His–tagged proteins were eluted with 150 mM imidazole in TNG buffer, then dialyzed twice against 500 ml TNG to remove the imidazole. The proteins (~95% purity as determined by SDS-PAGE) were quantified using the Bio-Rad Protein Assay reagent (Bio-Rad) and diluted to 0.0325 mg/ml.

Complementation of bead-bound GFP 1–10 OPT. A 100 μl aliquot of 50% v/v slurry of Talon resin was saturated with GFP 1–10 OPT bearing an N-terminal 6-His affinity tag (200 μl of 2 mg/ml protein). The beads (50 μl bed volume) were washed three times with 300 μl of TNG buffer to remove unbound GFP 1–10 OPT. The remaining buffer was aspirated and discarded, and the fluorescence was measured in a 96-well microtiter plate. Excess folding reporter GFP-GFP 11 M3 fusion protein (200 μl of 5 mg/ml protein) was added to the beads, mixed by pipetting for 15 s, rapidly transferred to a small 0.2 μm spin filtration column and washed three times with 0.5 ml aliquots of TNG to remove unbound folding reporter GFP-GFP 11 M3 protein. This procedure required approximately 5 min. Beads were transferred to a fresh well of the microtiter plate and the fluorescence measured at 3 min intervals for 12 h.

***In vitro* complementation assays.** For sensitivity experiments, a 96-well microplate was first blocked with a solution of 0.5% bovine serum albumin (BSA) in TNG for 10 minutes. Twofold serial dilutions of Talon resin-purified (Clontech) 6-His–SR-GFP 11 M3 fusion protein were performed in the same buffer. The dilutions spanned the range 0. 1 to 200 pmol per 20 μl aliquot; the aliquots were added to 96-well plates, and then complementation was performed using a large excess (800 pmol) of GFP 1–10 OPT (~0.5 mg/ml) added in a 180 μl aliquot such that the concentration of the large fragment was not limiting. To test the effect of crude *E. coli* lysate on the sensitivity of the reaction, in a separate experiment, samples were also spiked by addition of 20 μl of lysate from *E. coli* BL21 (DE3) expressing an irrelevant nontagged protein, before the addition of the GFP 1–10 OPT. Fluorescence kinetics (λ_{exc} = 488 nm /λ_{em} = 530 nm) were monitored with a FL600 Microplate Fluorescence Reader (Bio-Tek), at 3 min intervals for 15 h. The background fluorescence of a blank sample (20 μl of *E. coli* lysate expressing an irrelevant protein, 100 μl of 0.5 mg/ml GFP 1–10 OPT and 100 μl of 0.5% BSA in TNG buffer) was subtracted from final fluorescence values. The blank was less than 30% the signal from the lowest target concentration (0.1 pmol sulfite reductase–GFP 11). To assay soluble fractions of the eighteen *P. aerophilum* proteins for pET-expressed protein quantification tests and to perform assays on optima during directed evolution of the GPF 11 and GFP 1–10 variants, 20 μl of target protein soluble fractions were mixed with 180 μl of 0.35 mg/ml refolded GFP 1–10 OPT (~600 pmol) in a 96-well microplate (Nunc-Immuno plate, Nunc). To assay insoluble pellets, 50 μl of each resuspended insoluble fraction was centrifuged, the dried pellets were dissolved by addition of 50 μl of 9 M urea, and then 10 μl of the unfolded samples were assayed by rapid addition of 190 μl of 0.35 mg/ml GFP 1–10 OPT in TNG. The fluorescence values of the pellet assays were scaled by a factor of two to compensate for the lower volume relative to the soluble assays, allowing direct comparison with the soluble fraction assays. The final concentration of urea in the assay was ~0.4 M (see **Supplementary Fig. 4** online).

SDS-PAGE and solubility determinations. 15 μl of the soluble and pellet fractions were mixed with 15 μl of 2× SDS denaturing buffer containing 100 mM TRIS, 200 mM dithiothreitol, 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).

***In vivo* whole-cell plate complementation assays.** *E. coli* BL21 (DE3) cells coexpressing test proteins with an N-terminal 6-His and C-terminal GFP11 M3 tag from pTET-Spec^R plasmids (**Supplementary Fig. 1** online) and GFP 1–10 OPT from pET, were grown to saturation in LB containing 50 μg/ml kanamycin and 70 μg/ml spectinomycin and diluted in 20% glycerol at $OD_{600\ nm} = 1.0$ 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 cells were either induced sequentially or coinduced. For the sequential induction, cells on membranes were incubated for 1.5 h on a plate containing 250 ng/ml AnTet, 1 h on a resting plate and finally 1 h on 1 mM IPTG plate (note shorter induction times relative to those used for engineering GFP 11). For the coinduction protocol, the membranes were moved to plates containing both 600 ng/ml AnTET and 1 mM IPTG for 4 h at 37 °C to coexpress the GFP 11 fusions and the large GFP 1–10 fragment. The induced colonies on the plates were illuminated using an Illumatool Lighting System (LightTools Research) equipped with a 488 nm excitation filter and photographed with a DC290 digital camera (Kodak) through a colored glass filter (520 nm long pass; LightTools Research).

Talon resin binding of soluble GFP 11 M3 fusion proteins and analyses of soluble and pellet fractions by SDS-PAGE. To analyze soluble and pellet fractions of the same clones used for the *in vivo* whole-cell plate complementation assays, the clones were grown at 37 °C in a 1 ml 96-well culture plate. Cells were induced in the exponential phase with 250 ng/ml AnTET for 1 h, washed three times with LB and then induced with 1 mM IPTG for 1.5 h. After induction, the culture pellets were resuspended with 110 μl of TNG buffer and sonicated. The lysate was fractionated by centrifugation to yield the soluble and the pellet fractions. Forty microliters of the soluble extract of sequentially induced liquid cultures were mixed with an equal volume of a 50% v/v slurry of Talon metal affinity resin (Clontech) in TNG buffer and centrifuged briefly. The unbound fraction was removed by pipetting, and the beads were washed successively two times with an excess of TNG buffer. After the last centrifugation step, the buffer was discarded and 40 μl of 2×SDS denaturing buffer were added and heated for 15 min at 100 °C. The insoluble fraction was denatured as described. The Talon-bound and denatured samples were each resolved on a 4–20% gradient Criterion SDS-PAGE gel (Bio-Rad). The protein samples were stained using Gel Code Blue stain reagent (Pierce) and imaged using a GS-800 Calibrated Densitometer (Biorad).

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

ACKNOWLEDGMENTS

The authors wish to acknowledge B. Mark and J.-D. Pédelacq for helpful comments and the National Institutes of Health and Laboratory-directed research and development program for generous support.

COMPETING INTERESTS STATEMENT

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

Received 6 August; accepted 22 October 2004
Published online at http://www.nature.com/naturebiotechnology/

1. Ullmann, A., Jacob, F. & Monod, J. Characterization by *in vitro* complementation of a peptide corresponding to an operator-proximal segment of the beta-galactosidase structural gene of *Escherichia coli*. *J. Mol. Biol.* **24**, 339–343 (1967).
2. Wehrman, T., Kleaveland, B., Her, J.H., Balint, R.F. & Blau, H.M. Protein-protein interactions monitored in mammalian cells via complementation of beta-lactamase enzyme fragments. *Proc. Natl. Acad. Sci. USA* **99**, 3469–3474 (2002).
3. Nixon, A.E. & Benkovic, S.J. Improvement in the efficiency of formyl transfer of a GAR transformylase hybrid enzyme. *Protein Eng.* **13**, 323–327 (2000).
4. Wigley, W.C., Stidham, R.D., Smith, N.M., Hunt, J.F. & Thomas, P.J. Protein solubility and folding monitored *in vivo* by structural complementation of a genetic marker protein. *Nat. Biotechnol.* **19**, 131–136 (2001).
5. Kim, J.S. & Raines, R.T. Ribonuclease S-peptide as a carrier in fusion proteins. *Protein Sci.* **2**, 348–356 (1993).
6. Kelemen, B.R. *et al.* Hypersensitive substrate for ribonucleases. *Nucleic Acids Res.* **27**, 3696–3701 (1999).
7. Richards, F.M. & Vithayathil, P.J. The preparation of subtilisn-modified ribonuclease and the separation of the peptide and protein components. *J. Biol. Chem.* **234**, 1459–1465 (1959).
8. 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).
9. Bertens, P., Heijne, W., Van Der Wel, N., Wellink, J. & Van Kammen, A. Studies on the C-terminus of the Cowpea mosaic virus movement protein. *Arch. Virol.* **148**, 265–279 (2003).
10. 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).
11. Ghosh, I., Hamilton, A.D. & Regan, L. Antiparallel leucine zipper-directed protein reassembly: Application to the green fluorescent protein. *J. Am. Chem. Soc.* **122**, 5658–5659 (2000).
12. Hu, C.D. & Kerppola, T.K. Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat. Biotechnol.* **21**, 539–545 (2003).
13. Ozawa, T. & Umezawa, Y. Detection of protein-protein interactions *in vivo* based on protein splicing. *Curr. Opin. Chem. Biol.* **5**, 578–583 (2001).
14. Ozawa, T., Takeuchi, T.M., Kaihara, A., Sato, M. & Umezawa, Y. Protein splicing-based reconstitution of split green fluorescent protein for monitoring protein-protein interactions in bacteria: improved sensitivity and reduced screening time. *Anal. Chem.* **73**, 5866–5874 (2001).
15. Crameri, 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).
16. 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).
17. Fitz-Gibbon, S. *et al.* A fosmid-based genomic map and identification of 474 genes of the hyperthermophilic archaeon Pyrobaculum aerophilum. *Extremophiles* **1**, 36–51 (1997).
18. 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).
19. Tsien, R.Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998).
20. Tal, M., Silberstein, A. & Nusser, E. Why does Coomassie Brilliant Blue R interact differently with different proteins? A partial answer. *J. Biol. Chem.* **260**, 9976–9980 (1985).
21. Studier, F.W., Rosenberg, A.H., Dunn, J.J. & Dubendorff, J.W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89 (1990).
22. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1–I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
23. Baneyx, F. Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* **10**, 411–421 (1999).
24. Fahnert, B., Lilie, H. & Neubauer, P. Inclusion bodies: formation and utilisation. *Adv. Biochem. Eng. Biotechnol.* **89**, 93–142 (2004).
25. Gerstein, M., Edwards, A., Arrowsmith, C.H. & Montelione, G.T. Structural genomics: current progress. *Science* **299**, 1663 (2003).
26. Makrides, S.C. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* **60**, 512–538 (1996).
27. Yokoyama, S. Protein expression systems for structural genomics and proteomics. *Curr. Opin. Chem. Biol.* **7**, 39–43 (2003).
28. Fox, J.D., Kapust, R.B. & Waugh, D.S. Single amino acid substitutions on the surface of *Escherichia coli* maltose-binding protein can have a profound impact on the solubility of fusion proteins. *Protein Sci.* **10**, 622–630 (2001).
29. Pelletier, J.N., Campbell-Valois, F.X. & Michnick, S.W. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc. Natl. Acad. Sci. USA* **95**, 12141–12146 (1998).
30. Pelletier, J.N., Arndt, K.M., Pluckthun, A. & Michnick, S.W. An *in vivo* library-versus-library selection of optimized protein-protein interactions. *Nat. Biotechnol.* **17**, 683–690 (1999).
31. Goh, C.S. *et al.* Mining the structural genomics pipeline: identification of protein properties that affect high-throughput experimental analysis. *J. Mol. Biol.* **336**, 115–130 (2004).
32. Terwilliger, T.C. Structures and technology for biologists. *Nat. Struct. Mol. Biol.* **11**, 296–297 (2004).
33. Armstrong, N., de Lencastre, A. & Gouaux, E. A new protein folding screen: application to the ligand binding domains of a glutamate and kainate receptor and to lysozyme and carbonic anhydrase. *Protein Sci.* **8**, 1475–1483 (1999).
34. Waldo, G.S. Genetic screens and directed evolution for protein solubility. *Curr. Opin. Chem. Biol.* **7**, 33–38 (2003).
35. Knaust, R.K. & Nordlund, P. Screening for soluble expression of recombinant proteins in a 96-well format. *Anal. Biochem.* **297**, 79–85 (2001).
36. Ozawa, T., Sako, Y., Sato, M., Kitamura, T. & Umezawa, Y. A genetic approach to identifying mitochondrial proteins. *Nat. Biotechnol.* **21**, 287–293 (2003).
37. Zhang, S., Ma, C. & Chalfie, M. Combinatorial marking of cells and organelles with reconstituted fluorescent proteins. *Cell* **119**, 137–144 (2004).
38. Morris, M.C., Depollier, J., Mery, J., Heitz, F. & Divita, G. A peptide carrier for the delivery of biologically active proteins into mammalian cells. *Nat. Biotechnol.* **19**, 1173–1176 (2001).
39. 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).