Published in final edited form as:

Proteins. 2015 July ; 83(7): 1225–1237. doi:10.1002/prot.24699.

TGP, an extremely stable, non-aggregating fluorescent protein created by structure-guided surface engineering

Devin W. Close^{1,4}, Craig Don Paul², Patricia S. Langan¹, Matthew C.J. Wilce², Daouda A.K. Traore², Randal Halfmann³, Reginaldo C. Rocha⁴, Geoffery S. Waldo¹, Riley J. Payne⁵, Joseph B. Rucker⁵, Mark Prescott², and Andrew R.M. Bradbury^{1,*}

¹Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM, USA

²Department of Biochemistry and Molecular Biology, Monash University, Clayton Campus, Melbourne, Victoria, Australia

³Department of Biochemistry, UT Southwestern Medical Center, Dallas, TX, USA

⁴Center for Integrated Nanotechnologies, Los Alamos National Laboratory, Los Alamos, NM, USA

⁵Integral Molecular, Philadelphia PA, USA

Abstract

In this paper we describe the engineering and X-ray crystal structure of Thermal Green Protein (TGP), an extremely stable, highly soluble, non-aggregating green fluorescent protein. TGP is a soluble variant of the fluorescent protein eCGP123, which despite being highly stable, has proven to be aggregation-prone. The X-ray crystal structure of eCGP123, also determined within the context of this paper, was used to carry out rational surface engineering to improve its solubility, leading to TGP. The approach involved simultaneously eliminating crystal lattice contacts while increasing the overall negative charge of the protein. Despite intentional disruption of lattice contacts and introduction of high entropy glutamate side chains, TGP crystallized readily in a number of different conditions and the X-ray crystal structure of TGP was determined to 1.9 Å resolution. The structural reasons for the enhanced stability of TGP and eCGP123 are discussed. We demonstrate the utility of using TGP as a fusion partner in various assays and significantly, in amyloid assays in which the standard fluorescent protein, EGFP, is undesirable because of aberrant oligomerization.

Keywords

protein engineering; GFP structure; X-ray crystallography; thermal green protein; thermal stability; eCGP123; structure-guided mutagenesis

*Corresponding author: amb@lanl.gov; phone: +1 505 665 0281; fax: +1 505 665 3024; address: Los Alamos National Laboratory, Bioscience Division, HRL-1 TA-43, Los Alamos, NM 87545.

Introduction

Fluorescent proteins (FPs) are used in nearly every facet of biological research. The appeal of FPs comes in large part from the fact that fluorescence emission arises entirely from protein sequences encoded by DNA, without the requirement of exogenous ligands or co-factors other than molecular oxygen^{1–5}. This property enables the use of FPs as fusion tags for *in vivo* visualization in organisms and cell culture⁶. Since chromophore formation and fluorescence is dictated by the protein fold, which is encoded by the amino acid composition, even slight variations in the sequence can lead to dramatic changes in photophysical and biophysical characteristics⁵. Discovery of naturally occurring FP variants in parallel with optimization through random or rational mutagenesis, has led to an enormous number of FP-based proteins including a palette of colors⁷, photochromic versions⁸, photo switchable FPs^{9,10}, split FP based molecular reporters¹¹, inherently fluorescent biosensors¹², and highly stable variants^{13,14}.

An important characteristic of FPs is that fluorescence is directly associated with the properly folded protein². The fact that fluorescence can be used as a direct measure of the folded state of the protein allows rapid and facile screening in protocols that can then be easily adapted to non-fluorescent proteins. This has led to their use as reporters of *in vivo* or *in vitro* folding^{15,16} and has also facilitated development of methodologies for stabilization of non-fluorescent proteins^{13,14}. Stabilization and folding robustness (improved folding kinetics in a variety of hosts and conditions) are important as many proteins are increasingly being used in technological and cell biology applications that require high degrees of thermal and chemical robustness¹⁷. Compared to synthetic fluorophores (such as nanodots, quantum dots, and dyes), commonly available FPs are relatively unstable in conditions outside the typical physiological range, limiting their use in many applications. For this reason, more robust variants are desired to enable fluorescence emission in denaturing applications such as lysosome fusions¹⁸, amyloid fusions¹⁹, and in assays involving thermophilic organisms^{17,20}. Improving the stability and folding robustness of FPs not only allows for use in nominally denaturing conditions, but importantly, it has also been shown that improved stability leads to higher tolerance to random mutations and insertions^{12,14}, enabling further development of FPs with unique photophysical properties. A subset of FPs developed to have these qualities include superfolder GFP (sfGFP;¹⁴), dsRed, zFP506, mRFP1, DsRed^{21–23}, and the extremely stable FP eCGP123¹³.

The FP eCGP123 was engineered using directed evolution from a rationally designed and relatively unstable FP, consensus green protein (CGP)²⁴. Evolution of eCGP123 involved a recursive process whereby a destabilizing amino acid sequence (based on an antibody binding loop) was inserted at pre-defined positions between strands in the β -barrel leading to loss of fluorescence, followed by mutagenesis of the rest of the scaffold to recover fluorescence¹³. As beneficial mutations accumulated through DNA shuffling, the stringency (number of loops inserted simultaneously) was increased until a pool of FP-encoding genes containing three loops and accumulated stabilizing mutations was generated. When the destabilizing loops were excluded and consensus mutations were combined and included in a synthesized eCGP123 gene, the resulting protein exhibited exceptional thermal and chemical stability¹³.

In this work, we describe the X-ray crystal structure of eCGP123 and report the use of structure-guided engineering to generate a new protein, Thermal Green fluorescent Protein (TGP), with improved solubility and stability compared to eCGP123. We report the structure of TGP and discuss structural features of eCGP123 and TGP that help explain their observed stability. Finally, we demonstrate and discuss the use of TGP in biochemical assays including an amyloid assay not amenable to conventional FPs.

Materials and Methods

Cloning, Protein expression, and purification

DNA expression cassettes encoding each of the different FPs were cloned into the pETCK3 expression plasmid¹³. Proteins were expressed in *Escherichia coli* (*E. coli*) BL21 DE3 cells by culturing in autoinduction media²⁵ at 37°C for 5 hours followed by a further incubation 20°C for 18–20 hours. Cells were harvested and stored as pellets at –80°C. Cells were lysed using an Avestin Emusiflex Cell Homogenizer (Avestin) in Lysis Buffer (30mM Hepes pH 7.3, 300mM NaCl, 15mM Imidazole) and protein purified by affinity chromatography using Ni-NTA Agarose (Qiagen). Proteins were eluted from the Ni-NTA resin in Lysis Buffer + 300mM Imidazole, concentrated, then subjected to chromatography on a 320ml XK 26/60 Sephadex 200 size-exclusion column using an AKTA prime liquid chromatography system (GE healthcare). For biophysical and photophysical characterization, purified proteins were buffer exchanged by overnight dialysis into TNGE buffer (50mM Tris pH 7.5, 150mM NaCl, 10% Glycerol, 2mM EDTA).

Thermal Stability measurements

A qualitative assessment of thermal stability was made by observing the fluorescence emission of proteins through a 520nm filter and illuminated with 488nm lamp after incubation of protein for 1 hour at 80°C then 15–20 minutes at 20°C. Thermal stability at high temperatures was assessed by incubation in an ABI 7500 RT-PCR machine using the FAM filter set. Proteins in TNGE buffer, were held at constant temperatures of either 85°C or 90°C for up to 17 hours and fluorescence emission intensity acquired every 5 minutes.

Protein crystallization and data collection

Crystallization and data collection for eCGP123 have been described previously²⁶. Indexing and scaling using Mosflm²⁷ of the data in a monoclinic (or higher) Laue group indicated the data was consistent with space group P1 and *POINTLESS*²⁸ was used to confirm the correct Laue and space group. The eCGP123 structure was solved using molecular replacement and mAG (PDB entry: 3adf;²⁹) as a search ensemble. Crystals of TGP were grown using the hanging drop vapor diffusion method where 3μl of purified protein (27 mg/ml in 20mM Hepes pH 7.4 100mM NaCl) was mixed with 3μl of well solution (0.1 M Tris-HCl pH 7.0 and 20% w/v Polyethylene glycol monomethyl ether 2000). Green-fluorescent TGP crystals grew at 20° C over a period of 24–48 hours. Crystals were selected, cryoprotected in a solution containing the well solution with 20% (v/v) glycerol and flash frozen in liquid nitrogen, Diffraction data was collected at 100K using an in-house system consisting of a Rigaku Micromax-007HF generator (Rigaku), R-axis IV++ detector (Rigaku), and Oxford Liquid Nitrogen cryostream (Oxford Cryosystems). Data were collected using HKL3000³⁰,

indexed and integrated using iMosflm²⁷, then scaled using Scala within CCP4i suite³¹. Molecular replacement (using eCGP123 as a search model) and refinement were performed using PHENIX³². Simulated annealing was performed during initial rounds of refinement to remove model bias. Coot³³ and Molprobity³⁴ were used for model building, optimization, and validation. PyMOL (www.pymol.org) and Inkscape (www.inkscape.org) were used for graphical representation and figure generation. Surface and protomer interface calculations were performed using PDBe PISA³⁵. Data and refined models for TGP and eCGP123 have been deposited to the Worldwide Protein Data Bank (wwPDB) with accession numbers 4TZA and 4TZG (respectively).

Spectroscopy

UV-visible spectra were recorded using a Cary 300 spectrophotometer (Agilent). Extinction coefficients were determined using protein solutions with absorbance values < 0.2 (at $I_{\max}^{\text{Abs}}$). Fluorescence emission and excitation spectra were measured using a Jobin-Yvon Fluoromax-4 spectrofluorometer (Horiba) with a 2 nm excitation/emission slit-width. Quantum yields were determined by the comparative gradient method using fluorescein (in 0.1 M NaOH) as a standard ($\Phi_F = 0.91$) and excitation wavelength (I_{exc}) at 470 nm. For fluorescence lifetime measurements, the spectrofluorometer was coupled to a time-correlated single photon counting system (Horiba Jobin-Yvon) equipped with a 455 nm NanoLED operating at 1 MHz. Each measurement was terminated when a maximum peak preset of 20,000 photon counts was reached for the fluorescence monitored at 505 nm. The instrument response function was determined using a light scattering solution. Analysis of fluorescence decay profiles was performed using Horiba DAS6 software.

Amyloid fusions and partitioning assay

Fusions of Sup35 NM and Mot3PrD to TGP or EGFP were constructed using Gateway cloning. The TGP sequence was swapped with EGFP in destination vector pAG426GAL-ccdB-EGFP³⁶ to create destination vector pAG426GAL-ccdB-TGP. The Sup35 NM and Mot3PrD sequences were then introduced into each of these vectors through LR reactions with NM and Mot3PrD entry clones³⁷. The resulting expression clones encoding C-terminally tagged prion proteins were then transformed into [*PIN*+] yeast strain BY4741, or its isogenic [*pin*-] counterpart (RHY0734). All yeast manipulations were performed using standard techniques³⁷. Semi-denaturing detergent agarose gel electrophoresis (SDD-AGE) was performed as described in reference³⁷. Following electrophoresis, fluorescence was visualized in-gel using a Typhoon 9400 imager with 488nm excitation and 526nm emission.

Virus like particles

Virus-like particles (VLPs) were produced by transfection of HEK-293T cells with the retroviral (murine leukemia virus [MLV]) Gag protein, as previously described^{38,39}. Three different Gag proteins were utilized for this study: wild-type Gag (Gag-only), fusion proteins containing either eGFP (Gag-GFP), or TGP (Gag-TGP). Fluorescent gag constructs were cloned as in-frame fusions to the gag ORF to eliminate the pol gene as previously described⁴⁰. Forty-eight hours after transfection, supernatants containing VLPs were harvested, passed through a 0.45 μm filter, centrifuged through a 20% sucrose cushion and

resuspended in 10 mM HEPES, pH 7.5. Cell surface labeling (carbohydrate) by Dylight488 or biotin was performed as directed by the manufacturer. Each batch of VLPs produced was assayed for homogeneity by dynamic light scattering using a Proterion DynaPro DLS and for total protein concentration by bicinchoninic acid (BCA) protein assay reagent (Thermo Scientific, Waltham, MA). VLP fluorescent was measured using a Flexstation II 384 fluorescence plate reader (Molecular Devices).

Results and Discussion

Crystal Structure of eCGP123

We determined the crystal structure of eCGP123 to 2.1 Å (Figure 1) with crystallization and initial diffraction reported previously²⁶. The structure was determined using molecular replacement (data collection and refinement statistics are listed in Table I) in the P1 space group with eight molecules in the asymmetric unit (Figure 2A). The eight eCGP123 protomers are essentially identical to each other, with an overall (mean) RMSD of 0.25 Å for 217 Cα residues (calculated using PDBeFold⁴¹).

The overall fold of eCGP123 is an eleven-strand β-barrel consistent with other green FPs⁴² (Figure 1A). Within the barrel, a coaxial central helix is split by three amino acids (Gln62, Tyr63, and Gly64) that cyclize, dehydrate, and oxidize to form the mature QYG chromophore⁴. Electron density is well defined for the chromophore (Figure 1B) for all protomers including density for covalent bonding between the chromophore and residues Phe61 and Asn65. The chromophore is coordinated through extensive contacts with the interior of the β-barrel (detailed in Supplementary Table 1 and shown in Figure 1C). Critical residues for proper chromophore coordination and fluorescence include Arg66, Ser142, and His193, and are invariant with respect to similar proteins such as monomeric Azami Green (mAG)²⁹ and Dronpa^{10,43}. Maintaining planarity and rigidity of the 4-hydroxylbenzyl ring (Tyr63) relative to the imidazoline ring (Gln62, Gly64) of the chromophore is important for fluorescence⁴⁴, and its position is coordinated by a network of residues that line the chromophore cavity (Supplementary Table 1). As is the case for mAG and Dronpa, a central component of this network is His193, the imidazole ring of which is positioned to form a π-π stacking interaction with the 4-hydroxylbenzyl ring of the chromophore. His193 is held in a specific orientation through a hydrogen bond network that includes Glu144, Glu211, Arg66, Tyr177, and several well-ordered waters (Figure 1C). Outside of this network, Ser142 stabilizes the chromophore through hydrogen bonds to the 4-hydroxylbenzyl OH group. The majority of differences between eCGP123 and related proteins such as mAG lie outside the chromophore cavity, the implications of which for the stability of eCGP123 and TGP are discussed in more detail below.

Structure-guided surface engineering for improved solubility

Although eCGP123 is extremely thermostable, we observed the protein to be aggregation prone, and particularly so in phosphate buffered saline (PBS), even at concentrations as low as 1–2 mg/ml. This property may limit the use of eCGP123, especially for fusion to aggregation-prone proteins and in applications that require high protein concentrations and broad solubility profiles, such as the amyloid assays described below. We used the crystal

structure of eCGP123 to guide surface engineering for improved solubility using an approach that followed two guiding principles. First, we hypothesized that intermolecular surfaces observed in the crystal lattice correspond to those that lead to association and aggregation of the protein in solution. We surmised that disruption of these interfaces would result in a more soluble protein. In this respect, structure-guided surface engineering has successfully been used to reduce multimerization of other FPs such as dsRED^{45–48} and improve solubility profiles of other (non-fluorescent) proteins to facilitate crystallization⁴⁹. Second, we hypothesized that lack of solubility may be due at least in part, to the net charge, or calculated isoelectric point (pI), of eCGP123 being zero at neutral pH (7.0), a characteristic of aggregation-prone proteins^{50,51}. Further, recent reports of successes in “supercharging” proteins^{52,53}, suggest that increasing the net surface charge can lead to improved solubility at high concentrations. To combine these two principles, we decided to introduce selected amino acid substitutions that would both alter the net surface charge *and destabilize apparent interfaces observed in the crystal lattice*. Following this stratagem and starting with eCGP123, we generated Thermal Green Protein (TGP).

To generate TGP, lysine, arginine, or asparagine (K, R, or N) residues making extensive contacts through hydrogen bonds, salt bridges, or amino-aromatic interactions at intermolecular interfaces observed in the eCGP123 structure, were substituted with Glutamate (E) (Figure 2D). Substitution of positively charged side chains on the exposed surface of the β -barrel for negative ones (K45E, K73E, K117E, R149E, and N158E) enabled a higher net change in overall charge per mutation, with a greater likelihood of charge repulsion. Studies have shown that substitution of N, Q, or K with acidic residues can lead to the greatest net gain in solubility when compared to other substitutions^{49,54}. We also noted the presence of substantial interface forming contacts between the C-terminal tail residues of certain protomers and the barrel of others (Figure 2B). Due to the significant packing observed, and based on successful C-terminal tail engineering for other FPs⁴⁶, we mutated the C-terminal residues 219–225 from MLPSQAK to GGGSGGG, a common linker sequence with minimal interaction-promoting side chains. The combined substitutions resulted in a calculated charge of -10 (at neutral pH) for TGP.

Comparing solubility, thermal stability, and photophysical characteristics

A DNA expression cassette encoding TGP was synthesized, expressed in, and purified from *Escherichia coli* (see Materials and Methods) in parallel with eCGP123. Improvements in protein solubility were immediately apparent. TGP and eCGP123 protein purified by Ni-NTA chromatography was adjusted to a similar protein concentration (~ 1.3 mg/ml) and dialyzed against PBS. After overnight equilibration, eCGP123 was clearly aggregated while TGP showed no signs of visible aggregation (Figure 3A–B). It was necessary to increase the concentration of TGP to >50 mg/ml before any aggregation became visible. These observations demonstrate that our structure guided surface engineering approach resulted in a protein with greatly improved solubility.

We previously showed¹³ that eCGP123 has markedly improved thermal stability relative to mAG, despite an overall sequence identity to mAG of $\sim 88\%$ (Figure 2D). To assess whether improvements in TGP solubility may have arisen at the expense of thermal stability, we

examined the ability of TGP to maintain fluorescence with continued exposure to high temperatures. We performed a quantitative assessment of TGP, eCGP123, and mAG thermal stability by measuring fluorescence emission as a function of time while incubating at high temperature (85°C and 90°C). TGP appeared to have slightly improved thermal stability compared to eCGP123 (Figure 3C). The improvement in thermal stability was most marked when the proteins were held at 90°C for an extended period of time. In this case, mAG lost fluorescence immediately, whereas the time taken to reduce fluorescence by half was ~175 minutes for eCGP123 and ~380 minutes for TGP. This represents a ~2.2-fold improvement in apparent thermal stability for TGP compared to eCGP123.

We determined the absorption/emission spectra, quantum yield (Φ_F), and fluorescence lifetimes (τ) for TGP, eCGP123, and mAG (Table II and Figure 3D–E). The spectra are similar for all three proteins. The absorption maxima ($\lambda_{\max}^{\text{Abs}}$) for TGP and eCGP123 (both at 493 nm) are marginally red-shifted by 2 nm relative to mAG (491 nm), and the emission maximum for TGP (507 nm) shows a slight red-shift relative to mAG and eCGP123 (both at 504 nm). Quantum yield (Φ_F) decreases by a small margin for TGP and eCGP123 (0.66, 0.70) and both are somewhat lower than mAG (0.76), a result that is in agreement with other reports (¹³ and [Don Paul, 2014, in preparation]). The fluorescence lifetimes (τ) are essentially identical for the three proteins (3.0–3.2 ns). Collectively, these data demonstrate that a moderate degree of structure guided surface engineering resulted in a protein with improved solubility and thermal stability relative to eCGP123, without compromising key photophysical characteristics.

The X-ray crystal structure of TGP

In addition to understanding the mechanisms underlying its extreme stability, we believed the structure of TGP would also be interesting to study the effects of the solubilizing substitutions on crystallization. Our solubilizing strategy targeted crystal lattice interfaces of eCGP123 through substitution with glutamate, an amino acid with a low pKa, a high degree of conformational entropy, whose enrichment is typically negatively correlated with crystallization propensity^{55,56}. Interestingly, TGP crystallized readily under a range of different conditions and in different crystal forms (Supplemental Figure 1). We collected x-ray diffraction data on several crystals (not shown) and solved the structure of the protein using one of these data sets to a resolution of 1.9 Å in the space group $P2_12_12_1$ (statistics shown in Table I).

As expected, the TGP fold and chromophore interactions, are essentially identical to that of eCGP123 (0.37 Å over 211 C residues⁴¹ and Supplemental Table 1), although the major interfaces seen between eCGP123 protomers are absent in the TGP structure, with substituted residues making minimal contacts. For example, the extensive interface (~1462 Å² buried surface area (BSA)) between the C and E protomers (Figure 2B) of eCGP123 is not observed in the TGP structure, suggesting that the R149E and N158E mutations elicited the intended interface disrupting effects. In addition, the six C-terminal mutations (219–225) no longer make contacts with adjacent molecules and are disordered, as evidenced by lack of electron density for any of the four TGP protomers beyond residue 219. While at least four interfaces in the eCGP123 have BSA >1400 Å², the largest interface between protomers in

the TGP structure is $\sim 892 \text{ \AA}^2$. Although the differences in crystallization and diffraction could be due to specific conditions and crystal-to-crystal variability, we did not observe any crystals with space groups, unit cell dimensions, crystal forms or intermolecular interfaces comparable to the structure of eCGP123. This implies that surface engineering based on disruption of lattice contacts and addition of high entropy side chains does not necessarily abrogate crystal formation, but can lead to alternate packing which in this case, resulted in a high propensity for crystallization under many conditions, higher symmetry lattice formation, and improved resolution.

Interestingly, this appears contrary to proven strategies to promote crystallization, such as surface entropy reduction (SER), that work by replacing high entropy side chains (typically glutamate) with low entropy ones such as alanine^{56–58}. Moreover, larger-scale protein crystallization studies suggest that an increase in the number of surface exposed glutamates leads to a decrease in the likelihood of crystallization⁵⁵. That said, striking the right balance appears to be crucial, as other studies indicate that improving protein solubility by surface engineering can result in improved crystallization and diffraction (see⁴⁹ for a review), although these do not use lattice packing to inform mutagenesis, as is the case in our study. In the case of TGP, targeted disruption of crystal lattice contacts by reversing charge (from positive to negative) through the addition of glutamates, led to crystallization under a number of different conditions (examples shown in Supplementary Figure 1), higher symmetry, and improved resolution. This suggests that in certain instances, particularly in the case of proteins with poor solubility, an increase in solubility at the cost of increased surface entropy can be beneficial, most likely due to the ability to perform crystal trials at much higher protein concentrations. Although additional examples are required before the generality of the strategy can be claimed, a similar model of disruption of lattice contacts could be employed in specific cases where improved resolution, higher symmetry, or crystallization in a broader range of conditions is desired, particularly for aggregation-prone proteins.

Structural clues for stability

The crystal structures of TGP and eCGP123 were used to gain insight into how the stabilizing substitutions confer higher stability when compared to similar, but far less stable green FPs such as mAG⁴⁸ and CGP²⁴. Although TGP and eCGP123 differ from mAG at more than 25 positions (Figure 2D), only seven of these positions are also altered in CGP. These seven mutations (K30I, A53S, T59P, V60A, T82A, K190E, and K208R; Figure 4) are therefore the key contributors to thermostabilization and are the primary focus of our structural analyses.

The stabilizing mutations occur throughout the eCGP123 and TGP proteins (Figure 4A) and are likely to contribute to stability through a variety of mechanisms. Three of the mutations (A53S, T59P, and V60A) are buried and are constituents of the coaxial α -helix N-terminal to the chromophore. In the case of A53S, the Ser53 O γ expands a hydrogen bond network consisting of Thr136, Asp55 and well-ordered solvent (Figure 4D). The expanded hydrogen bond network likely stabilizes the central helix and/or the packing interaction of the helix with residues 133–137, mediated through Thr136 (Figure 4D). T59P and V60A (Figure 4B)

were some of the earliest mutations to arise in the directed evolution of eCGP123 from CGP¹³. Both substitutions may have a stabilizing effect primarily due to favorable entropic contributions and stabilization of the central helix^{59,60}. The central helix is distorted with 3_{10} character, so it may be that Pro and Ala at these positions are energetically favored due to the tighter dihedral angle observed in 3_{10} helices. Solvent is reordered due to T59P/V60A and may contribute to increased stability (Figure 4B). In mAG, a well-ordered water is coordinated by the carbonyl oxygen of Asp55 and the O γ of Thr59. The substitution T59P results in steric repulsion of this water by Pro59 that could lead to improved van der Waals interaction between Pro59 and Trp139. Interestingly, the T59P/V60A mutation results in a larger cavity and new ordered water molecules observed in a hydrogen bonding network between Arg91 and the carbonyl oxygen of Ile56 (Figure 4B). The overall rearrangement of solvent in this region results in a larger hydrogen bonding network accompanied by more efficient packing within the interior of the protein^{61,62}.

Both K190E and K208R are surface exposed substitutions that may have a stabilizing effect through inter-strand hydrogen bonding, a feature that has been reported previously¹⁴. This is especially obvious for K208R, which forms a side chain salt bridge with Asp41 (Figure 4C). Other stabilizing substitutions are T82A and K30I, the effects of which are not entirely clear. The polar side chain of Thr82 is located at the end of a short helix with its side chain buried adjacent to the bulky aromatics sidechains of Phe83 and Tyr87. Loss of the polar side chain at this position in the T82A may improve packing together with more favorable entropy due to a truncated sidechain and helical preference for Ala over Thr⁶³. The reasons for thermostabilization conferred by K30I are not clear, but the mutation could be favoring β -strand structure and disruption of potential charge-charge repulsion with the side-chain of an adjacent Lys9 residue. Alternatively, T82A and K30I might simply eliminate off-pathway folding intermediates, as suggested for superfolder GFP Y145F and I171V¹⁴.

TGP fusions in amyloid assays and lipoparticles

When fused to other proteins of interest, FPs that aggregate or form oligomers can interfere with the biology or biochemistry of the process under investigation. Indeed, it was recently demonstrated that most FPs stabilize even weak homo-oligomers, resulting in aberrant clustering and mislocalization of their fusion partners⁶⁴. Prion-forming proteins are one class of proteins with a tendency to form oligomers *in vivo*, a property important for their biological activity. Prion-forming proteins are characterized by the ability to switch between soluble and aggregated forms. Aberrant oligomerization driven by FP fusion partners may alter the frequency of aggregate formation and preclude accurate assessments of prion behavior. Chimeric versions of the prion-forming region (NM) from the yeast prion protein, Sup35, were constructed in fusion with either TGP or the widely used FP, EGFP. Yeast cells expressing either fusion were lysed and the lysates separated on an SDS-containing agarose gel (SDD-AGE;⁶⁵) as shown in Figure 5. Direct visualization of fluorescence revealed that in addition to monomeric and polymerized species, NM-EGFP also formed a discrete and highly fluorescent SDS-stable oligomer. The oligomer was even more prevalent when NM-EGFP was expressed in the absence of prion templates (Figure 5A), suggesting that it may compete with amyloid polymerization. This species is an artifact of fusion to EGFP, since over-expressed full length Sup35 does not form discrete SDS-resistant oligomers⁶⁶. Both the

size and the extraordinary fluorescent intensity of the oligomer interfere with fluorescence quantification of the monomeric and amyloid species and therefore limit the utility of SDD-AGE. In contrast, NM-TGP populated only monomeric and amyloid species, and not the oligomer (Figure 5B). We performed similar experiments with the prion-forming region (PrD) of a second yeast prion protein, Mot3. As for NM, Mot3PrD formed a highly fluorescent, SDS-stable oligomeric species when fused to EGFP, but not when fused to TGP (Figure 5C). Immunoblotting confirmed that the two fusion proteins were expressed to comparable levels.

The tendency of EGFP to form oligomers when fused to NM or Mot3PrD may result from protein folding interference (especially since the fusion partners are appended to the N-terminus of EGFP and will therefore be translated first), although the fact the oligomers are highly fluorescent would argue against this possibility. Interestingly, the “monomerizing” mutant of GFP, A206K, fails to prevent the formation of discrete SDS-resistant oligomers when fused to NM (Viknesh Sivanathan and Ann Hochschild, personal communication) or Mot3PrD (Halfmann, unpublished data), suggesting that FP monomer species that interact very weakly (such as A206K) may nevertheless promote cooperative aggregation when fused to interacting species. The lack of oligomerization observed by TGP could be due to its increased, or more distributed charge (charge of EGFP at pH 7.0 = -5.9), reflecting its lack of aggregation, even at very high concentrations. The apparent resistance to oligomer formation in this assay suggests that TGP fusions may also be very useful in other assays in which aggregation is problematic, such as super-resolution microscopy techniques studying receptor aggregation following stimulation.

In order to assess the utility of TGP in other experiments in which FP fusions are used, we tested its ability to act as a Gag fusion partner for retroviral virus-like particle (VLP) formation. Retrovirus particles contain 1500–2000 copies of a structural Gag protein which self-assemble and bud from the plasma membrane of producer cells⁶⁷. While Gag-GFP fusions are known to properly bud from cells, non-aggregating fusion proteins such as TGP may be better fusion partners for retroviral Gag or other viral structural proteins. To test this, TGP was fused to the C-terminus of the Gag protein (Gag-TGP) of murine leukemia virus (MLV) and used to produce VLPs in HEK-293T cells. VLPs containing a fusion with EGFP (Gag-GFP) or Gag alone (Null) were also produced in parallel. All three constructs behaved similarly for production of VLPs suggesting that the differences in aggregation between TGP and EGFP were not sufficient to alter the self-assembly of MLV Gag. The fluorescence intensities of Gag-TGP particles and Gag-GFP particles were also equivalent and mirrored the similar fluorescence properties of the two proteins. As an additional control, Gag-TGP particles were compared with VLPs fluorescently labeled via their extracellular carbohydrates. As expected, Gag-TGP VLPs, with 1500 to 200 copies of TGP per particle were more fluorescent than cell surface labeled VLPs. Thus TGP is equivalent to EGFP as a fusion partner for viral structural proteins and may have advantages for assays that are sensitive to non-specific interactions, e.g. subcellular distribution⁶⁸.

Conclusion

In this work, we have determined the crystal structure of the extremely thermostable FP eCGP123 and used the information to carry out structure-guided surface engineering, leading to the improved protein, TGP. This protein has increased solubility while maintaining important photophysical characteristics and thermal stability. Improved solubility and stability are important characteristics for many proteins with applications in biotechnological and industrial processes, in large part because stable proteins can work at increased temperatures, can be active for longer, can be recycled, and in turn, lower cost of goods⁶⁹. In addition, it has been shown that stable or ‘superfolder’ proteins, are also more tolerant to mutagenesis compared to their less stable counterparts, enabling further engineering outcomes that, in some cases, may not have been otherwise possible^{11,14}. In this regard, both eCGP123 and TGP have been used to generate a range of photoswitching and photochromic proteins⁸, [Langan, et al 2014, in preparation]. High solubility and efficient folding are desirable properties when the FP is used as a fusion partner in order to prevent non-specific interactions and aggregation, that may create artifacts of subcellular distribution¹⁴. This is particularly important for some applications, perhaps best illustrated by the SDD-AGE amyloid assay described here, in which the “monomeric” A206K variant of EGFP still induces oligomerization, which is overcome by use of TGP. We have also used TGP as a fluorescent transcription factor fusion tag in a protein-binding microarray in which TGP’s lack of dimerization, intrinsic fluorescence and overall negative charge (preventing non-specific DNA interactions), were significant advantages over the standard fusion partner, GST⁷⁰. In addition to the advantages provided by its increased stability, solubility and monomeric status, we also show that TGP can be used as a standard FP fusion partner as a TGP-Gag fusion in the formation of fluorescent virus-like lipoparticles (VLPs).

The crystal structures of eCGP123 and TGP suggest that stabilizing mutations are likely to contribute to stability by different mechanisms. Some appear to increase stability by an easily observable mechanism (A53S leads to extended, internal hydrogen bonding network (Figure 4D)), while for others, the mechanism is not clear, such as T82A and K30I. Future studies will focus on the individual mutations to assess to what extent each contributes to stability. Surprisingly, the disruption of intermolecular interfaces in the eCGP123 structure led not only to improved solubility, but improved crystallization outcomes including crystallization in a wider variety of conditions (data not shown), higher symmetry and higher resolution diffraction. This strategy may be an alternative to approaches such as SER, particularly where the protein of interest has poor solubility, where improvement of low resolution (e.g. >3.0 Å) structures are desired, or in situations where higher symmetry and fewer molecules in the asymmetric unit are advantageous, such as in neutron crystallography⁷¹.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Moniquetta Hall for help with VLP production and Chidananda Sulli for cloning the Gag-fusion constructs. Funding was provided by DTRA (CBS.MEDBIO.04.10.LA.008) to ARMB, LANL LDRD-ER (20120449ER) to GSW, NIH Director's Early Independence Award (DP5-OD009152) to RH, and NIH Program Project Grant (GM063210; PI TC Terwilliger) to ARMB and GSW.

References

1. Chudakov DM, Matz MV, Lukyanov S, Lukyanov KA. Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues. *Physiol Rev.* 2010; 90(3):1103–1163. [PubMed: 20664080]
2. Tsien RY. The green fluorescent protein. *Annual Review of Biochemistry.* 1998; 67:509–544.
3. Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ. Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene.* 1992; 111(2):229–233. [PubMed: 1347277]
4. Miyawaki A, Nagai T, Mizuno H. Mechanisms of protein fluorophore formation and engineering. *Current Opinion in Chemical Biology.* 2003; 7(5):557–562. [PubMed: 14580558]
5. Pakhomov AA, Martynov VI. GFP Family: Structural Insights into Spectral Tuning. *Chemistry & Biology.* 2008; 15(8):755–764. [PubMed: 18721746]
6. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green fluorescent protein as a marker for gene expression. *Science.* 1994; 263(5148):802–805. [PubMed: 8303295]
7. Rizzo MA, Davidson MW, Piston DW. Fluorescent Protein Tracking and Detection: Fluorescent Protein Structure and Color Variants. *Cold Spring Harb Protoc.* 2009; 2009(12)
8. Don Paul C, Kiss C, Traore DAK, Gong L, Wilce MCJ, Devenish RJ, Bradbury A, Prescott M, Phanta: A Non-Fluorescent Photochromic Acceptor for pcFRET. *PLoS ONE.* 2013; 8(9)
9. Chang H, Zhang M, Ji W, Chen J, Zhang Y, Liu B, Lu J, Zhang J, Xu P, Xu T. A Unique Series of Reversibly Switchable Fluorescent Proteins with Beneficial Properties for Various Applications. *Proceedings of the National Academy of Sciences.* 2012; 109(12):4455–4460.
10. Wilmann PG, Turcic K, Battad JM, Wilce MCJ, Devenish RJ, Prescott M, Rossjohn J. The 1.7 Å Crystal Structure of Dronpa: A Photoswitchable Green Fluorescent Protein. *Journal of Molecular Biology.* 2006; 364(2):213–224. [PubMed: 17010376]
11. Cabantous S, phanie, Terwilliger TC, Waldo GS. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nature Biotechnology.* 2004; 23(1):102–107.
12. Pavoov TV, Cho YK, Shusta EV. Development of GFP-based biosensors possessing the binding properties of antibodies. *Proceedings of the National Academy of Sciences.* 2009; 106(29):11895–11900.
13. Kiss C, Temirov J, Chasteen L, Waldo GS, Bradbury ARM. Directed evolution of an extremely stable fluorescent protein. *Protein Engineering Design and Selection.* 2009; 22(5):313–323.
14. Pédelacq J-D, Cabantous S, Tran T, Terwilliger TC, Waldo GS. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology.* 2006; 24(1):79–88.
15. Pedelacq J-D, Nguyen HB, Cabantous S, Mark BL, Listwan P, Bell C, Friedland N, Lockard M, Faille A, Mourey L, Terwilliger TC, Waldo GS. Experimental mapping of soluble protein domains using a hierarchical approach. *Nucleic Acids Research.* 2011; 39(18):e125–e125. [PubMed: 21771856]
16. Cabantous S, Waldo GS. In vivo and in vitro protein solubility assays using split GFP. *Nat Meth.* 2006; 3(10):845–854.
17. Cava F, De Pedro MA, Blas-Galindo E, Waldo GS, Westblade LF, Berenguer J. Expression and use of superfolder green fluorescent protein at high temperatures in vivo: a tool to study extreme thermophile biology. *Environmental Microbiology.* 2008; 10(3):605–613. [PubMed: 18190515]
18. Huang L, Pike D, Sleat DE, Nanda V, Lobel P. Potential Pitfalls and Solutions for Use of Fluorescent Fusion Proteins to Study the Lysosome. *PLoS ONE.* 2014; 9(2)
19. Bence NF, Sampat RM, Kopito RR. Impairment of the Ubiquitin-Proteasome System by Protein Aggregation. *Science.* 2001; 292(5521):1552–1555. [PubMed: 11375494]

20. Henche A-L, Koerdt A, Ghosh A, Albers S-V. Influence of cell surface structures on crenarchaeal biofilm formation using a thermostable green fluorescent protein. *Environmental Microbiology*. 2012; 14(3):779–793. [PubMed: 22059595]
21. Stepanenko OV, Stepanenko OV, Shcherbakova DM, Kuznetsova IM, Turoverov KK, Verkhusha VV. Modern fluorescent proteins: from chromophore formation to novel intracellular applications. *BioTechniques*. 2011; 51(5):313–327. [PubMed: 22054544]
22. Stepanenko OV, Verkhusha VV, Kazakov VI, Shavlovsky MM, Kuznetsova IM, Uversky VN, Turoverov KK. Comparative Studies on the Structure and Stability of Fluorescent Proteins EGFP, zFP506, mRFP1, “dimer2”, and DsRed1†. *Biochemistry*. 2004; 43(47):14913–14923. [PubMed: 15554698]
23. Verkhusha VV, Kuznetsova IM, Stepanenko OV, Zaraisky AG, Shavlovsky MM, Turoverov KK, Uversky VN. High Stability of Discosoma DsRed As Compared to Aequorea EGFP†. *Biochemistry*. 2003; 42(26):7879–7884. [PubMed: 12834339]
24. Dai M, Fisher HE, Temirov J, Kiss C, Phipps ME, Pavlik P, Werner JH, Bradbury ARM. The creation of a novel fluorescent protein by guided consensus engineering. *Protein Engineering Design and Selection*. 2007; 20(2):69–79.
25. Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein expression and purification*. 2005; 41(1):207–234. [PubMed: 15915565]
26. Don Paul C, Traore DAK, Byres E, Rossjohn J, Devenish RJ, Kiss C, Bradbury A, Wilce MCJ, Prescott M. Expression, purification, crystallization and preliminary X-ray analysis of eCGP123, an extremely stable monomeric green fluorescent protein with reversible photoswitching properties. *Acta Crystallographica Section F Structural Biology and Crystallization Communications*. 2011; 67(10):1266–1268. [PubMed: 22102044]
27. Leslie AGW, Powell HR. *Evolving Methods for Macromolecular Crystallography - The Structural Path to the Understanding*. 2007
28. Evans PR. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D Biological Crystallography*. 2011; 67(4):282–292. [PubMed: 21460446]
29. Ebisawa T, Yamamura A, Kameda Y, Hayakawa K, Nagata K, Tanokura M. The structure of mAG, a monomeric mutant of the green fluorescent protein Azami-Green, reveals the structural basis of its stable green emission. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2010; 66(Pt 5):485–489.
30. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. *HKL* – 3000: the integration of data reduction and structure solution – from diffraction images to an initial model in minutes. *Acta Crystallographica Section D Biological Crystallography*. 2006; 62(8):859–866. [PubMed: 16855301]
31. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. Overview of the *CCP4* suite and current developments. *Acta Crystallographica Section D Biological Crystallography*. 2011; 67(4):235–242. [PubMed: 21460441]
32. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography*. 2010; 66(2):213–221. [PubMed: 20124702]
33. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of *Coot*. *Acta Crystallographica Section D Biological Crystallography*. 2010; 66(4):486–501. [PubMed: 20383002]
34. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. *MolProbity*: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography*. 2009; 66(1):12–21. [PubMed: 20057044]
35. Krissinel E, Henrick K. Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*. 2007; 372(3):774–797. [PubMed: 17681537]

36. Alberti S, Gitler AD, Lindquist S. A suite of Gateway® cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. *Yeast*. 2007; 24(10):913–919. [PubMed: 17583893]
37. Alberti S, Halfmann R, King O, Kapila A, Lindquist S. A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. *Cell*. 2009; 137(1):146–158. [PubMed: 19345193]
38. Hoffman TL, Canziani G, Jia L, Rucker J, Doms RW. A biosensor assay for studying ligand-membrane receptor interactions: Binding of antibodies and HIV-1 Env to chemokine receptors. *Proceedings of the National Academy of Sciences*. 2000; 97(21):11215–11220.
39. Sulli C, Banik SSR, Schilling J, Moser A, Xiang X, Payne R, Wanless A, Willis SH, Paes C, Rucker JB, Doranz BJ. Detection of Proton Movement Directly across Viral Membranes To Identify Novel Influenza Virus M2 Inhibitors. *J Virol*. 2013; 87(19):10679–10686. [PubMed: 23885079]
40. Davidoff C, Payne RJ, Willis SH, Doranz BJ, Rucker JB. Maturation of the Gag core decreases the stability of retroviral lipid membranes. *Virology*. 2012; 433(2):401–409. [PubMed: 22995186]
41. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D Biological Crystallography*. 2004; 60(12):2256–2268. [PubMed: 15572779]
42. Örmö M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal Structure of the *Aequorea victoria* Green Fluorescent Protein. *Science*. 1996; 273(5280):1392–1395. [PubMed: 8703075]
43. Stiel Andre C, Trowitzsch S, Weber G, Andresen M, Eggeling C, Hell Stefan W, Jakobs S, Wahl Markus C. 1.8 Å bright-state structure of the reversibly switchable fluorescent protein Dronpa guides the generation of fast switching variants. *Biochemical Journal*. 2007; 402(1)
44. Niwa H, Inouye S, Hirano T, Matsuno T, Kojima S, Kubota M, Ohashi M, Tsuji FI. Chemical nature of the light emitter of the *Aequorea* green fluorescent protein. *PNAS*. 1996; 93(24):13617–13622. [PubMed: 8942983]
45. Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, Zacharias DA, Tsien RY. A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences*. 2002; 99(12):7877–7882.
46. Bevis BJ, Glick BS. Rapidly maturing variants of the *Discosoma* red fluorescent protein (DsRed). *Nature Biotechnology*. 2002; 20(1):83–87.
47. Baird GS, Zacharias DA, Tsien RY. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *PNAS*. 2000; 97(22):11984–11989. [PubMed: 11050229]
48. Karasawa S, Araki T, Yamamoto-Hino M, Miyawaki A. A Green-emitting Fluorescent Protein from *Galaxeidae* Coral and Its Monomeric Version for Use in Fluorescent Labeling. *Journal of Biological Chemistry*. 2003; 278(36):34167–34171. [PubMed: 12819206]
49. Derewenda ZS. Application of protein engineering to enhance crystallizability and improve crystal properties. *Acta Crystallogr D Biol Crystallogr*. 2010; 66(Pt 5):604–615. [PubMed: 20445236]
50. Fields GB, Alonso DOV, Stigter D, Dill KA. Theory for the aggregation of proteins and copolymers. *J Phys Chem*. 1992; 96(10):3974–3981.
51. Krebs Mark RH, Domike Kristin R, Donald Athene M. Protein aggregation: more than just fibrils. *Biochemical Society Transactions*. 2009; 37(4)
52. Lawrence MS, Phillips KJ, Liu DR. Supercharging Proteins Can Impart Unusual Resilience. *Journal of the American Chemical Society*. 2007; 129(33):10110–10112. [PubMed: 17665911]
53. Der BS, Kluwe C, Miklos AE, Jacak R, Lyskov S, Gray JJ, Georgiou G, Ellington AD, Kuhlman B. Alternative Computational Protocols for Supercharging Protein Surfaces for Reversible Unfolding and Retention of Stability. *PLoS ONE*. 2013; 8(5)
54. Dale GE, Broger C, Langen H, Arcy AD, Stüber D. Improving protein solubility through rationally designed amino acid replacements: solubilization of the trimethoprim-resistant type S1 dihydrofolate reductase. *Protein Engineering*. 1994; 7(7):933–939. [PubMed: 7971955]
55. Price Li WN, Chen Y, Handelman SK, Neely H, Manor P, Karlin R, Nair R, Liu J, Baran M, Everett J, Tong SN, Forouhar F, Swaminathan SS, Acton T, Xiao R, Luft JR, Lauricella A, DeTitta GT, Rost B, Montelione GT, Hunt JF. Understanding the physical properties that control

- protein crystallization by analysis of large-scale experimental data. *Nature Biotechnology*. 2009; 27(1):51–57.
56. Cooper DR, Boczek T, Grelewska K, Pinkowska M, Sikorska M, Zawadzki M, Derewenda Z. Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallographica Section D Biological Crystallography*. 2007; 63(5):636–645. [PubMed: 17452789]
 57. Longenecker KL, Garrard SM, Sheffield PJ, Derewenda ZS. Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. *Acta Crystallographica Section D Biological Crystallography*. 2001; 57(5):679–688. [PubMed: 11320308]
 58. Garrard SM, Longenecker KL, Lewis ME, Sheffield PJ, Derewenda ZS. Expression, Purification, and Crystallization of the RGS-like Domain from the Rho Nucleotide Exchange Factor, PDZ-RhoGEF, Using the Surface Entropy Reduction Approach. *Protein Expression and Purification*. 2001; 21(3):412–416. [PubMed: 11281715]
 59. Fujiwara K, Toda H, Ikeguchi M. Dependence of α -helical and β -sheet amino acid propensities on the overall protein fold type. *BMC Structural Biology*. 2012; 12(1)
 60. Sawle L, Ghosh K. How Do Thermophilic Proteins and Proteomes Withstand High Temperature? *Biophysical Journal*. 2011; 101(1):217–227. [PubMed: 21723832]
 61. Takano K, Funahashi J, Yamagata Y, Fujii S, Yutani K. Contribution of water molecules in the interior of a protein to the conformational stability. *Journal of Molecular Biology*. 1997; 274(1): 132–142. [PubMed: 9398521]
 62. Matthews BW, Liu L. A review about nothing: Are apolar cavities in proteins really empty? *Protein Science*. 2009; 18(3):494–502. [PubMed: 19241368]
 63. Chou PY, Fasman GD. Empirical Predictions of Protein Conformation. *Annual Review of Biochemistry*. 1978; 47(1):251–276.
 64. Landgraf D, Okumus B, Chien P, Baker TA, Paulsson J. Segregation of molecules at cell division reveals native protein localization. *Nature Methods*. 2012; 9(5):480–482. [PubMed: 22484850]
 65. Halfmann R, Jarosz DF, Jones SK, Chang A, Lancaster AK, Lindquist S. Prions are a common mechanism for phenotypic inheritance in wild yeasts. *Nature*. 2012; 482(7385):363–368. [PubMed: 22337056]
 66. Salnikova AB, Kryndushkin DS, Smirnov VN, Kushnirov VV, Ter-Avanesyan MD. Nonsense Suppression in Yeast Cells Overproducing Sup35 (eRF3) Is Caused by Its Non-heritable Amyloids. *Journal of Biological Chemistry*. 2005; 280(10):8808–8812. [PubMed: 15618222]
 67. Swanstrom, R.; Wills, JW. Synthesis, Assembly, and Processing of Viral Proteins. In: Coffin, JM.; Hughes, SH.; Varmus, HE., editors. *Retroviruses*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1997.
 68. Engelenburg SBV, Shtengel G, Sengupta P, Waki K, Jarnik M, Ablan SD, Freed EO, Hess HF, Lippincott-Schwartz J. Distribution of ESCRT Machinery at HIV Assembly Sites Reveals Virus Scaffolding of ESCRT Subunits. *Science*. 2014; 343(6171):653–656. [PubMed: 24436186]
 69. Woodley JM. Protein engineering of enzymes for process applications. *Current Opinion in Chemical Biology*. 2013; 17(2):310–316. [PubMed: 23562542]
 70. Maity TS, Close DW, Valdez YE, Nowak-Lovato K, Marti-Arbona R, Nguyen TT, Unkefer PJ, Hong-Geller E, Bradbury ARM, Dunbar J. Discovery of DNA operators for TetR and MarR family transcription factors from *Burkholderia xenovorans*. *Microbiology*. 2012; 158(2):571–582. [PubMed: 22117006]
 71. Blakeley MP, Langan P, Niimura N, Podjarny A. Neutron crystallography. *Current opinion in structural biology*. 2008; 18(5):593–600. [PubMed: 18656544]

Figure 1.

The crystal structure of eCGP123. (A) Three views of the structure with a cutaway view showing the chromophore (shown in yellow) and the adjacent coaxial helix. (B) The QYG chromophore shown in stick representation with $2mF_o-DF_c$ electron density from the refined structure contoured at 1.5σ . (C) The chromophore environment with critical residues discussed in the text is shown. Red spheres represented ordered water molecules modeled in the structure.

- Putative stabilizing mutations. Mutations present in TGP and eCGP123 but not in either mAG or CGP.

Figure 2.

eCGP123 was subjected to structure guided surface engineering to generate the more soluble TGP. (A) There are eight protomers in the asymmetric unit of the eCGP123 crystal structure. Interfaces between protomers and between symmetry-related molecules were disrupted through selected amino acid substitutions. (B) An example is shown of a substantial interface between protomers C and E observed in the crystal lattice. C-terminal residues 217–223 are well ordered in the structure due to stabilizing interactions with the adjacent protomer. Amino acids in this region were substituted to improve solubility. (C) Several surface locations were targeted in an effort to disrupt protomer interfaces. For example, stabilization of the E:C protomer interface through contacts between Arg149 (protomer C) and Glu96 and His168 (protomer E) was reduced by the Arg149Glu substitution. (D) A total of 11 amino acid substitutions were made to eCGP123 to generate TGP. The amino acid sequence of eCGP123 and TGP are shown aligned to mAG⁴⁸. Amino acids in positions enclosed by red dashed boxes are those considered likely to contribute to

the increased thermal stability of eCGP123 and TGP based on comparison with mAG and CGP.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3.

Some optical and biophysical properties of eCGP123 and TGP.

(A) Fluorescence images of eCGP123 and TGP solutions each at 1.5 mg/ml after an overnight dialysis against PBS at 4°C. The cassette containing eCGP123 (left) is precipitating whereas TGP (right) is not. (B) Centrifugation of the samples in panel (A), shows that nearly all of the eCGP123 has precipitated out of solution. (C) The time dependent changes in fluorescence emission of TGP, eCGP123, and mAG on incubation at 85°C or 90°C in buffer consisting of 50mM Tris pH 7.5, 150mM NaCl, 10% Glycerol, 2mM

EDTA. (E) The excitation and emission spectra (D) and molar absorptivities (E) of TGP, eCGP123 and mAG are shown.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4.

TGP structure and thermal stability. (A) The overall fold of TGP showing the location of amino acid substitutions (maroon spheres) that contribute to thermal stability. Detailed views showing TGP (green sticks) and mAG (blue sticks; PDB: 3ADF) superposed to highlight the substitutions Thr59Pro, Val60Ala (B), Lys208Arg (C) and Ala53Ser (D). Side chains corresponding to mAG are labeled in *italics*. In panels B–D, $2mF_o-DF_c$ electron density is shown for the fully refined TGP structure contoured at 1.0σ . See main text for detailed discussion.

Figure 5.

Comparison of EGFP and TGP as fusions in the SDD-AGE amyloid assay. (A) NM-EGFP was expressed in yeast cells either harboring a template for NM amyloid formation ([PIN+]) or not ([pin-]). Cells were lysed and resolved by SDD-AGE into monomeric, oligomeric, and amyloid species, as labeled. Proteins were visualized in the gel by fluorescence. Fluorescent oligomers are more prevalent in the [pin-] sample, suggesting that they may compete with amyloid polymerization. (B) [PIN+] cells expressing either NM-EGFP or NM-TGP were compared as in (A). (C) [PIN+] cells expressing either Mot3PrD-EGFP or

Mot3PrD-TGP were compared as in (A), except that the lysates in SDS-containing sample buffer were incubated at either 25°C or 95°C for ten minutes prior to loading. Following fluorescence visualization, proteins were transferred to nitrocellulose and probed for the natural hexahistidine motif in Mot3PrD to assess expression levels. Note that Mot3PrD amyloids react much more strongly to the antibody than do other species of Mot3PrD. Very large amyloids near the top of the gel did not transfer. The dotted line denotes intervening lanes that were removed for clarity. (D) Virus-like particles (VLPs) were produced using the retroviral core Gag protein from murine leukemia virus (Gag-only) or a fusion protein containing either EGFP (Gag-GFP) or TGP (Gag-TGP) and assayed for fluorescence intensity (485nm/535nm) using a Flexstation II 384 plate reader (Molecular Devices). (E) Virus-like particles were produced using Gag-nsTGP or surface labeled with either Dylight488 fluorophore or biotin and were assayed for fluorescence intensity (485nm/535nm). Surface labeling with Dylight488 was performed at two different fluorophore densities.

Table I

Data collection and refinement statistics

	eCGP123	TGP
Wavelength (Å)	1.542	1.542
Resolution range (Å)	29.85 – 2.10 (2.18 – 2.10)	28.59 – 1.90 (1.97 – 1.90)
Space group	<i>P</i> 1	<i>P</i> 2 ₁ 2 ₁
Unit-cell parameters (Å, °)	74.6 75.4 84.5 91.0 89.8 104.0	94.2 141.2 69.1 90.090.090.0
No. total reflections	195423 (18992)	353093 (29185)
No. unique reflections	98476 (9572)	72746 (6770)
Multiplicity	2.0 (2.0)	4.9 (4.3)
Completeness (%)	94.4 (91.7)	99.2 (93.4)
$\langle I/\sigma(I) \rangle$	9.6 (2.3)	12.3 (3.1)
Wilson B-factor (Å ²)	33.5	23.8
<i>R</i> _{merge} (%)	4.2 (30.4)	6.2 (34.9)
CC _{1/2}	0.997 (0.855)	0.996 (0.912)
CC*	0.999 (0.960)	0.999 (0.977)
Rwork/Rfree (%)	16.8/20.3	17.4/20.0
No. of non-hydrogen atoms	15372	7843
Protein	14149	6920
Ligands	192	96
Solvent	1031	816
Protein residues	1757	858
R.m.s deviations		
Bond lengths (Å)	0.004	0.007
Bond angles (°)	0.930	1.230
Ramachandran favored (%)	99.6	99.5
Ramachandran outliers (%)	0	0
Clashscore	0.88	1.97
Average B-factor (Å ²)	41.1	32.1
Protein	41.2	31.7
Ligands	32.8	22.3
Solvent	41.2	36.7
PDB code	4TZG	4TZA

Table II

Photophysical Data

Protein	Absorption		Fluorescence		
	λ_{max} (nm)	ϵ ($\text{M}^{-1} \text{cm}^{-1}$)	λ_{max} (nm)	Φ_{F}	τ (ns)
TGP	493	6.4×10^4	507	0.66	3.1
eCGP123	493	6.8×10^4	504	0.70	3.0
mAG	491	5.2×10^4	504	0.76	3.2