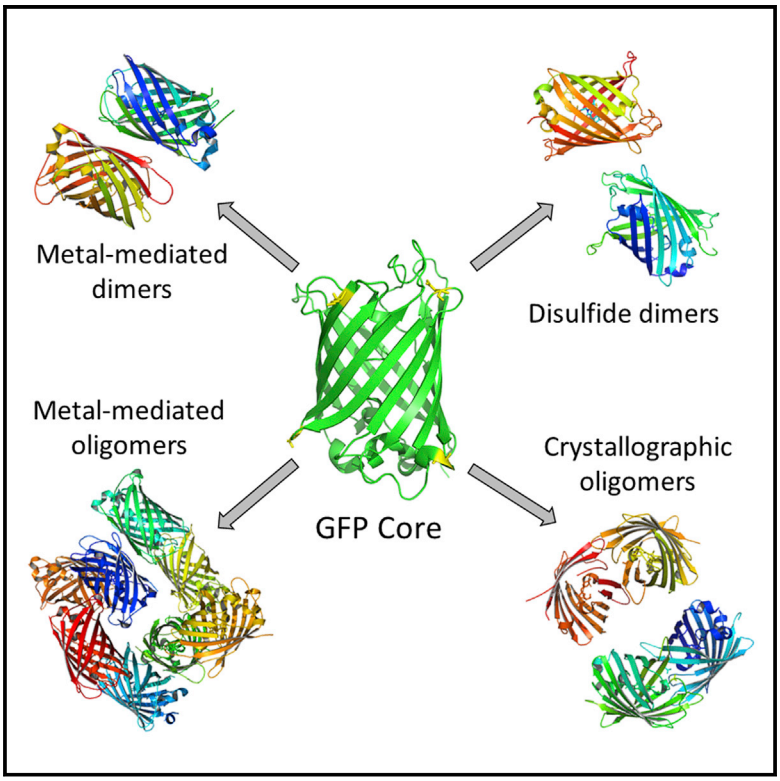
Resource

Structure

A Suite of Engineered GFP Molecules for Oligomeric Scaffolding

Graphical Abstract



Authors

David J. Leibly, Mark A. Arbing, Inna Pashkov, ..., Geoffrey S. Waldo, Thomas C. Terwilliger, Todd O. Yeates

Correspondence

yeates@mbi.ucla.edu

In Brief

Leibly et al. have generated a large suite of GFP oligomers with varying spatial arrangements of subunits, which are elucidated in 33 distinct crystal forms. These new GFP forms have potential applications ranging from synthetic biology to protein crystallization by connecting multiple proteins together in predefined spatial relationships.

Highlights

- A large suite of GFP mutants has been developed to allow diverse oligomerization
- 33 distinct crystal structures of designed GFP oligomers have been determined
- Target proteins can be oligomerized in diverse ways by attachment to the GFP molecules

Accession Numbers

4W69	4W6N	4W77
4W6A	4W6O	4W7A
4W6B	4W6P	4W7C
4W6C	4W6R	4W7D
4W6D	4W6S	4W7E
4W6F	4W6T	4W7F
4W6G	4W6U	4W7R
4W6H	4W72	
4W6I	4W73	
4W6J	4W74	
4W6K	4W7X	
4W6L	4W75	
4W6M	4W76	

Leibly et al., 2015, Structure *23*, 1754–1768
September 1, 2015
http://dx.doi.org/10.1016/j.str.2015.07.008

CellPress

CellPress

Structure
Resource

A Suite of Engineered GFP Molecules for Oligomeric Scaffolding

David J. Leibly,[1,2] Mark A. Arbing,[2] Inna Pashkov,[2] Natasha DeVore,[3] Geoffrey S. Waldo,[3] Thomas C. Terwilliger,[3] and Todd O. Yeates[1,2,*]

[1]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA
[2]UCLA-DOE Institute of Genomics and Proteomics, University of California, Los Angeles, CA 90095, USA
[3]Bioscience Division, Los Alamos National Laboratory, MS M888, Los Alamos, NM 87545, USA
*Correspondence: yeates@mbi.ucla.edu
http://dx.doi.org/10.1016/j.str.2015.07.008

SUMMARY

Applications ranging from synthetic biology to protein crystallization could be advanced by facile systems for connecting multiple proteins together in predefined spatial relationships. One approach to this goal is to engineer many distinct assembly forms of a single carrier protein or scaffold, to which other proteins of interest can then be readily attached. In this work we chose GFP as a scaffold and engineered many alternative oligomeric forms, driven by either specific disulfide bond formation or metal ion addition. We generated a wide range of spatial arrangements of GFP subunits from 11 different oligomeric variants, and determined their X-ray structures in a total of 33 distinct crystal forms. Some of the oligomeric GFP variants show geometric polymorphism depending on conditions, while others show considerable geometric rigidity. Potential future applications of this system are discussed.

INTRODUCTION

The general idea of connecting and spatially organizing multiple proteins is an emerging theme in synthetic biology. Notable applications include the spatial organization of multiple enzymes for metabolic pathway optimization (Conrado et al., 2008; Dueber et al., 2009; Lee et al., 2012), the organization of signaling molecules (Good et al., 2011; Zeke et al., 2009), and the creation of large self-assembling protein architectures (Lai et al., 2012). Another area under exploration is the synthetic organization of protein molecules into various symmetric forms to expand the chances of being able to induce them to form well-ordered crystals (Laganowsky et al., 2011). Facile systems for enabling the specific spatial organization of arbitrary proteins of interest could therefore advance research along various lines.

Ongoing efforts toward engineering proteins for improved crystallization stem from the generally low success rate and unpredictability of macromolecular crystallization (Sundstrom et al., 2006; Stacy et al., 2011). Regardless of the varied explanation for why many proteins are difficult to crystallize, the chances for a successful outcome might be improved by promoting the formation of intermolecular contacts that are compatible with crystal symmetry. Various methods for engineering proteins to improve their likelihood of forming good crystal contacts through surface residue mutations or fusion to a carrier protein have been described and reviewed (Banatao et al., 2006; Derewenda and Vekilov, 2006; Salgado et al., 2008; Forse et al., 2011; Corsini et al., 2008; Moon et al., 2010; Zou et al., 2012) including fusion to engineered GFPs (Suzuki et al., 2010).

Synthetic symmetrization, the engineering of artificially symmetric forms of a given protein molecule, has been promoted as one method for explicitly increasing the likelihood that a protein will be able to form a crystal lattice (Banatao et al., 2006). Two potential advantages have been articulated. First, geometric arguments and analysis of observed crystallization patterns suggests that a modest advantage can be gained by building symmetry into an otherwise asymmetric protein molecule by forcing it to oligomerize. Second and perhaps more important, the ability to produce multiple distinct symmetric forms of a target protein is a major advantage for crystallization. If the protein under study is the subject of crystallization trials, then each of the oligomeric constructs (e.g. specific dimers) is in effect a distinct molecular species with new opportunities to form lattice contacts in the context of a crystal. Distinct dimeric forms of a protein, for example, can be constructed by introducing single cysteine residues at various surface-exposed residues in a protein (Banatao et al., 2006; Forse et al., 2011). In another approach, metal-binding half-sites can be designed by introducing two potential metal-ligating residues (e.g. histidines) at proximal positions on the protein surface (Laganowsky et al., 2011). These experiments have shown that proteins engineered in such ways form oligomers that are rigid enough for facile crystallization, and that many new opportunities are opened up for the crystallization of a single given protein. In many cases, the new interactions introduced into the target protein contribute to the symmetry of the crystal (Banatao et al., 2006; Chruszcz et al., 2008).

Despite the potential for synthetic symmetrization to expand the opportunities for growing protein crystals, the method as it has been applied so far is experimentally burdensome. Its advantages are offset by the need to engineer multiple variants of the target protein, whose structure may be unknown, leading to potential challenges in conferring favorable properties without disrupting its fold. In this study, we explore a route for circumventing this obstacle. The essential idea is to apply the protein engineering work (i.e. to introduce varied forms of synthetic

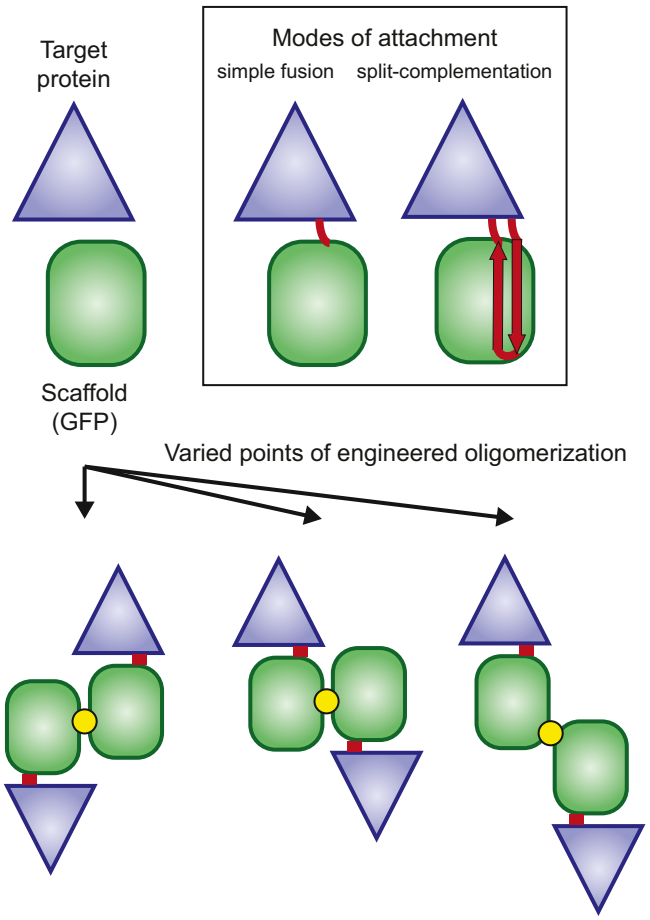


Figure 1. Concept of Scaffold-Mediated Synthetic Symmetrization
Here, GFP serves as a scaffold to induce synthetic symmetry. (Top) Multiple modes for attaching a target protein to GFP are indicated, including simple fusion and split-form complementation where the target protein is fused to a fragment of GFP, either strand 11 or strands 10–11. (Bottom) GFP (or another scaffold) is engineered in multiple ways to create varied oligomeric forms. When a target protein is connected (by fusion or complementation) to the engineered GFP molecules, varied oligomeric forms of the target protein are created.

symmetrization) to a model protein that can subsequently serve as a general carrier or scaffold for attaching otherwise arbitrary proteins. In this way a target protein can be driven into varied oligomeric forms with distinct opportunities to crystallize, without having to substantially compromise its native sequence. As long as the target protein is not much smaller than the scaffold to which it is attached, both components can be expected to participate in ordered contacts in a crystal.

Multiple strategies are possible for attaching a target protein to a scaffold protein, including by direct genetic fusion. Other possibilities are presented by a scaffold that can be produced and then reconstituted from two separate fragments. Here, we investigate the use of GFP as a scaffold for oligomerization, since GFP, particularly when accompanied by stabilizing mutations, can be expressed in split form and then functionally reconstituted from a large fragment and a small fragment (Cabantous et al., 2005, 2013; Huang and Bystroff, 2009; Nguyen et al., 2014). The key elements of the approach are illustrated in Figure 1. The use of monomeric split-GFP to complement and then crystallize another protein that is fused to the small GFP fragment has been already demonstrated in recent work (Nguyen et al., 2014). Here, the oligomerization element of the overall strategy is demonstrated by the construction and crystallographic investigation of several engineered variants of GFP. This large suite of engineered GFP proteins provides a foundation for various future developments, including those in the broad area of synthetic biology as well as in protein crystallization.

RESULTS

Rationale for GFP-Mediated Symmetrization

Engineered "split" forms of GFP have gained widespread use in the laboratory setting as biosensors (March et al., 2003) or fusion partners to probe for protein solubility (Cabantous et al., 2005, 2013). These mutants of GFP can be expressed without one or more terminal β strands of the 11 strands composing the GFP β barrel. Using circular permutants of a full-length GFP containing mutations developed for the split form of GFP (Cabantous et al., 2005), Huang and Bystroff (2009) created additional split-GFP pairs (with other tagging or "left-out" strands such as β strand 7). The partial core can then be complemented by addition of another protein that has been engineered to carry the missing GFP β strand(s), as either a terminal fusion or a loop insertion. Once complementation occurs the full β barrel is restored, and formation of the native chromophore provides a convenient readout of complex formation.

These previous developments make GFP well suited as a general carrier protein for implementing a new approach to the idea of synthetic symmetrization. The particular form of GFP used in our study can be split after strand 9, resulting in the GFP (strands 1–9) core and GFP (strands 10–11) hairpin (Cabantous et al., 2005; Nguyen et al., 2014). With this system, the hairpin formed by strands 10 and 11 can be engineered into a target protein, which will then complement GFP(1–9). In the simplest scenario, the (10–11) hairpin can be fused as an extension at either the N or C terminus of the target protein. However, the two-stranded hairpin allows for another particularly advantageous kind of construction. If the hairpin can be inserted at an internal sequence position on an exposed loop in the target protein, then the protein complex formed upon complementation will possess a two-chain crossing between the reconstituted GFP domain and the target protein structure (Figure 1). This is expected to enforce a much more rigid spatial arrangement between the two components, which could be an advantage, particularly where crystallization is the ultimate goal. In fact this has been demonstrated in one recent study, where a crystal structure revealed two copies of the molecular complex in the asymmetric unit in very nearly the same configuration, suggesting a limited range of motion when using the (10–11) hairpin insertion approach (Nguyen et al., 2014). Anticipating the ultimate advantage of the GFP(1–9) plus (10–11) hairpin approach, we focused our efforts on engineering oligomerizing variants of GFP in the strand 1–9 core region at positions remote from the (10–11) hairpin.

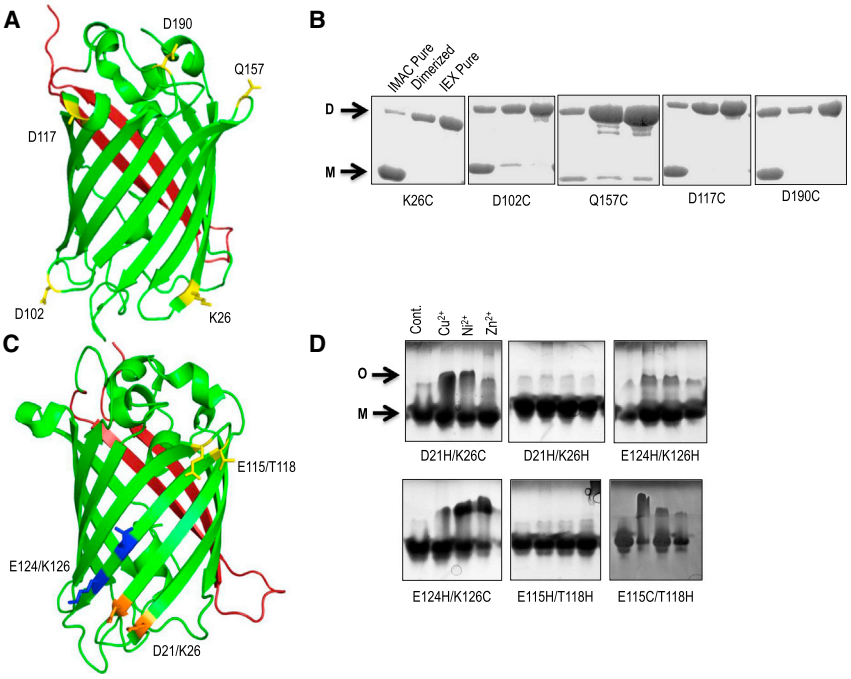


Figure 2. Locations of Point Mutations Introduced on Full-Length Split-GFP to Induce Oligomerization

(A) Locations of the individual point mutations to cysteines (yellow) on the GFP(1–9) core (green) on the opposite face of the β barrel from the GFP(10–11) hairpin (red).

(B) Each cysteine point mutant was purified in non-reducing conditions, and dimer formation was visualized on a non-reducing SDS-PAGE gel. After an initial IMAC step, GFP variants were dimerized with Cu^{2+}. The dimeric form, D, was then separated from the monomer, M, via anion exchange chromatography and used for crystallization experiments.

(C) Locations of the metal-half-site mutations on GFP (yellow, orange or blue); each site involves a pair of spatially proximal mutations (indicated). Color coding as in (A).

(D) Native PAGE screening of each metal-chelating mutation in the presence of Cu^{2+}, Ni^{2+}, and Zn^{2+}. This screen showed apparent oligomer formation for the D21H/K26C, E115C/T118H, E124H/K126H, and E124H/K126C variants, as determined by a mobility shift from the monomeric (M) band to the assumed oligomeric (O) band.

Design and Structure of Cysteine-Based GFP Dimers

In our first approach to engineering oligomerizing variants of GFP, individual cysteine residues were introduced at surface positions. Each such engineered protein was expected to produce a distinctly different dimeric structure upon oxidative disulfide formation. The utility of the disulfide-based approach to synthetic symmetrization has been demonstrated previously (Banatao et al., 2006; Forse et al., 2011). For application of the idea to GFP, we created five cysteine point mutations—K26C, D102C, D117C, Q157C, and D190C—as well as two sets of mutations to serve as either disulfide or metal-mediated oligomers (discussed subsequently): E115C/T118H and E124H/K126C. These amino acids were selected for mutation based on their polarity, surface location, and distance from strands 10–11 in order to limit interference with complementation when ultimately expressed in the split form (Figure 2). As the starting or wild-type sequence for design of the point mutations, we chose the sequence of split-GFP in its full-length form (Cabantous et al., 2013) using the superfolder GFP structure as a reference for point mutations in solvent-exposed locations (Pédelacq et al., 2006). Two native cysteines at positions C48 and C70 were first mutated to alanine to prevent subsequent interference with disulfide-based dimerization; one exception was an initial experiment and crystal structure of the K26C mutant of the superfolder form (PDB: 4W6B) in which only the cysteine at position 48 had been removed. The ultimate goal of our study is to use engineered versions of the truncated GFP(1–9) to synthetically symmetrize target proteins bearing the (10–11) hairpin, but we judged it prudent to first conduct the GFP engineering experiments against the background of the complete GFP(1–11) construct. Full-length GFP constructs bearing the single engineered cysteine residue were therefore expressed, purified, and oxidized to form homogeneous dimers (Figure 2). For all five of the cysteine sites chosen, pure dimers could be obtained in good yield with ~20–50 mg of protein obtained from 2 l of auto-induction media.

With the exception of Q175C, crystals grew readily in 1–7 days. Depending on the mutant, diffraction-quality crystals grew in as few as one condition for K126C or in more than 20 for D102C and D190C. Due to the large numbers of crystals that grew in the initial experiments, it was not feasible to screen X-ray diffraction in all crystals or to optimize all the crystal hits that were observed. We took the approach of screening crystals that appeared morphologically unique and large enough to mount for X-ray diffraction experiments. In some cases where initial crystals did not diffract despite having good morphology, minor optimization was performed, but otherwise crystals were taken directly from initial screens. Across the many crystal forms examined for the various mutants, the diffraction resolution ranged from 1.7 Å to poorer than 3.5 Å (Table 1). Rather than striving to maximize the resolution for the many crystal forms obtained, we focused on investigating the variety of crystal packing arrangements that these dimers could explore, and the degree to which they appeared to have well-ordered modes of dimerization.

In addition to the cases where we intentionally designed a disulfide bond to make GFP dimers, there were cases whereby we had anticipated the formation of a metal-binding site between GFP monomers involving a combination of an inserted histidine and cysteine pair, but obtained instead GFP dimers connected by a simple disulfide bond when the metal ion was added. These were mutant pairs D21H/K26C, E115C/T118H, and E124H/K126C. In these cases a disulfide bond was seen in the electron density map, but without evidence for metal binding at the dimer interface. These fortuitous dimers were not explored in depth to try to produce additional crystal forms, so their abilities to form alternative crystal lattices were not established.

Table 1. Summary of New GFP Crystal Forms

PDB	Mutation	Type	Space Group	Resolution (Å)	ASU[a]
4W69	Q157C	disulfide	P 43 21 2	3.98	2
4W6A	Q157C	disulfide	P 32 2 1	2.99	2
4W6B	K26C[b]	disulfide	P 21 21 21	1.90	2
4W6C	D21H/K26C[c]	disulfide	P 21 21 21	2.49	2
4W6D	K26C	disulfide	P 32 2 1	3.45	2
4W6F	D21H/K26C	disulfide	P 32 2 1	2.70	2
4W6G	D190C	disulfide	P 61	3.02	2
4W6H	D190C	disulfide	P 65	1.95	2
4W6I	D190C	disulfide	P 21 21 21	2.63	2
4W6J	D117C	disulfide	P 31 2 1	1.70	2
4W6K	D117C	disulfide	P 41 21 2	2.88	2
4W6L	D117C	disulfide	I 41 2 2	2.45	1
4W6M	D117C	disulfide	P 63	2.79	4
4W6N	D117C	disulfide	C 1 2 1	3.38	6
4W6O	D117C	disulfide	P 64 2 2	2.60	1
4W6P	D102C	disulfide	P 21 21 21	3.09	8
4W6R	D102C[c]	disulfide	P 1	3.47	16
4W6S	D124H/K126C	disulfide	P 43 21 2	3.10	2
4W6T	E115H/T118H	Cu-mediated contacts	P 43 21 2	1.60	1
4W6U	E115H/T118H	Ni-mediated contacts	P 21 21 21	2.28	4
4W72	E115C/T118H	disulfide + metal contacts	P 21 21 21	1.99	2
4W73	E115C/T118H	disulfide	P 21 21 21	2.79	2
4W74	E115C/T118H	Zn crystal contacts	P 1 21 1	2.10	8
4W7X	E115C/T118H	disulfide	P 1 21 1	2.80	4
4W75	D21H/K26C[c]	disulfide + metal contacts	P 21 21 21	3.47	2
4W76	D21H/K26C[c]	disulfide + metal contacts	P 21 21 21	2.35	2
4W77	D21H/K26C[c]	disulfide + metal contacts	P 21 21 21	3.10	2
4W7A	D21H/K26C[c]	disulfide + metal contacts	P 21 21 21	3.60	4
4W7C	D21H/K26C[c]	disulfide + metal contacts	C 1 2 1	2.50	4
4W7D	D21H/K26H	Cu crystal contacts	P 21 21 21	1.80	2
4W7E	D21H/K26H	Cu crystal contacts	P 41 21 2	2.59	1
4W7F	D124H/K126H	Cu crystal contacts	C 2 2 21	2.90	1
4W7R	D124H/K126H	Cu dimers	P 1 21 1	1.80	4

[a]Number of GFP chains in the asymmetric unit.
[b]Superfolder GFP C48A backbone mutation.
[c]Split-GFP C48A backbone mutation. All other sequences have the double mutations of C48A and C70A.

In all, we were able to characterize 20 distinctly different crystal forms of the GFP disulfide dimers and solve their structures (Table 1), with an additional six dimers containing both a disulfide bond and metal contacts. In all these structures, we modeled disulfide bonds into the electron density maps where possible, tabulating standard geometric terms and bond energies for the observed disulfide bonds (Tables 2 and S1) (Katz and Kossiakoff, 1986). In some cases where the resolution was limited this was not possible, and in at least two cases it appeared that the disulfide bond had been broken during the course of the X-ray diffraction experiment due to synchrotron radiation damage, as has been observed before (Carugo and Carugo, 2005; Weik et al., 2000).

The occurrence of multiple crystal forms for individual mutants, and the presence in several cases of multiple crystallographically independent GFP dimers in the unit cell, made it possible to analyze the range of conformations and degree of flexibility in these engineered dimers (Figure 3). An analysis of the symmetry and variations due to disulfide bond flexibility was performed for each cysteine mutation by comparing all dimers that were observed for a given point mutation (Figure 4; Table 2). In each case, we calculated the angle of rotation between the two subunits connected by the engineered disulfide bond to judge whether the synthetically generated dimers were nearly symmetric (i.e. related by a 180° rotation) (Table 2). Then, to evaluate how rigidly connected the two subunits were, we examined the degree of geometric variability between multiple instances of the same dimer as observed across different crystal forms or different asymmetric units of the same crystal form. A complete analysis is provided in Tables S2 and S3, and

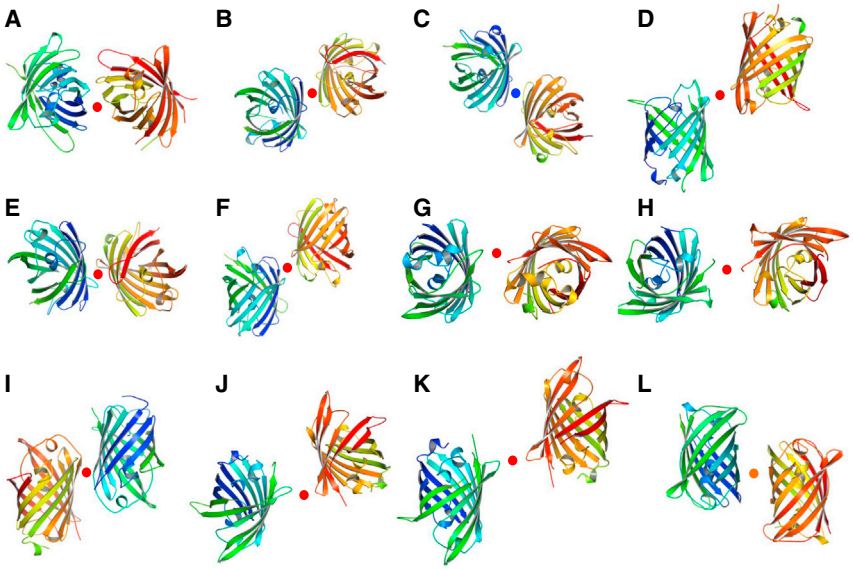


Figure 3. Observed Examples of the GFP Dimer

The internal rotation axis relating the subunits of each dimer is shown (red dot for disulfide dimers, blue for the mixed dimer (L), and orange for the metal-mediate dimer (C)). For each dimer the rotation axis corresponds to the location of the engineered disulfide bond, or metal-mediated crystal contact.

(A–L) The 12 dimers shown are from structures PDB: (A) 4W6B, (B) 4W6C, (C) 4W7C, (D) 4W6R, (E) 4W7X, (F) 4W6M, (G) 4W6G, (H) 4W6I, (I) 4W6S, (J) 4W69, (K) 4W6K, and (L) 4W7R. These dimers are representative of the complete set of 43 total dimers visualized in this work.

summarized in Figure 4 and Table 2. A description of the individual disulfide-bonded GFP structures is as follows.

K26C

Four crystal forms of K26C dimers were observed (PDB: 4W6B, 4W6C, 4W6D, and 4W6F), two in each of the space groups $P2_12_12_1$ and $P3_221$. Of these, PDB: 4W6C was the most symmetric (175.6°) while PDB: 4W6F was the least symmetric (144.3°). PDB: 4W6C, 4W6D, and 4W6F were most similar to each other with a maximum variation of 33.3°, while PDB: 4W6B varied by a rotation of up to 140.4° when overlaid on the others (Table 2; Figure 4B). Two of the structures (PDB: 4W6F and 4W6C) in which GFP dimers were obtained through a disulfide bond at position 26 arose from a D21H/K26C double mutant initially designed for metal chelation. Unexpectedly, addition of Ni^{2+} resulted in formation of a disulfide bond between residues 26C from two protein molecules during the crystallization experiments on these variants.

D102C

Two crystal forms were observed for this mutant, one in space group P1 (PDB: 4W6R) and one in $P2_12_12_1$ (PDB: 4W6P). Crystals appearing in the P1 morphology (thin plates) were obtained in numerous conditions containing PEG polymers as the precipitant. We were able to solve the structure of PDB: 4W6R to 3.47 Å; this was the highest resolution we were able to obtain from all the crystals screened of the D102C mutant. This P1 crystal form had a total of eight disulfide-bonded dimers in the crystal asymmetric unit with an average angle between the chains of 167°. The eight dimers were remarkably similar, with a maximum angular variation of only 8° (Figure 4C; Table 2). Due to this small range of variation, the CCP4 program Zanuda (Winn et al., 2011) was used to investigate and rule out the possibility that some higher crystallographic symmetry had been missed in the initial structure determination. The PDB: 4W6P structure also contained four dimers in the asymmetric unit of $P2_12_12_1$. These dimers are less symmetric

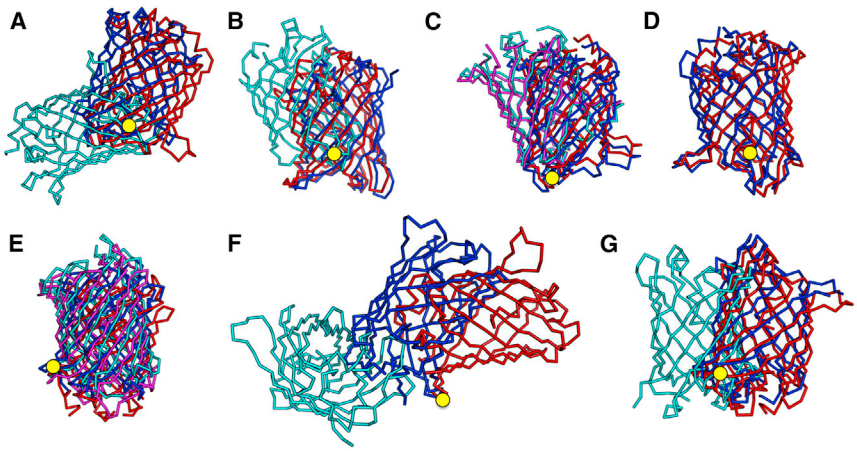


Figure 4. Chain Angle Ranges for Dimers

Depicted is the range of variation between the chain orientations for each disulfide-bonded dimer. Chain A of each dimer was first aligned to visualize the difference in the orientation of the distinct versions of chain B. Only the chain B backbone traces are depicted. Each panel illustrates the multiple conformations observed for one specific cysteine mutant. The blue and red traces represent the range of orientations the chains adopted. When a single outlier is found it is shown in cyan. When two disparate groups of conformations are present they are shown in red and blue, and cyan and magenta. When more than one dimer was observed in the asymmetric unit, instances representing the extremes in conformation were chosen. The rotation axis that relates the two molecules, and which coincides roughly with the position of the point mutation(s), is indicated by a yellow circle.

(A–G) The PDB codes for structures and dimer chains displayed are (A) K26C: red, PDB: 4W6C; blue, PDB: 4W6F; cyan, PDB: 4W6B. (B) D21H/K26C: red, PDB: 4W7A AB dimer; blue, PDB: 4W7A CD dimer; cyan, PDB: 4W75. (C) D102C: red, PDB: 4W6P CD dimer; blue, PDB: 4W6P FG dimer; cyan, PDB: 4W6R AN dimer; magenta, PDB: 4W6R KL dimer. (D) E115C: red, PDB: 4W72; blue, PDB: 4W73. (E) D117C: red, PDB: 4W6O; blue, 4W6K; cyan, PDB: 4W6N BF dimer; magenta, PDB: 4W6J. (F) Q157C: red, PDB: 4W69; blue, PDB: 4W6A A dimers; cyan, PDB: 4W6A B dimer. (G) D190C: red, 4W6H; blue, 4W6I; cyan, 4W6G.

Table 2. GFP Disulfide Dimer Characterizations

Mutant	PDB	Dimer	Disulfide Cα Distance (Å)	Dimer Angle (°)	Grouped PDB	Chain "B" Variation Range (°)
K26C	4W6B	AB	6.4	151.66	group: 4W6C, 4W6D, 4W6F	group 4W6C–4W6F = 33.3 maximum range:
	4W6C	AB	6.2	175.55		
	4W6D	AB	6.2	158.12		
	4W6F	AB	5.6	144.29	outlier: 4W6B	4W6B–4W6D = 140.4
D21H/K26C	4W7A	AB	5.8	169.72	group: 4W7A, 4W7C, 4W76 4W77 outlier: 4W75	group: 4W7A AB–4W7A CD = 6.3 maximum range: 4W7A CD–4W75 = 32.1
	4W7A	CD	6.2	177.95		
	4W7C	AB	5.9	173.38		
	4W7C	CD	6.4	171.85		
	4W75	AB	6.2	151.90		
	4W76	AB	6.4	174.64		
	4W77	AV	6.1	173.00		
D102C	4W6P	AB	4.5	143.38	group 1: 4W6P group 2: 4W6R	group 1: 4W6P CD–4W6P FG = 8.3 group 2: 4W6R AN–4W6R KL = 7.7 maximum range: 4W6P FG–4W6R KL = 32.4
	4W6P	CD	4.6	146.21		
	4W6P	EH	4.6	143.79		
	4W6P	FG	4.6	139.64		
	4W6R	AN	5.2	165.37		
	4W6R	BI	4.7	165.15		
	4W6R	CD	4.1	170.66		
	4W6R	EJ	4.4	167.73		
	4W6R	FO	4.7	166.16		
	4W6R	GO	4.9	163.96		
	4W6R	HM	4.9	166.20		
	4W6R	KL	4.3	170.91		
E115C	4W7X	AB	6.2	166.40	group: 4W7X, 4W72, 4W73	4W72–4W73 = 12.3
	4W7X	CD	5.4	163.93		
	4W72	AB	5.9	159.85		
	4W73	AB	6.4	170.95		
D117C	4W6J	AB	5.7	154.89	group 1: 4W6K, 4W6L, 4W6M 4W6O group 2: 4W6J, 4W6N	group 1: 4W6O–4W6K = 16.4 group 2: 4W6N BF–4W6J = 10.8 maximum range: 4W6N BF–4W6M AC = 34.8
	4W6K	AB	5.7	166.82		
	4W6L	AB	5.5	180.0		
	4W6M	AC	5.6	178.44		
	4W6M	BD	6.5	178.14		
	4W6N	AD	6.1	148.41		
	4W6N	BF	6.3	146.59		
	4W6N	CE	6.4	146.87		
	4W6O	AB	5.5	179.97		
K126C	4W6S	AB	6.00	177.96	–	–
K126H	4W7R	AB	–	179.1	–	AB–CD = 1.7
	4W7R	CD		179.15		
Q157C	4W69	AB	5.5	141.18	–	4W96–4W6A B = 129
	4W6A	A	5.78	180.0		
	4W6A	B	11.7[a]	180.0		
D190C	4W6G	AB	5.8	140.95	group: 4W6G, 4W6H outlier: 4W6I	group: 4W6G–4W6H = 6.3 maximum range: 4W6H–4W6I = 41.4
	4W6H	AB	5.8	135.23		
	4W6I	AB	6.4	171.21		

[a]Potential disulfide broken during crystallization.

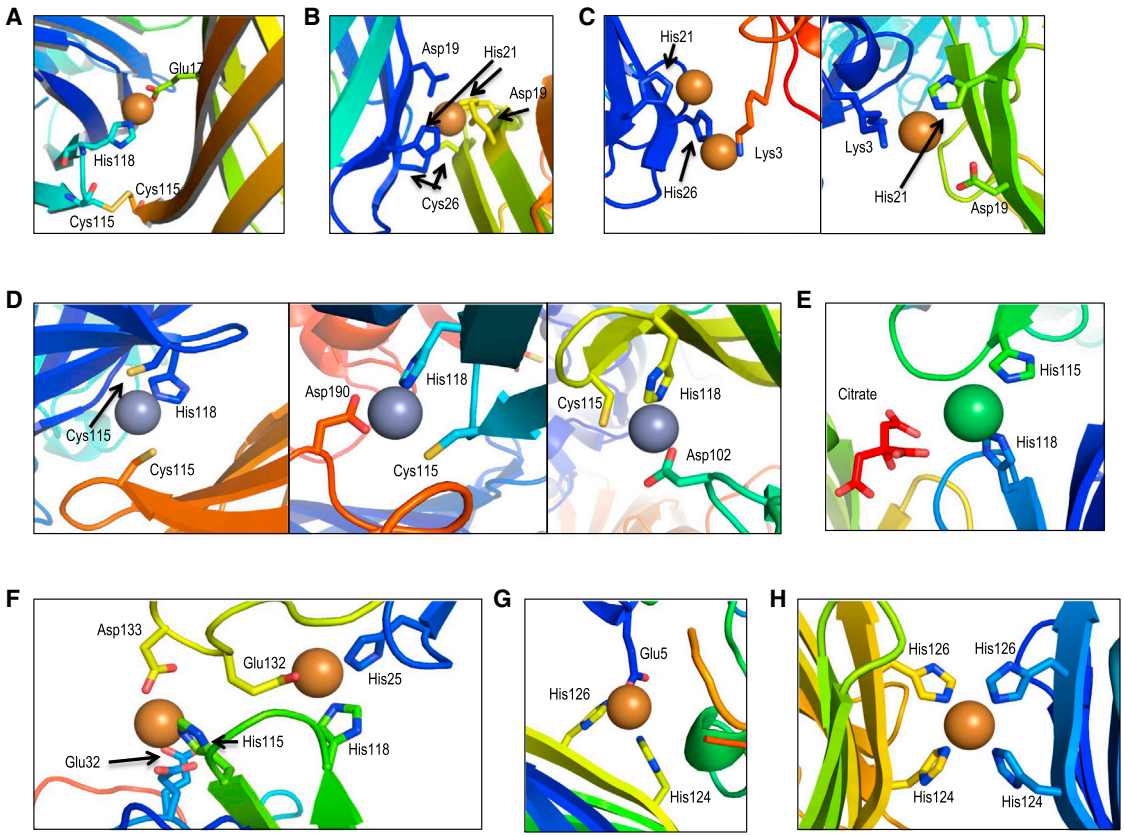


Figure 5. Observed Metal-Mediated Crystal Contacts

(A) Structure PDB: 4W72: a disulfide bond is formed in addition to the copper-binding site.

(B) Example of a mixed dimer from structure PDB: 4W76. Here the copper ion is chelated by histidine and aspartate residues from both molecules, and a disulfide bond is also formed.

(C) The two forms of metal-mediated contacts in PDB: 4W7D.

(D) The three observed zinc-mediated contacts found in PDB: 4W74. (Left) Cys115/His118 from one chain and Cys115 from another chain chelate the zinc. (Middle and right) Cys115/His118 from one chain chelate the zinc ion along with an aspartate (Asp190 or Asp102) from another chain.

(E) A nickel-mediated crystal contact in the structure of PDB: 4W6U involving histidines from the two proteins and a citrate molecule.

(F) A double copper-mediated crystal contact in the structure of PDB: 4W6T, both involving a combination of histidine and carboxylates.

(G) A copper-mediated crystal contact in the structure of PDB: 4W7F. His124 and His126 chelate the copper ion with Glu5 of the symmetry mate.

(H) Copper chelation by His124 and His126 of the symmetric dimer of PDB: 4W7R.

than those observed in the P1 form (average internal angle between subunits of ~143°). In comparison with the other dimeric forms in the same crystal asymmetric unit of this mutant, one dimer was a minor outlier, having a relative chain rotation between subunits of 5°–8°. The uniqueness of this dimer effectively rules out the possibility of any higher symmetry in the crystal.

E115C

Originally intended to serve as a metal-binding half-site, the mutated pair of residues, E115C/T118H, revealed disulfide-bonded dimer formation under crystallization conditions with the addition of metal ions. Four structures were obtained: three disulfide dimers (PDB: 4W72, 4W73, and 4W7X), and one structure with metal-mediated contacts only (PDB: 4W74, discussed subsequently). The three disulfide dimers feature an average rotation angle between subunits of 165°, with a variation up to 12° (Figure 4E; Table 2). Interestingly, PDB: 4W72 features a metal-mediated contact as well, involving the chelation of a copper ion by His118 of one chain A and Glu17 of another (Figure 5A).

D117C

This mutant resulted in six crystal forms, each in a different space group. The six dimers fall into two groups (Figure 4E; Table 2). Three of the dimeric forms observed (PDB: 4W6L, 4W6M, and 4W6O) are either perfectly symmetric with the two subunits related by crystal symmetry (PDB: 4W6L and 4W6O), or very nearly symmetric PDB: 4W6M, 179° rotation). PDB: 4W6J and 4W6M feature similarly asymmetric dimers (average internal angle of 149°), and PDB: 4W6K contains a dimer with an internal angle of 167°. It is notable that D117C dimers are rigid enough to form very well-ordered crystal lattices, diffracting up to 1.7 Å. Yet they are not locked into one conformation, and the permissible angular variation allows for multiple distinct lattices.

K126C

An intended metal-half-site pair, E124H/K126C (PDB: 4W6S) apparently underwent disulfide oxidation in the crystal drop, leading to a symmetric dimer (178°). Copper was added to the protein immediately prior to the crystallization experiment, and

no copper ions were observed in the crystal structure. No further efforts were undertaken to explore the possibility of additional space groups for this dimer.

Q157C

Two structures were solved from this mutant (PDB: 4W69 and 4W6A), and only after screening and optimization of crystal conditions. This is likely a result of the point mutation being located on a somewhat flexible loop of the GFP core. The best crystals diffracted to a resolution of 4 Å (PDB: 4W69). PDB: 4W6A represents an interesting and somewhat mysterious crystal form. Two chains are in the asymmetric unit, and they contribute to two different symmetric dimers sitting on axes of crystallographic symmetry, but the expected disulfide bonds are not present. The distance between the cysteine Cα positions of the two subunits is ~11 Å. The crystals took over 6 months to grow, and we suspect that the formate in the crystallization mixture may have slowly reduced the disulfide bonds initially present (Gibson, 1969). Based on the difficulties crystallizing this mutant, we do not view it as a favorable candidate for future crystallization experiments.

D190C

As with the Q157C point mutation, D190C is located in a flexible loop that is found to be disordered in many of the GFP structures presented in this study. This mutant resulted in >20 conditions with poorly diffracting crystals. We were still able to determine the structures of three D190C mutants (PDB: 4W6G, 4W6H, and 4W6I). PDB: 4W6I was the most symmetric dimer (171°) while PDB: 4W6G and 4W6H were asymmetric at 141° and 135°, respectively (Figure 4G; Table 2).

Taking all the observed disulfide dimers together, we note that only two of these are perfectly symmetric by virtue of lying on crystallographic axes of 2-fold symmetry. Of those that did not fall on symmetry axes, another nine had internal angles between the chains >170° (11 of 36 disulfide dimers observed). The remaining majority of dimers were substantially asymmetric. This contrasts with the trend toward nearly symmetric dimers noted in earlier studies on synthetically symmetrized proteins (Banatao et al., 2006; Forse et al., 2011) that had been connected primarily through α-helical segments rather than a β-sheet conformation as in GFP.

Design and Structure of Metal-Mediated GFP Oligomers

In addition to disulfide dimerization, we explored the possibility of forming dimers or higher oligomers of GFP by designing metal-binding half-sites on its surface. This second approach follows from the work conducted by the groups of Tezcan and Kuhlman (Salgado et al., 2008, 2010; Der et al., 2012). Here, the idea is that introducing a metal half-site into the surface of a protein will lead to assembly upon addition of metal. The utility of the metal-mediated approach to synthetic symmetrization has been demonstrated before, whereby it was found that, in addition to the intended dimeric forms, varied modes of assembly can be realized upon metal addition (Laganowsky et al., 2011). Previous efforts exploring engineered metal-mediated oligomer formation have focused on mutations in α-helical proteins. In those cases, residues i and i+4 can be mutated to metal-chelating residues. The mutations are typically to His/His or His/Cys pairs in an attempt to replicate native chelation motifs. We investigated whether a variation of the approach could be applied to GFP, which consists mainly of a single β barrel. We selected residues in three distinct regions of the protein to mutate to either His/His or His/Cys pairs. These mutations were residues i and i+2 on one β strand (E124/K126) or two residues on adjacent strands (D21/K26 and E115/T118) (Figure 2C). To evaluate their ability to form oligomers in the presence of metal ions, we analyzed purified proteins in the presence of Cu^{2+}, Ni^{2+}, and Zn^{2+} salts using native gel-shift assays. Additional metal ions (e.g. Fe^{2+}, Fe^{3+}, Cd^{2+}, and Co^{3+}) were screened, but these metals either indicated no oligomer formation or had non-reproducible results by our native gel-shift assay and were not pursued for crystallization studies. We determined that mutant pairs D21H/K26C, E115C/T118H, E124H/K126C, and E124H/K126H were all able to form oligomers in the presence of each of the ions (Figure 2D). All of these mutant-metal combinations were then used for crystallization experiments to determine their ability to sample different space groups and form metal-mediated crystal contacts. Although D21H/K26H and E115H/T118H did not show shifts on the native gel assay, we proceeded with the crystallization experiments to determine whether they could still form metal-mediated contacts during the crystallization process.

From these metal-mediated variants we solved seven unique structures that were dependent on metal chelation to form. As with the disulfide and mixed disulfide-metal dimers, an ability to crystallize in a variety of conditions was observed. In a range of other cases, however, the metal ions established crystal contacts between different GFP molecules through a combination of the engineered residues and other native residues (typically Asp and Glu) on the protein surface. Only one of these structures (PDB: 4W7R) formed a symmetric dimer, whereas the other cases involved more complex spatial arrangements. In several cases, owing to low resolution and poor electron density, it was difficult to determine the exact chelation of the metal ion by the protein side chains. In some instances this likely results from exposure to synchrotron radiation, which can change the oxidation state of metal ions or damage carboxylic acid groups in the chelating aspartic acid side chains (Carugo and Carugo, 2005; Weik et al., 2000).

D21H/K26C

The designed metal half-site mutation D21H/K26C resulted in either disulfide dimers discussed previously or a mixed dimer containing the disulfide and a chelated metal ion (PDB: 4W75, 4W76, 4W77, 4W7A, and 4W7C). Adjacent to the disulfide bond, a copper ion is chelated by residues Asp19 and His21 from both participating protein chains (Figure 5B); the mutated histidine was intended for chelation whereas the aspartate was fortuitous. Some of the structures have poor electron density for the Asp19 and His21 side chains, and it appears in some instances that only one of the residues from each chain is involved in the metal chelation. Structures PDB: 4W76, 4W77, 4W7A, and 4W7C are close to being symmetric (average angle of 173.4°), with PDB: 4W75 being more asymmetric at a 152° inter-subunit rotation. The symmetric structures are very similar to each other, with a variation upon overlap of 2°–8° (Figure 4B; Table 2).

D21H/K26H

Two structures of this variant were obtained having copper-mediated crystal contacts. In PDB: 4W7E, Asp19 and His21 of one chain and Gln184 of the symmetry mate chelate the copper

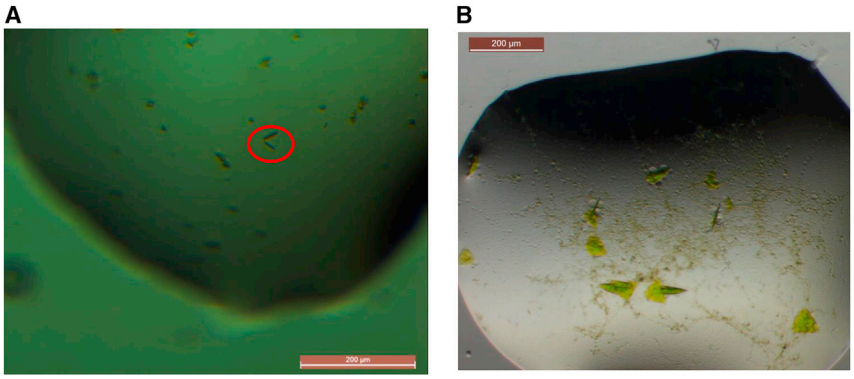


Figure 6. Crystals of Split-GFP with a Novel Crystallization Target

(A) Crystals of the STARD9-10/11-GFP1-9 (D21H/K26C) complex were obtained in a condition composed of 10% v/v 2-propanol, 0.1 M MES (pH 6.0), and 0.2 M $Ca(OAc)_2$. The protein complex was mixed in a 1:1 molar ratio with $CuSO_4$ immediately prior to the crystallization experiments. The green color of the crystals is used as an indication of the complex formation; the largest crystals observed to date (~20 µM in the largest dimension) are highlighted by the red circle.

(B) Crystals of a designed protein with an engineered internal (10–11) hairpin in complex with GFP1-9 (D117C). The triangular plate crystals (~50–75 µM) grew in a condition containing 0.1 M SPG buffer (pH 5.0) and 25% w/v PEG1500.

ion. This mutant crystallized in the presence of imidazole, leading to one imidazole molecule also being bound to the copper ion. Structure PDB: 4W7D features two different copper-mediated contacts (Figure 5C), and two chains are present in the asymmetric unit. Chain A makes contacts with two different protein molecules in the crystal using side chains that were engineered into this mutant. First, His21 and His26 chelate two copper ions and form a crystal contact to Lys3 of one neighboring molecule. A crystal contact to a different molecule is through Lys2 of chain A and Asp19 and His21 of the other protein, similar to the metal chelation observed in the D21H/K26C structures. The high pH (9.5) of this crystallization condition allows the lysine side chain to participate in the chelation of the copper ion.

E115C/T118H

In addition to the observed disulfide dimers of this mutant, structure PDB: 4W74 forms a complex system of metal-mediated crystal contacts between the eight protein chains in the asymmetric unit and six zinc ions via three different coordination sites (Figure 5D). The mutated Cys115/His118 half-site is found to chelate the zinc to a lone Cys115 in two cases; between chain A (Cys115/His118) and chain G (Cy115), and chain D (Cys115/His118) to chain F (Cys115). The Cys115/His118 half-site and an aspartic acid residue from a neighboring protein molecule chelate the other four zinc ions in arrangements that are generally similar to each other.

E115H/T118H

Two crystal forms of the E115H/T118H mutant with two different metal-mediated contacts were solved. PDB: 4W6U contains four chains in the asymmetric unit, yet only chains A and B feature a nickel-mediated contact. His118 of chain A and His115 of chain B are the residues responsible for metal chelation, along with a citrate molecule from the crystallization buffer (Figure 5E). A second nickel atom is chelated by residues His25 and Glu132 of chain A alone. In the structure PDB: 4W6T there is one chain in the asymmetric unit, which makes contact with other protein molecules through two copper ions (Figure 5F). His115 of the first chain and His25 and Glu132 of the symmetry mate chelate the first copper atom. His118 and Glu32 of the first chain and Asp133 of the symmetry mate chelate the second copper atom.

E124H/K126H

From the final mutant we determined two crystal structures, PDB: 4W7F and 4W7R. PDB: 4W7F contains one chain in the asymmetric unit with the copper-mediated contact formed between His124/His126 of the first chain and Glu5 of the symmetry mate (Figure 5G). The only symmetric metal-mediated dimer for which we obtained a structure is PDB: 4W7R. In this case the His124/His126 pair from chain A chelates a copper ion together with the His124/His126 pair from chain B. Two copper-mediated dimers (four subunits in total) are found in the asymmetric unit, and both dimers are nearly symmetric with chains orientated 179° apart. The two dimers are virtually identical, with only a 2° variation when aligned.

GFP Oligomers as a Crystallization Scaffold

After establishing in a previous study that a complex between the split-GFP(1–9) and a protein containing the (10–11) hairpin could form diffraction-quality crystals (Nguyen et al., 2014), we set out to crystallize a novel protein that had failed to crystallize in previous experiments. We attempted this with the motor domain of STARD9 (Torres et al., 2011), a monomeric kinesin that could serve as a target for novel anti-mitotic drug development. We co-expressed a construct of STARD9 as an N-terminal fusion to the GFP(10–11) hairpin together with the four metal-chelating GFP(1–9) mutants that consistently showed oligomerization in the native gel experiments (K26C/D21H, E124H/K126H, E124H/K126C, and E115C/T118H). Of the four experiments attempted, only K26C/D21H&E124H/K126H gave robust complementation. We were able to obtain crystals of the STARD9-10/11 and GFP1-9 (D21H/K26C) complex after approximately 3 months (Figure 6A). However, these crystals are small (~20 µM in the largest dimension) and have not produced well-ordered diffraction to date; optimization efforts are under way.

A second computationally designed 271-amino-acid protein (to be published) containing the (10–11) hairpin as a loop insertion was co-expressed with the cysteine mutant suite of split-GFPs, all of which resulted in robust complementation. After 7 months, triangular plate crystals (~50–75 µM) (Figure 6B) were observed containing the designed protein in complex with the GFP1-9 (D117C). As with the STARD9-10/11 constructs, optimization efforts of these crystals are under way.

DISCUSSION

The structural results presented here characterize a suite of engineered GFP molecules comprising a wide range of oligomeric

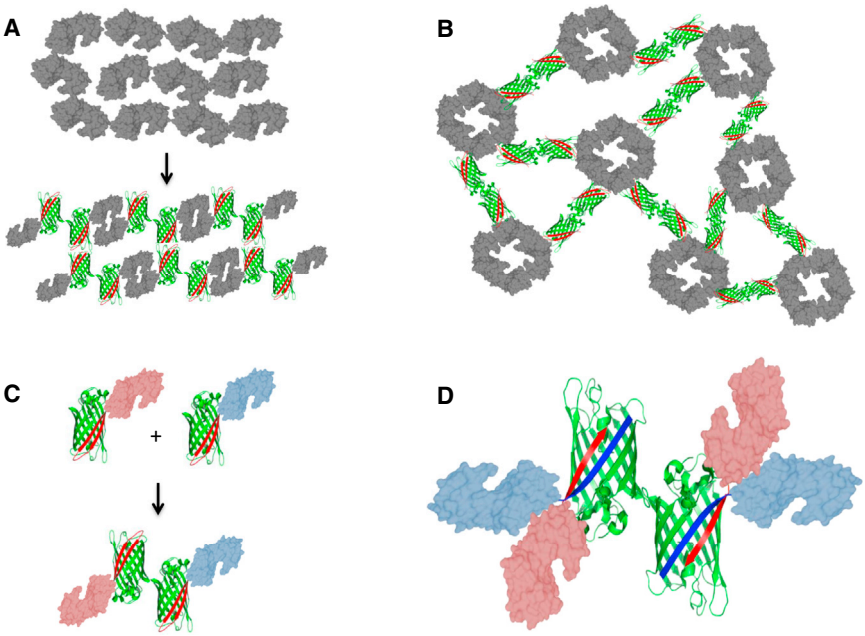


Figure 7. Alternative Applications for Engineered Oligomeric GFPs

Beyond their proposed utility as carriers for the crystallization of novel proteins, other potentially useful applications are possible.

(A) Fusion to GFP dimers could be used to change the crystal forms of existing proteins. Here a disordered crystal (top) can form a different and possibly better-ordered lattice (bottom) through fusion to one of the GFP oligomers in the available suite.

(B) Fusion to a multimeric enzyme, in this example a tetramer, could be used to create an enzymatically active amorphous gel for facile separation of enzymes and products for in vitro reaction systems.

(C) With the split form or through terminal fusions, the GFP dimers could be used to create a heterodimer for co-localization of enzymes for substrate channeling or co-crystallization experiments.

(D) Expanding on the idea from (C), two proteins can be forced into close proximity and further symmetrized, by separate genetic fusion of strand 10 to one protein and strand 11 to the other, then allowing them to complement for various applications.

forms, most of which appear highly amenable to crystallization on their own. We obtained 20 new crystal forms of seven disulfide-bonded dimers, plus 13 metal-mediated structures from five combinations of metal-chelating mutations in the presence of different metals. The 33 crystal forms are all distinct from each other (Table 1). Many of the engineered GFPs formed additional crystal forms in numerous conditions that were not pursued for structure determination. In analyzing individual GFP variants that were observed in multiple crystal forms, it was found that some of the oligomeric GFPs show strong geometric constraints between the disulfide-bonded subunits, while others display considerable geometric polymorphism. The K26C dimer was especially variable; among four instances observed for that dimer, the smallest angular deviation between any pair was 33°. The D21H/K26C and D102 mutants were the most rigid. Several instances of those dimers showed common conformations within about 8° deviation, although individual outliers were also obtained in both cases. The oligomeric GFP molecules designed here have not yet been used to successfully crystallize a target protein that was otherwise recalcitrant to crystallization. Which of the GFP constructs might ultimately be most useful in such a context is therefore presently unknown. However, it is notable that a few of the constructs formed an unusual number of distinct crystal forms readily. Among the disulfide-based dimers, the D117C construct formed the most (six) distinct crystal forms. Among the metal-mediated designs, the E115H/T118H and D21H/K26H constructs each also formed six distinct crystal forms.

The suite of oligomerizing GFP constructs designed here could be used for crystallizing target proteins by direct fusion. Alternatively, as noted above, our GFP constructs were engineered to be compatible with use in split form so that engineered variants of the GFP(1–9) construct can be reconstituted with a target protein bearing the (10–11) hairpin. In principle, this reconstitution can be performed in vivo (by co-expression) or in vitro (after separate purifications). Initial in vitro experiments using the split forms of the oligomerizing GFP constructs (not presented here) suggest that further optimization of the GFP(1–9) core may be important for stabilization in the context of the various mutations introduced into the GFP sequence. The counterbalancing advantages and disadvantages of the present system will also have to be compared with other strategies. For example, in some crystallization approaches the target protein is potentially stabilized by its fusion to an intact scaffold protein; attaching a small GFP fragment to a target protein (in the split-complementation approach) is not likely to provide such an advantage.

A principal long-term motivation for the present work is the crystallization of novel proteins, but other diverse applications in synthetic biology are likely to emerge for these oligomeric variants of GFP (Figure 7). One prospective application would be in attaching metabolically coupled enzymes together in different geometries through metal-mediated interactions or in vitro oxidized cysteines. They could be used as oligomerizing scaffolds for bringing together homo- or hetero-pairs of proteins into close proximity, in different spatial arrangements, and in ways that can be triggered by the addition of metal ions (Figures 7C and 7D). To promote formation of strictly heteromeric assemblies, future experiments would be required to design asymmetric versions of an oligomerizing carrier protein. Another avenue for future applications will be in using oligomerizing carrier proteins (GFP and others that could be developed) to drive other proteins or enzymes to form extended materials or amorphous gels (Figure 7B). While the motivating application emphasized in the present study (protein crystallization) applies primarily to target proteins that are naturally monomeric, we envision that extended materials, most likely with irregular structures, could be formed by complementing various oligomeric forms of the split-GFP(1–9) with naturally oligomeric proteins or enzymes bearing the (10–11) hairpin. In most cases this

Table 3. X-Ray Diffraction Data and Refinement Statistics

PDB:	4W69	4W6A	4W6B	4W6C	4W6D	4W6F	4W6G	4W6H	4W6I	4W6J	4W6K
Wavelength (Å)	1.0717	0.9789	0.97918	0.9789	0.9793	0.9792	0.9793	1.0717	1.0717	0.9793	1.0717
Resolution range (Å)	94.58–3.975 (4.117–3.975)	77.02–2.991 (3.098–2.991)	44.69–1.895 (1.963–1.895)	71.3–2.492 (2.581–2.492)	87.16–3.447 (3.57–3.447)	84.34–2.701 (2.798–2.701)	69.09–3.024 (3.132–3.024)	82.72–1.953 (2.023–1.953)	53.59–2.625 (2.719–2.625)	98.5–1.702 (1.763–1.702)	75.46–2.877 (2.98–2.877)
Space group	P 43 21 2	P 32 2 1	P 21 21 21	P 21 21 21	P 32 2 1	P 32 2 1	P 61	P 65	P 21 21 21	P 31 2 1	P 41 21 2
Unit cell	133.76 133.76 88.92 90 90 90	88.93 88.93 135.76 90 90 120	50.151 90.356 102.83 90 90 90	51.33 88.37 120.69 90 90 90	123.11 123.11 151.32 90 90 120	121.98 121.98 140.09 90 90 120	93.38 93.38 132.97 90 90 120	95.52 95.52 132.5 90 90 120	57.35 67.6 150.58 90 90 90	113.74,113.74 82.46 90 90,120	106.72 106.72 97.45 90 90 90
Total reflections	110,307 (10,414)	252,345 (23,254)	251,856 (13,789)	81,636 (7,060)	89,887 (8,957)	335,541 (32,238)	66,094 (6,525)	508,993 (47,571)	88,235 (6,817)	339,028 (33,029)	167,091 (13,967)
Unique reflections	7,344 (715)	12,990 (1,219)	37,298 (3,260)	19,311 (1,713)	17,869 (1,730)	33,538 (3,281)	12,859 (1,284)	49,488 (4,865)	17,334 (1,613)	67,258 (6,633)	13,200 (1,210)
Multiplicity	15.0 (14.6)	19.4 (19.1)	6.8 (4.2)	4.2 (4.1)	5.0 (5.2)	10.0 (9.8)	5.1 (5.1)	10.3 (9.8)	5.0 (4.2)	5.0 (5.0)	12.7 (11.5)
Completeness (%)	99.90 (99.31)	99.52 (95.08)	98.40 (87.52)	97.33 (88.79)	99.26 (98.69)	99.95 (99.64)	99.74 (99.46)	99.80 (98.06)	95.43 (82.54)	99.47 (98.82)	99.25 (93.51)
Mean $I/\sigma(I)$	16.9 (2.1)	20.3 (2.5)	12.3 (4.5)	7.3 (1.9)	11.91 (1.7)	5.6 (2.0)	16.7 (2.4)	16.4 (2.9)	5.9 (1.0)	16.6 (2.1)	22.1 (2.6)
Wilson B factor	162.1	72.3	20.6	60.3	112.8	64.4	95.5	25.4	54.4	25.2	91.6
R_{merge}	0.144 (1.663)	0.162 (1.385)	0.104 (0.376)	0.104 (0.660)	0.118 (1.023)	0.304 (0.578)	0.064 (0.729)	0.109 (0.819)	0.246 (1.047)	0.051 (0.702)	0.083 (1.265)
R_{meas}	0.149	0.167	0.113	0.118	0.132	0.321	0.072	0.114	0.273	0.057	0.087
$CC_{1/2}$	0.999 (0.714)	0.999 (0.767)	0.995 (0.871)	0.989 (0.881)	0.998 (0.607)	0.955 (0.871)	0.998 (0.853)	0.998 (0.811)	0.983 (0.853)	0.999 (0.782)	0.999 (0.811)
CC*	1 (0.913)	1 (0.932)	0.999 (0.965)	0.997 (0.968)	1 (0.869)	0.988 (0.965)	1 (0.959)	1 (0.947)	0.996 (0.959)	1 (0.937)	1 (0.946)
R_{work}	0.307 (0.457)	0.191 (0.304)	0.167 (0.201)	0.248 (0.443)	0.236 (0.363)	0.204 (0.258)	0.248 (0.384)	0.166 (0.200)	0.268 (0.504)	0.189 (0.259)	0.249 (0.351)
R_{free}	0.335 (0.398)	0.240 (0.389)	0.202 (0.280)	0.276 (0.445)	0.267 (0.318)	0.238 (0.290)	0.270 (0.361)	0.190 (0.229)	0.316 (0.606)	0.212 (0.289)	0.294 (0.395)
No. of non-hydrogen atoms	3,458	3,574	3,867	3,553	3,550	3,604	3,505	3,884	3,558	3,925	3,037
Macromolecules	3,414	3,530	3,599	3,509	3,505	3,539	3,461	3,635	3,514	3,623	2,993
Ligands	44	44	47	44	45	65	44	44	44	96	44
Water	0	0	221	0	0	0	0	205	0	206	0
Protein residues	434	446	454	443	445	445	437	457	443	451	378
RMS (bonds)	0.011	0.018	0.009	0.01	0.01	0.011	0.011	0.011	0.014	0.011	0.011
RMS (angles)	1.29	1.6	1.08	1.29	1.69	1.37	1.37	1.21	1.3	1.21	0.95
Ramachandran favored (%)	97	97	98	96	98	97	97	98	97	99	97
Ramachandran outliers (%)	0	0.23	0	0.23	0.7	0.23	0.48	0	0	0	0.29
Clashscore	21.2	14.2	1.4	12.2	17.5	6.2	21.2	1.5	5.4	4.9	9.3
Average B factor	191.0	64.0	25.0	63.0	124.1	68.6	171.7	26.9	56.1	33.9	94.8
Macromolecules	191.5	64.1	24.9	63.2	124.3	68.4	171.7	26.8	56.3	33.5	35.1
Ligands	147.9	54.6	18.5	50	107.6	77.1	175.5	20.1	42.6	38.9	78.3
Solvent	–	–	27.8	–	–	–	–	30.7	–	37.1	–

(Continued on next page)

would lead to runaway oligomerization, yielding materials with potentially novel properties and uses. Other synthetic biology applications may benefit from higher-order oligomers. Based on our crystal structures, there are possible interfaces that could be mutated to achieve this purpose. As an example, a novel tetrameric form of GFP could be based on the structure of the D117C mutant PDB: 4W6M. This structure features a tetramer composed of two symmetric dimers in the asymmetric unit of the crystal. Further mutations in the region of the fortuitous interaction between dimers (residues I206, S146, and N147), either via metal-mediated interactions or by computational sequence design of a more extensive interface, could create a higher oligomeric form of GFP.

EXPERIMENTAL METHODS

Cloning

Unless otherwise stated, primers were ordered from Valuegene, enzymes were from New England Biolabs, and DNA sequencing was performed by Genewiz. The plasmid construct containing the split-GFP (Cabantous et al., 2005, 2013) was used as a template to generate a construct with a C-terminal hexahistidine tag and the C terminus: … TAAGITHHHHHH. The GFP gene was PCR amplified with Phusion DNA polymerase using the primers GFP.For and GFP.Rev, which include NdeI and HindIII restriction sites, respectively, in the primer extensions. The PCR-amplified segment was purified, digested with NdeI and HindIII, and ligated into pET24a, which had been restriction digested with the same two enzymes. Colony PCR using T7 and T7 terminator primers was performed to identify putative positive clones whose DNA sequences were subsequently confirmed by DNA sequencing. Two cysteine residues (Cys48, Cys70) were mutagenized to alanine using the primers C48A.For.New./C48A.Rev.New. and C70A/C70A_antisense to eliminate the possibility of unintended disulfide bonds. The C48A mutation was made by linear PCR amplification of the target vector with Phusion DNA polymerase, followed by DpnI digestion of the template plasmid and subsequent phosphorylation of the gel-extracted DNA with T4 polynucleotide kinase and ligation with T4 DNA ligase. The C70A mutation was made using Pfu Turbo AD polymerase (Agilent) using the Quikchange mutagenesis procedure. Additional mutations were made in the GFP construct containing the C48A/C70A mutations by the Quikchange method to generate the following GFP mutant proteins: C48A/C70A/D102C, C48A/C70A/D117C, C48A/C70A/Q157C, C48A/C70A/K26C, C48A/C70A/D190C, C48A/C70A/E124H/K126H, and C48A/C70A/E115C/T118H.

Table 3. *Continued*

4W6L	4W6M	4W6N	4W6O	4W6P	4W6R	4W6S	4W6T	4W6U	4W72	4W73	4W74
1.0717	0.9793	0.9537	0.9793	0.9793	0.9792	0.9789	0.9795	0.9795	0.9795	0.9789	0.9795
76.67–2.45 (2.538–2.45)	73.83–2.793 (2.893–2.794)	88.89–3.375 (3.496–3.375)	67.51–2.6 (2.693–2.6)	79.58–3.085 (3.195–3.085)	89.88–3.471 (3.595–3.471)	68.28–3.1 (3.211–3.1)	74.46–1.604 (1.661–1.604)	82.99–2.278 (2.36–2.278)	57.12–1.996 (2.067–1.996)	52.18–2.787 (2.887–2.787)	88.27–2.099 (2.174–2.099)
I 41 2 2	P 63	C 1 2 1	P 64 2 2	P 21 21 21	P 1	P 43 21 2	P 43 21 2	P 21 21 21	P 21 21 21	P 21 21 21	P 1 21 1
108.43 108.43 101.47 90 90 90	170.5 170.5 79.57 90 90 120	181.21 102.68 84.13 90 101.44 90	77.95 77.95 178.88 90 90 120	86.05 117.86 209.1 90 90 90	92.42 92.56 124.53 94.94 96.17 102.25	91.51 91.51 205.11 90 90 90	105.3 105.3 69.61 90 90 90	47.64 116.58 165.98 90 90 90	72.54 74.4 89.15 90 90 90	69.74 70.58 77.5 90 90 90	67.44 119.79 130.57 90 89.99 90
148,335 (14,256)	63,420 (1,852)	70,644 (6,362)	160,637 (16,471)	261,903 (23,019)	87,921 (8,257)	418,083 (42,445)	653,315 (63,113)	279,406 (25,897)	238,183 (20,955)	62,325 (4,525)	411,267 (39,095)
11,419 (1,106)	32,658 (378)	20,957 (1964)	10,518 (991)	39,143 (3,478)	46,082 (4,386)	16,549 (1,612)	51,580 (4,985)	42,930 (4,011)	33,399 (3,272)	9,864 (848)	119,482 (11,566)
13.0 (12.9)	5.7 (4.9)	3.4 (3.2)	15.3 (16.6)	6.7 (6.6)	1.9 (1.9)	25.3 (26.3)	12.7 (12.7)	6.5 (6.5)	7.1 (6.4)	6.3 (5.3)	3.4 (3.4)
99.83 (98.57)	99.15 (95.61)	97.89 (91.99)	99.94 (99.70)	97.98 (88.91)	89.44 (85.24)	99.95 (99.94)	99.78 (97.98)	99.28 (94.55)	99.74 (98.55)	98.67 (87.69)	98.81 (95.89)
23.9 (3.4)	4.8 (3.8)	6.6 (1.5)	26.0 (3.4)	10.8 (2.2)	5.7 (1.4)	18.0 (3.6)	17.8 (1.6)	14.2 (1.9)	16.6 (2.4)	10.1 (1.7)	8.1 (1.5)
61.8	71.3	84.6	68.9	70.7	101.5	76.2	26.5	44.6	41.2	73.9	32.2
0.064 (0.840)	0.766 (0.789)	0.205 (0.778)	0.072 (0.934)	0.149 (0.847)	0.095 (0.467)	0.257 (1.844)	0.072 (1.129)	0.105 (0.906)	0.060 (0.821)	0.109 (0.851)	0.104 (0.775)
0.067	0.823	0.244	0.075	0.162	0.134	0.262	0.075	0.115	0.065	0.119	0.123
0.999 (0.979)	0.683 (0.49)	0.992 (0.746)	0.999 (0.938)	0.994 (0.76)	0.989 (0.773)	0.998 (0.947)	0.999 (0.821)	0.997 (0.739)	0.999 (0.895)	0.996 (0.70)	0.996 (0.747)
1 (0.995)	0.901 (0.811)	0.998 (0.924)	1 (0.984)	0.999 (0.929)	0.997 (0.934)	1 (0.986)	1 (0.95)	0.999 (0.922)	1 (0.972)	0.999 (0.908)	0.999 (0.925)
0.252 (0.375)	0.261 (0.389)	0.316 (0.400)	0.262 (0.348)	0.232 (0.315)	0.307 (0.409)	0.223 (0.284)	0.180 (0.260)	0.210 (0.277)	0.190 (0.292)	0.221 (0.366)	0.212 (0.303)
0.278 (0.450)	0.285 (0.394)	0.363 (0.469)	0.311 (0.415)	0.279 (0.363)	0.357 (0.431)	0.276 (0.372)	0.207 (0.289)	0.250 (0.319)	0.235 (0.314)	0.297 (0.464)	0.235 (0.330)
1,635	6,752	10,419	1,662	12,960	25,002	3,538	2,074	7,317	3,817	3,519	14,583
1,613	6,652	10,331	1,637	12,828	25,002	3,442	1,846	7,083	3,570	3,469	14,200
22	100	88	22	132	0	96	69	111	45	50	182
0	0	0	3	0	0	0	159	123	202	0	201
205	842	1,306	208	1,618	3,133	436	227	890	450	437	1,793
0.01	0.011	0.009	0.013	0.009	0.008	0.012	0.019	0.012	0.011	0.013	0.013
1.23	1.2	1.49	1.85	1.42	1.76	1.55	1.7	1.36	1.24	1.42	1.51
96	99	95	98	97	96	95	99	98	98	97	98
0	0.12	0.08	0	0.13	0.44	0.47	0	0.12	0	0	0
8.1	14.0	30.0	20.8	12.6	20.0	21.6	6.3	4.5	3.2	13.7	7.2
93.1	91.1	33.4	106.4	80.1	113.2	84.2	32	47.4	46.4	67.8	39.1
93.3	91.4	99.6	106.6	80.2	113.2	84	31	47.7	46.2	67.9	39.1
79.2	72	72.8	97.2	70.6	–	91	42	36.4	42.4	55.9	36.3
–	–	–	72	–	–	–	39.7	43.8	50.5	–	37.1

(Continued on next page)

Proteins with an N-terminal tobacco etch virus (TEV) protease cleavable His6 tag were constructed by cloning the existing GFP mutants in pET24 into a modified pET28 vector with N-terminal cleavable tag to add the N-terminal sequence: MGSDKIHHHHHHENLYFQG. In brief, the primers GFP.pMA507-star.For. and GFP.pMA507-star.Rev. were used to PCR amplify the mutated GFP DNA segments; the DNA was gel extracted and cloned into pMA507star by the Gibson ISO assembly method (Gibson et al., 2009). pMA507-star was PCR amplified with the primers PIPE.Vec.For. and PIPE.Vec.Rev. to generate compatible DNA overhangs. Primer sequences used are presented in Table S4.

Protein Expression

Plasmids containing mutant GFP genes were transformed into *Escherichia coli* BL21-DE3 expression cells (New England Biolabs). 10-ml starter cultures were grown with overnight shaking at 37°C in LB media containing appropriate antibiotics. The starter culture was used to inoculate 1 l of terrific broth medium supplemented with 20 ml 50× 5052 auto-induction sugars (Studier, 2005) and appropriate antibiotics. Cultures were grown for 4 hr at 37°C. The temperature was then reduced to 30°C, and cultures were allowed to grow for approximately 20 hr. After growth, the cultures were centrifuged at 5,000 × *g* for 30 min at 4°C. Harvested cell paste was stored at −80°C until purification.

Protein Purification

Cell paste was thawed at room temperature in a lysis buffer of 20 mM Tris (pH 8.0), 200 mM NaCl, 10 mM $MgCl_2$, 30 mM Imidazole, 400 μg/ml lysozyme, 10 μg/ml DNAse, and 1 mM AEBSF (4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride). Once the pellet was thawed, cells were lysed via sonication. Lysed cells were incubated at room temperature for 15 min prior to centrifugation to remove all insoluble material, and lysates were clarified at 25,000 × *g* for 30 min at 4°C. The soluble lysate fraction was applied to a 5 ml Ni-nitrilotriacetic acid (IMAC) column, rinsed with 10 column volumes of wash buffer consisting of 20 mM Tris (pH 8.0), 250 mM NaCl, and 30 mM imidazole. The protein was eluted from the column with wash buffer containing 250 mM imidazole. Elution fractions were pooled and then concentrated until the final volume was approximately 1 ml. For the disulfide dimers, the protein was exchanged into a buffer consisting of 20 mM Tris (pH 9.0) and 100 mM NaCl. Cysteines where then oxidized to form dimers by the addition of 10 ml of dimerization buffer (20 mM Tris [pH 9.0], 100 mM NaCl, 5 mM $CuSO_4$). This oxidation reaction was incubated at room temperature for 15 min before being quenched by the addition of 50 mM EDTA. To separate newly formed dimers from remaining monomers, the protein was dialyzed overnight at 4°C into anion exchange buffer (10 mM Tris [pH 9.5], 1 mM EDTA). The protein was applied to an anion exchange column and then eluted via a salt gradient of 0–1 M NaCl in anion exchange buffer. The major peak for each cysteine mutant was assessed for

Table 3. *Continued*

4W7X	4W75	4W76	4W77	4W7A	4W7C	4W7D	4W7E	4W7F	4W7R
0.9789	1.0717	0.9792	0.9789	0.9792	0.9795	0.9792	0.9792	0.9789	0.9789
66.77–2.8 (2.9–2.8)	69.13–3.47 (3.597–3.473)	60.5–2.345 (2.429–2.345)	60.79–3.1 (3.211–3.1)	96.28–3.603 (3.731–3.603)	96.15–2.5 (2.59–2.5)	66.57–1.799 (1.863–1.799)	67.92–2.592 (2.685–2.592)	48.76–2.9 (3.004–2.9)	92.07–1.799 (1.863–1.799)
P 1 21 1	P 21 21 21	P 21 21 21	P 21 21 21	P 21 21 21	C 1 2 1	P 21 21 21	P 41 21 2	C 2 2 21	P 1 21 1
66.84 70.47 116.78 90 92.56 90	59.86 83.93 121.88 90 90 90	64.16 66.84 121 90 90 90	62.12 68.29 121.58 90 90 90	120.83 121.33 192.56 90 90 90	205.49 69.42 102.81 90 110.73 90	56.42 82.05 113.87 90 90 90	96.05 96.05 69.96 90 90 90	68.2 69.75 82.57 90 90 90	62.67 87.19 92.07 90 90.01 90
183,888 (18,230)	104,617 (8,741)	145,381 (13,147)	63,735 (6,380)	220,621 (21,172)	321,515 (31,440)	326,510 (29,374)	134,951 (13,211)	23,617 (1523)	306,630 (30,073)
26,887 (2,649)	8,254 (755)	22,306 (2,125)	9,841 (949)	33,244 (3,177)	46,757 (4,596)	49,634 (4,736)	10,580 (999)	4,558 (378)	90,790 (8,888)
6.8 (6.9)	12.7 (11.6)	6.5 (6.2)	6.5 (6.7)	6.6 (6.7)	6.9 (6.8)	6.6 (6.2)	12.8 (13.2)	5.4 (4.0)	3.4 (3.4)
99.66 (99.62)	98.78 (94.83)	99.23 (96.33)	99.87 (99.79)	99.40 (95.49)	99.06 (97.93)	99.57 (96.26)	99.68 (97.18)	99.52 (99.55)	98.83 (97.73)
8.2 (1.6)	13.6 (1.7)	10.1 (1.8)	9.9 (2.6)	13.9 (2.3)	15.6 (1.9)	8.8 (1.0)	19.8 (2.0)	8.0 (4.0)	6.4 (1.2)
57.1	125.6	47.9	69.4	112.5	63.2	25.9	58.9	82.4	23.5
0.190 (1.297)	0.150 (1.419)	0.110 (1.015)	0.151 (0.748)	0.122 (0.799)	0.073 (1.03)	0.138 (1.868)	0.113 (1.67)	0.178 (0.405)	0.122 (1.021)
0.205	0.156	0.119	0.165	0.133	0.079	0.15	0.118	0.196	0.145
0.991 (0.616)	0.999 (0.944)	0.998 (0.924)	0.996 (0.814)	0.998 (0.804)	0.999 (0.917)	0.997 (0.451)	0.999 (0.792)	0.98 (0.826)	0.994 (0.71)
0.998 (0.873)	1 (0.985)	1 (0.98)	0.999 (0.947)	0.999 (0.944)	1 (0.978)	0.999 (0.789)	1 (0.94)	0.995 (0.951)	0.998 (0.911)
0.217 (0.317)	0.301 (0.444)	0.233 (0.424)	0.217 (0.260)	0.278 (0.336)	0.226 (0.408)	0.179 (0.317)	0.207 (0.3534)	0.264 (0.404)	0.223 (0.376)
0.269 (0.386)	0.345 (0.377)	0.288 (0.459)	0.291 (0.377)	0.302 (0.337)	0.254 (0.428)	0.221 (0.342)	0.262 (0.441)	0.332 (0.439)	0.253 (0.425)
7,089	3,181	3,639	3,474	7,085	7,028	4,014	1,820	1,726	7,625
7,001	3,180	3,588	3,473	6,994	6,938	3,603	1,766	1,703	7,166
88	1	45	1	91	90	103	28	23	146
0	0	6	0	0	0	309	26	0	313
882	396	452	432	881	873	224	222	215	677
0.009	0.004	0.011	0.01	0.011	0.013	0.01	0.012	0.01	0.012
0.93	0.85	1.28	1.4	1.22	1.33	1.26	1.46	1.37	1.27
97	98	99	94	96	99	97	97	97	99
0.23	0	0	0.24	0.35	0	0	0	0	0
11.3	4.1	9.1	12.4	8.0	15.0	4.0	7.1	10.9	6.0
50.7	161.6	64.7	65.2	117.2	100.1	21.3	52.7	74.7	33.5
50.8	161.6	64.8	65.2	118.2	100.3	30.4	52.8	75	33.2
43.3	196.4	61.1	64.9	42.2	83.9	39.9	53.4	51	37.1
–	–	56.8	–	–	–	39	46.2	–	38

$CC_{1/2}$, correlation coefficient between intensities of crystallographic random half-datasets; CC*, correlation coefficient of the full dataset derived from $CC_{1/2}$.

dimer purity by non-reducing SDS-PAGE. Fractions of homogeneous dimers were pooled, buffer exchanged into GFP crystallization buffer (10 mM Tris, 100 mM NaCl), then concentrated to 20 mg/ml. Aliquots of protein were flash-frozen in liquid nitrogen and stored at −80°C for subsequent crystal trials.

Metal-mediated mutants were purified using the same method, up to the IMAC purification, where the hexahistidine tag was cleaved off with TEV protease overnight at 4°C in TEV cleavage buffer (10 mM Tris [pH 8.0], 100 mM NaCl, 5 mM DTT, 1 mM EDTA). Cleaved protein was then subject to a second IMAC step to remove the TEV protease, cleaved histidine tag, and any uncleaved protein. All unbound protein was pooled, buffer exchanged into crystallization buffer, concentrated to 40 mg/ml, flash-frozen, and stored at −80°C for future crystal trials.

Co-expression with Target Proteins

The STARD9-10/11 construct consisted of the N-terminal TEV protease cleavable His6 tag (MGSDKIHHHHHHENLYFQG) followed by the (10–11) hairpin sequence, DLPDDHYLSTQTILSKDLNEKRDHMVLLEYVTAAGITDAS, with the "DAS" serving as a linker between the hairpin and target protein as previously described (Nguyen et al., 2014). Only the first 391 amino acids (Met1–Asn391) corresponding to the putative motor domain of the protein were used in this construct.

For the prospective designed protein construct, the GFP (10–11) hairpin was inserted into a presumptive loop between Ser135 and Thr136 of the native 271-amino-acid protein. This construct features a non-cleavable C-terminal His6 tag, and as such was not used for the metal-mediated experiments.

The expression and purification methods for the co-expressed GFP(1–9) and crystallization targets with the (10–11) hairpin were essentially the same as for the GFPs alone. After size-exclusion chromatography, the fractions with approximate 1:1 molar ratio of GFP(1–9) and target protein (visualized by SDS-PAGE) were used for the crystallography experiments.

Crystallization

The GFP oligomers were crystallized using hanging-drop vapor diffusion. Initial experiments were carried out at the UCLA crystallization facility using commercial sparse matrix screens in a 96-well format. All initial screening trays were set using a Mosquito liquid handling device (TPP LabTech). Limited optimizations were performed manually in some cases using 24-well Linbro plates. Each disulfide dimer was screened initially with four commercial sparse matrix screens JCSG+ (Qiagen), SaltRx (Hampton Research), Crystal Screen I + II (Hampton Research), and Wizard I + II (EmeraldBio). Metal-mediated mutations were screened with JCSG+ and Wizard only. The final concentration of protein in all crystallization experiments was 20 mg/ml. Metal-mediated mutants were mixed with the metal ions (Ni^{2+}, Zn^{2+}, or Cu^{2+}, in three separate

screens) immediately before setting crystal trays, at a final concentration of 20 mg/ml protein and 2 mM metal ion salts. Trays were set at room temperature and checked periodically over 30 days. Single crystals were mounted with CrystalCat HT Cryoloops (Hampton Research), cryoprotected as needed, flash-frozen with liquid nitrogen, and screened for diffraction. All diffracting crystals were stored for later data collection. All diffraction data were collected at 100 K at APS-NECAT beamline 24-ID-C on a DECTRIS-PILATUS 6M detector. The crystallization and cryoprotectant conditions are reported in Table S5.

Structure Determination

Datasets from individual crystals were indexed, integrated, and scaled using XDS/XSCALE (Kabsch, 2010), with the resolution limit selected to balance completeness, calculated I/σ, R_{sym}, and $CC_{1/2}$ of the highest-resolution shell with emphasis on I/σ values of >1.5 and $CC_{1/2}$ values of >0.9. Structures were solved by molecular replacement using the program Phaser (McCoy et al., 2007), with the superfolder GFP (Pédelacq et al., 2006) protein (PDB: 2B3P) as the search model. To accelerate the model building and refinement, molecular replacement solutions were initially refined with the PDB_REDO server (Joosten et al., 2011). Final iterative rounds of model building and refinement were carried out using Coot (Emsley et al., 2010) and PHENIX (Adams et al., 2010) with TLS refinement (Painter and Merritt, 2006). Structures were validated with PROCHECK (Laskowski et al., 1993), ERRAT (Colovos and Yeates, 1993), MolProbity (Davis et al., 2007), and VERIFY3D (Luthy et al., 1992). Atomic coordinates and structure factors for all 33 structures were deposited in the PDB. Figures depicting the structures were made with PyMOL (Schrödinger). Data collection and refinement statistics are given in Table 3.

Structure Comparison Procedure

To compare multiple observed instances of the same disulfide-bonded dimer, one structure was first chosen as the reference. Then one chain of a subsequent dimer was aligned to chain A of the reference dimer, and the transformation required for overlapping those two chains was applied to the second chain. Both possible assignments to chain A versus chain B were tested for each dimer, and the best match was retained for comparison. These optimal chain assignments do not necessarily correspond to chain assignments in the deposited PDB files.

ACCESSION NUMBERS

Atomic coordinates and structure factors for all 33 structures in this study were deposited, under accession numbers PDB: 4W69, 4W6A, 4W6B, 4W6C, 4W6D, 4W6F, 4W6G, 4W6H, 4W6I, 4W6J, 4W6K, 4W6L, 4W6M, 4W6N, 4W6O, 4W6P, 4W6R, 4W6S, 4W6T, 4W6U, 4W72, 4W73, 4W74, 4W7X, 4W75, 4W76, 4W77, 4W7A, 4W7C, 4W7D, 4W7E, 4W7F, and 4W7R.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five tables and can be found with this article online at http://dx.doi.org/10.1016/j.str.2015.07.008.

ACKNOWLEDGMENTS

This work was supported by NIH grant P01 GM098177 (to T.C.T.). D.J.L. was supported by Ruth L. Kirschstein National Research Service Award T32GM007185. The authors thank Michael Sawaya, Duilio Cascio, and Michael Thompson for X-ray data collection at APS beamline 24-ID-C. We thank Michael Collazo for help with the crystallization trials, and Dan McNamara for help with structure determinations. The UCLA macromolecular structure facilities are supported by the BER program of the DOE Office of Science (award DE-FC02-02ER63421). We thank David Baker and Fabio Parmeggiani for providing the designed protein as a target for fusion-based crystallization experiments. We thank the staff of the NECAT synchrotron beamline, including Jon Schuermann, Igor Kourinov, and Malcolm Capel, and for helpful discussions. X-ray data collection was supported by DOE Grant DE-FC02-02ER63421 and the NECAT beamlines of the Advanced Photon Source, which are supported by NIH Grant RR-15301(NCRR). Use of the Advanced Photon Source is supported by the DOE, Office of Basic Energy Sciences, under Contract DE-AC02-06CH11357.

Received: February 25, 2015
Revised: June 8, 2015
Accepted: July 7, 2015
Published: August 13, 2015

REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Banatao, D.R., Cascio, D., Crowley, C.S., Fleissner, M.R., Tienson, H.L., and Yeates, T.O. (2006). An approach to crystallizing proteins by synthetic symmetrization. Proc. Natl. Acad. Sci. USA *103*, 16230–16235.

Cabantous, S., Terwilliger, T.C., and Waldo, G.S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nat. Biotechnol. *23*, 102–107.

Cabantous, S., Nguyen, H.B., Pedelacq, J.D., Koraïchi, F., Chaudhary, A., Ganguly, K., Lockard, M.A., Favre, G., Terwilliger, T.C., and Waldo, G.S. (2013). A new protein-protein interaction sensor based on tripartite split-GFP association. Sci. Rep. *3*, 2854.

Carugo, O., and Carugo, K.D. (2005). When X-rays modify the protein structure: radiation damage at work. Trends Biochem. Sci. *30*, 213–219.

Chruszcz, M., Potrzebowski, W., Zimmerman, M.D., Grabowski, M., Zheng, H., Lasota, P., and Minor, W. (2008). Analysis of solvent content and oligomeric states in protein crystals—does symmetry matter? Protein Sci. *17*, 623–632.

Colovos, C.Y., and Yeates, T.O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. *2*, 1511–1519.

Conrado, R.J., Varner, J.D., and DeLisa, M.P. (2008). Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. Curr. Opin. Biotechnol. *19*, 492–499.

Corsini, L., Hothorn, M., Scheffzek, K., Sattler, M., and Stier, G. (2008). Thioredoxin as a fusion tag for carrier-driven crystallization. Protein Sci. *17*, 2070–2079.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S., and Richardson, D.C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. *35*, W375–W383.

Der, B.S., Machius, M., Miley, M.J., Mills, J.L., Szyperski, T., and Kuhlman, B. (2012). Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. J. Am. Chem. Soc. *134*, 375–385.

Derewenda, Z.S., and Vekilov, P.G. (2006). Entropy and surface engineering in protein crystallization. Acta Crystallogr. D Biol. Crystallogr. *62*, 116–124.

Dueber, J.E., Wu, G.C., Malmirchegini, G.R., Moon, T.S., Petzold, C.J., Ullal, A.V., Prather, K.L., and Keasling, J.D. (2009). Synthetic protein scaffolds provide modular control over metabolic flux. Nat. Biotechnol. *27*, 753–759.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Forse, G.J., Ram, N., Banatao, D.R., Cascio, D., Sawaya, M.R., Klock, H.E., Lesley, S.A., and Yeates, T.O. (2011). Synthetic symmetrization in the crystallization and structure determination of CelA from *Thermotoga maritima*. Protein Sci. *20*, 168–178.

Gibson, H.W. (1969). Chemistry of formic acid and its simple derivatives. Chem. Rev. *69*, 673–692.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Good, M.C., Zalatan, J.G., and Lim, W.A. (2011). Scaffold proteins: hubs for controlling the flow of cellular information. Science *332*, 680–686.

Huang, Y.M., and Bystroff, C. (2009). Complementation and reconstitution of fluorescence from circularly permuted and truncated green fluorescent protein. Biochemistry *48*, 929–940.

Joosten, R.P., Joonsten, K., Cohen, S.X., Vriend, G., and Perrakis, A. (2011). Automatic rebuilding and optimization of crystallographic structures in the Protein Data Bank. Bioinformatics *27*, 3392–3398.

Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Katz, B.A., and Kossiakoff, A. (1986). The crystallographically determined structures of atypical strained disulfides engineered into subtilisin. J. Biol. Chem. *261*, 15480–15485.

Laganowsky, A., Zhao, M., Soriaga, A.B., Sawaya, M.R., Cascio, D., and Yeates, T.O. (2011). An approach to crystallizing proteins by metal-mediated synthetic symmetrization. Protein Sci. *20*, 1876–1890.

Lai, Y.T., King, N.P., and Yeates, T.O. (2012). Principles for designing ordered protein assemblies. Trends Cell Biol. *22*, 653–661.

Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993). Procheck—a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.

Lee, H., DeLoache, W.C., and Dueber, J.E. (2012). Spatial organization of enzymes for metabolic engineering. Metab. Eng. *14*, 242–251.

Luthy, R., Bowie, J.U., and Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. Nature *356*, 83–85.

March, J.C., Rao, G., and Bentley, W.E. (2003). Biotechnological applications of green fluorescent protein. Appl. Microbiol. Biotechnol. *62*, 303–315.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

Moon, A.F., Mueller, G.A., Zhong, X., and Pedersen, L.C. (2010). A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. Protein Sci. *19*, 901–913.

Nguyen, H.B., Hung, L.W., Yeates, T.O., Terwilliger, T.C., and Waldo, G.S. (2014). Split green fluorescent protein as a modular binding partner for protein crystallization. Acta Crystallogr. D Biol. Crystallogr. *69*, 2513–2523.

Painter, J.M., and Merritt, E.A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr. D Biol. Crystallogr. *62*, 439–450.

Pédelacq, J.D., Tran, T., Terwilliger, T.C., and Waldo, G.S. (2006). Engineering and characterization of a superfolder green fluorescent protein. Nat. Biotechnol. *24*, 79–88.

Salgado, E.N., Lewis, R.A., Faraone-Mennella, J., and Tezcan, F.A. (2008). Metal-mediated self-assembly of protein superstructures: influence of secondary interactions on protein oligomerization and aggregation. J. Am. Chem. Soc. *130*, 6082–6084.

Salgado, E.N., Ambroggio, X.I., Brodin, J.D., Lewis, R.A., Kuhlman, B., and Tezcan, F.A. (2010). Metal templated design of protein interfaces. Proc. Natl. Acad. Sci. USA *107*, 1827–1832.

Stacy, R., Begley, D.W., Phan, I., Staker, B.L., Van Voorhis, W.C., Varani, G., Buchko, G.W., Stewart, L.J., and Myler, P.J. (2011). Structural genomics of infectious disease drug targets: the SSGCID. Acta Crystallogr. F Struct. Biol. Cryst. Commun. *67*, 979–984.

Studier, F.W. (2005). Protein production by auto-induction in high-density shaking cultures. Protein Expr. Purif. *41*, 207–234.

Sundstrom, M., Norin, M., and Edwards, A. (2006). Structural Genomics and High Throughput Structural Biology (CRC Press).

Suzuki, N., Hiraki, M., Yamada, Y., Matsugaki, N., Igarashi, N., Kato, R., Dikic, I., Drew, D., Iwata, S., Wakatsuki, S., and Kawasaki, M. (2010). Crystallization of small proteins assisted by green fluorescent protein. Acta Crystallogr. D Biol. Crystallogr. *66*, 1059–1066.

Torres, J.Z., Summers, M.K., Peterson, D., Brauer, M.J., Lee, J., Senese, S., Gholkar, A.A., Lo, Y.C., Lei, X., Jung, K., Anderson, D.C., et al. (2011). The STARD9/Kif16a kinesin associates with mitotic microtubules and regulates spindle pole assembly. Cell *147*, 1309–1323.

Weik, M., Ravelli, R.B., Kryger, G., McSweeney, S., Raves, M.L., Harel, M., Gros, P., Silman, I., Kroon, J., and Sussman, J.L. (2000). Specific chemical and structural damage to proteins produced by synchrotron radiation. Proc. Natl. Acad. Sci. USA *97*, 623–628.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235–242.

Zeke, A., Lukac, M., Lim, W.A., and Remenyi, A. (2009). Scaffolds: interaction platforms for cellular signalling circuits. Trends Cell Biol. *19*, 364–374.

Zou, Y., Weiss, W.I., and Kobilka, B.K. (2012). N-Terminal T4 lysozyme fusion facilitates crystallization of a G protein coupled receptor. PLoS One *7*, e46039.

