UCLA
UCLA Previously Published Works

Title

A Suite of Engineered GFP Molecules for Oligomeric Scaffolding

Permalink

https://escholarship.org/uc/item/82h5w7nw

Journal

Structure, 23(9)

ISSN

1359-0278

Authors

Leibly, David J
Arbing, Mark A
Pashkov, Inna
et al.

Publication Date

2015-09-01

DOI

10.1016/j.str.2015.07.008

Peer reviewed

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

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 green fluorescent protein (GFP) as a scaffold, and engineered many alternate 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 in order 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 towards 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 & Vekilov, 2006; Salgado et al., 2008; Forse et al., 2011; Corsini et al., 2008; Moon et al., 2010; Zou & Kobilka, 2012) including fusion to engineered green fluorescent proteins (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). Those 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 that obstacle. The essential idea is to apply the protein engineering work (i.e. to introduce varied forms of synthetic symmetrization) to a model protein that can subsequently serve as a general carrier or scaffold for attaching otherwise arbitrary proteins. In that 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 & 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 demonstrated already 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 and 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 & Bentley, 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 beta-strands of the eleven strands composing the GFP beta barrel. Using circular permutants of a full-length GFP containing mutations developed for the split-form of GFP (Cabantous et al. 2005), Bystroff and co-workers created additional split-GFP pairs (with other tagging or "left-out" strands such as beta strand 7) (Huang & Bystroff, 2009). The partial core can then be complemented by addition of another protein that has been engineered to carry the missing GFP beta strand(s), either as a terminal fusion or as a loop insertion. Once complementation occurs, the full beta 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 nine, 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 10–11 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 two-chain crossing between the reconstituted GFP domain and the target protein structure (Fig. 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.

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 before (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, their surface location, and their distance from strands 10–11 in order to limit interference with complementation when ultimately expressed in the split form (Fig. 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 (Pedelacq 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 in the background of the complete GFP (1–11) construct. Full-length GFP constructs bearing the single engineered cysteine residue were therefore expressed, purified, and then oxidized to form homogenous dimers (Figure 2). For all five of the cysteine sites chosen, pure dimers could be obtained in good yield with ∼20–50mg of protein obtained from 2L of auto-induction media.

With the exception of Q175C, crystals grew readily in one to seven days. Depending on the mutant, diffraction quality crystals grew in as few as one condition for K126C or in more than twenty 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 where 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.

In all, we were able to characterize 20 distinctly different crystal forms of the GFP disulfide dimers and solve their structures (Tables 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 (Table 2, Table S1) (Katz & 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 & 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 together all dimers that were observed for a given point mutation (Fig. 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 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 depositions 4W6B, 4W6C, 4W6D and 4W6F), two in each of the space groups $P2_12_12_1$ and $P3_22_11$. Of these, 4W6C was the most symmetric (175.6°) whil e 4W6F was the least (144.3°). 4W6C, 4W6D and 4W6F were most similar with a maximum variation of 33.3° while 4W6B varied by a rotation of up to 140.4° when over laid on the others (Table 2, Figure 4B). Two of the structures (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 (4W6R) and one in P2$_1$ 2$_1$ 2$_1$ (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 4W6R to 3.47Å; 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 4W6P structure also contained four dimers in the asymmetric unit of P2$_1$2$_1$2$_1$. These dimers are less symmetric than those observed in the P1 form (average internal angle between subunits of ~143°). In comparison to 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 effe ctively 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 (4W72, 4W73 and 4W7X) and one structure with metal-mediated contacts only (4W74, discussed subsequently). The three disulfide dimers feature an average rotation angle between subunits of 165°, wi th a variation up to 12° (Figure 4E, Table 2). Interestingly, 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 (4W6L, 4W6M and 4W6O) are either perfectly symmetric with the two subunits related by crystal symmetry (4W6L and 4W6O), or very nearly symmetric (4W6M, 179° rotation). 4W6J and 4W6N feature simila rly asymmetric dimers (average internal angle of 149°), and 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 1.7Å. 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 (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, 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 (4W69). 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 (4W6G, 4W6H and 4W6I). 4W6I was the most symmetric dimer (171°) while 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 towards 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 alpha helical segments rather than a beta sheet conformation as in GFP.

Design and structure of metal-mediated GFP oligomers

In additional 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 of Tezcan et al. and Kuhlman et al. (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, where 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 alpha 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 on the approach could be applied to GFP, which consists mainly of a single beta-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 beta 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 if 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 (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 & 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 (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 4W76, 4W77, 4W7A and 4W7C are close to being symmetric (average angle of 173.4°) with 4W75 being more asymmetric at a 152° inter-subunit rotat ion. 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 4W7E, Asp19 and His21 of one chain and Gln184 of the symmetry mate chelate the copper ion. This mutant crystallized in the presence of imidazole, leading to one imidazole molecule also being bound to the copper ion. Structure 4W7D features two different copper-mediate 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 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. 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 4W6T structure 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, 4W7F and 4W7R. 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 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 oriented 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-GFP1–9 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 GFP1–9 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 and E124H/K126H gave robust complementation. We were able to obtain crystals of the STARD9–10/11 and GFP1–9 (D21H/K26C) complex after approximately three months (Fig 6A).

However, these crystals are small (∼20μM in the largest dimension) and have not produced well-ordered diffraction to date; optimization efforts are underway.

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 seven months, triangular plate crystals (∼50–75μM) (Fig 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 underway.

DISCUSSION

The structural results presented here characterize a suite of engineered GFP molecules comprising a wide range of oligomeric forms, most of which appear highly amenable to crystallization on their own. We obtained 20 new crystal forms of seven disulfide-bonded dimers, plus thirteen 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, though individual outliers were also obtained in both cases. The oligomeric GFP molecules designed here have not been used yet to successfully crystallize a target protein that was otherwise recalcitrant to crystallization. Which of the GFP constructs might ultimately be most useful in that 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 (6) distinct crystal forms. Among the metal mediated designs, the E115H/T118H and the 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 GFP1–9 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 to 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 (Fig 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 (Fig. 7 C, D). In order 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 (Fig. 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 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 that purpose. As an example, a novel tetrameric form of GFP could be based on the structure of the D117C mutant 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.

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) 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.

Proteins with an N-terminal 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. Briefly, 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 DG, 2009). pMA507star 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 *E. coli* BL21-DE3 expression cells (New England Biolabs). 10ml starter cultures were grown with overnight shaking at 37°C in LB media containing appropriate antibiotics. The starter culture was used to inoculate 1L of terrific broth media supplemented with 20ml 50× 5052 auto-induction sugars (Studier, 2005) and appropriate antibiotics. Cultures were grown for 4 hours at 37°C. The temperature was then reduced to 30°C, and cultures were allowed to grow for approximately 20 hours. After growth, the cultures were centrifuged at $5000 \times g$ for 30 minutes 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 20mM Tris pH 8.0, 200mM NaCl, 10mM $MgCl_2$, 30mM Imidazole, 400ug/ml lysozyme, 10ug/ml DNAse and 1mM AEBSF. 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 xg for 30 min at 4°C. The soluble lysate fraction was applied to a 5ml Ni-NTA (IMAC) column, rinsed with 10 column volumes of wash buffer consisting of 20mM Tris pH 8.0, 250 mM NaCl, 30mM imidazole. The protein was eluted from the column with wash buffer containing 250mM imidazole. Elution fractions were pooled and then concentrated until the final volume was approximately 1ml. For the disulfide dimers, the protein was exchanged into a buffer consisting of 20mM Tris pH 9.0, 100mM NaCl. Cysteines where then oxidized to form dimers by the addition of 10ml of dimerization buffer (20mM Tris pH 9.0, 100mM NaCl, 5mM $CuSO_4$). This oxidation reaction was incubated at room temperature for 15 minutes before being quenched by the addition of 50mM EDTA. To separate newly formed dimers from remaining monomers, the protein was dialyzed overnight at 4 °C into anion exchange buffer (10mM Tris pH 9.5, 1mM EDTA). The protein was applied to an anion exchange column and then

eluted via a salt gradient of 0–1M NaCl in anion exchange buffer. The major peak for each cysteine mutant was assessed for dimer purity by non-reducing SDS-PAGE. Fractions of homogeneous dimers were pooled, buffer exchanged into GFP crystallization buffer (10mM tries, 100mM NaCl), then concentrated to 20mg/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 (10mM Tris pH 8.0, 100mM NaCl, 5mM DTT, 1mM 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 40mg/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, DLPDDHYLSTQTILSKDLNEKRDHMVLLEYVTAAGIT**DAS**, 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 long 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 GFP1–9 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 ration of GFP1–9 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 20mg/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, Aliso Viejo, CA), cryoprotected as needed, flash frozen with liquid nitrogen, and then screened for diffraction. All diffracting crystals were stored for later data collection. All diffraction data were collected at 100K 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

Data sets 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 CC1/2 of the highest resolution shell with emphasis on I/σ values of >1.5 and CC1/2 values > 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), PHENIX (Adams et al., 2010) with TLS refinement (Painter et al., 2006). Structures were validated with PROCHECK (Laskowski et al., 1993), ERRAT (Colovos & 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 with codes; 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. Figures depicted the structures were made with PyMOL (Schrödinger, LLC). 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 vs 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Abbreviations

GFP green fluorescent protein

PDB Protein Data Bank

References

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, 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 Crystallogr D Biol Crystallogr. 2010; 66:213–221. [PubMed: 20124702]

Banatao DR, Cascio D, Crowley CS, Fleissner MR, Tienson HL, Yeates TO. An approach to crystallizing proteins by synthetic symmetrization. Proc Natl Acad Sci. 2006; 103:16230–16235. [PubMed: 17050682]

Cabantous S, Terwilliger TC, Waldo GS. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nature Biotechnology. 2005; 23:102–107.

Cabantous S, Nguyen HB, Pedelacq JD, Koraïchi F, Chaudhary A, Ganguly K, Lockard MA, Favre G, Terwilliger TC, Waldo GS. A new protein-protein interaction sensor based on tripartite split-GFP association. Sci Rep. 2013; 3:2854. [PubMed: 24092409]

Carugo O, Carugo KD. When X-rays modify the protein structure: radiation damage at work. Trends Biochem Sci. 2005; 30:213–219. [PubMed: 15817398]

Chruszcz M, Potrzebowski W, Zimmerman MD, Grabowski M, Zheng H, Lasota P, Minor W. Analysis of solvent content and oligomeric states in protein crystals—does symmetry matter? Protein Sci. 2008; 17:623–632. [PubMed: 18359856]

Colovos CY, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 1993; 2:1511–1519. [PubMed: 8401235]

Corsini L, Hothorn M, Scheffzek K, Sattler M, Stier G. Thioredoxin as a fusion tag for carrier-driven crystallization. Protein Sci. 2008; 17:2070–2079. [PubMed: 18780816]

Conrado RJ, Varner JD, DeLisa MP. Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. Curr Opin Biotechnol. 2008; 19:492–499. [PubMed: 18725290]

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucl Acids Res. 2007; 35:W375–W383. [PubMed: 17452350]

Der BS, Machius M, Miley MJ, Mills JL, Szyperski T, Kuhlman B. Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. J Am Chem Soc. 2012; 134:375–385. [PubMed: 22092237]

Derewenda ZS, Vekilov PG. Entropy and surface engineering in protein crystallization. Acta. Cryst. D. 2006; 62:116–124. [PubMed: 16369101]

Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, Ullal AV, Prather KL, Keasling JD. Synthetic protein scaffolds provide modular control over metabolic flux. Nature biotechnology. 2009; 27:753–759.

Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010; 66:486–501. [PubMed: 20383002]

Forse GJ, Ram N, Banatao DR, Cascio D, Sawaya MR, Klock HE, Lesley SA, Yeates TO. Synthetic symmetrization in the crystallization and structure determination of CelA from Thermotoga maritima. Protein Sci. 2011; 20:168–178. [PubMed: 21082721]

Gibson DGYL, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009; 6:343–345. [PubMed: 19363495]

Gibson HW. Chemistry of formic acid and its simple derivatives. Chem Rev. 1969; 69:673–692.

Good MC, Zalatan JG, Lim WA. Scaffold proteins: hubs for controlling the flow of cellular information. Science. 2011; 332:680–686. [PubMed: 21551057]

Huang YM, Bystroff C. Complementation and reconstitution of fluorescence from circularly permuted and truncated green fluorescent protein. Biochemistry. 2009; 48:929–940. [PubMed: 19140681]

Joosten RP, Joonsten K, Cohen SX, Vriend G, Perrakis A. Automatic rebuilding and optimization of crystallographic structures in the Protein Data Bank. Bioinformatics. 2011; 27:3392–3398. [PubMed: 22034521]

Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr. 2010; 66:125–132. [PubMed: 20124692]

Kim C, Basner J, Lee B. Detecting internally symmetric protein structures. BMC Bioinformatics. 2010; 11:303. [PubMed: 20525292]

Katz BA, Kossiakoff A. The crystallographically determined structures of atypical strained disulfides engineered into subtilisin. J Biol Chem. 1986; 261:15480–5. [PubMed: 3096989]

Laganowsky A, Zhao M, Soriaga AB, Sawaya MR, Cascio D, Yeates TO. An approach to crystallizing proteins by metal-mediated synthetic symmetrization. Protein Sci. 2011; 20:1876–1890. [PubMed: 21898649]

Lai YT, King NP, Yeates TO. Principles for designing ordered protein assemblies. Trends Cell Biol. 2012; 22:653–661. [PubMed: 22975357]

Laskowski RA, Macarthur MW, Moss DS, Thornton JM. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. J Appl Crystallogr. 1993; 26:283–291.

Lee H, DeLoache WC, Dueber JE. Spatial organization of enzymes for metabolic engineering. Metab Eng. 2012; 14:242–251. [PubMed: 21946160]

Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature. 1992; 356:83–85. [PubMed: 1538787]

March JC, Rao G, Bentley WE. Biotechnological applications of green fluorescent protein. Appl Microbiol Biotechnol. 2003; 62:303–315. [PubMed: 12768245]

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007; 40:658–674. [PubMed: 19461840]

Moon AF, Mueller GA, Zhong X, Pedersen LC. A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. Protein Sci. 2010; 19:901–913. [PubMed: 20196072]

Nguyen HB, Hung LW, Yeates TO, Terwilliger TC, Waldo GS. Split green fluorescent protein as a modular binding partner for protein crystallization. Acta Crystallogr D Biol Crystallogr. 2014; 69:2513–2523. [PubMed: 24311592]

Painter JM, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr. 2006; 62:439–450. [PubMed: 16552146]

Pedelacq JD, Tran T, Terwilliger TC, Waldo GS. Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol. 2006; 24:79–88. [PubMed: 16369541]

Salgado EN, Ambroggio XI, Brodin JD, Lewis RA, Kuhlman B, Tezcan FA. Metal templated design of protein interfaces. Proc Natl Acad Sci. 2008; 107:1827–32. [PubMed: 20080561]

Salgado EN, Lewis RA, Faraone-Mennella J, Tezcan FA. Metal-Mediated Self-Assembly of Protein Superstructures: Influence of Secondary Interactions on Protein Oligomerization and Aggregation. J Am Chem Soc. 2008; 130:6082–6084. [PubMed: 18422313]

Stacy R, Begley DW, Phan I, Staker BL, Van Voorhis WC, Varani G, Buchko GW, Stewart LJ, Myler PJ. Structural genomics of infectious disease drug targets: the SSGCID. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2011; 67:979–984.

Studier FW. Protein production by auto-induction in high-density shaking cultures. Prot Expres Pur. 2005; 41:207–234.

Sundstrom, M.; Norin, M.; Edwards, A. Structural Genomics and High Throughput Structural Biology. Boca Raton, FL, USA: CRC Press; 2006.

Suzuki N, Hiraki M, Yamada Y, Matsugaki N, Igarashi N, Kato R, Dikic I, Drew D, Iwata S, Wakatsuki S, Kawasaki M. Crystallization of small proteins assisted by green fluorescent protein. Acta Crystallogr D Biol Crystallogr. 2010; 66:1059–1066. [PubMed: 20944239]

Torres JZ, Summers MK, Peterson D, Brauer MJ, Lee J, Senese S, Gholkar AA, Lo YC, Lei X, Jung K, Anderson DC, Davis DP, Belmont L, Jackson PK. The STARD9/Kif16a kinesin associates with mitotic microtubules and regulates spindle pole assembly. Cell. 2011; 147:1309–1323. [PubMed: 22153075]

Weik M, Ravelli RB, Kryger G, McSweeney S, Raves ML, Harel M, Gros P, Silman I, Kroon J, Sussman JL. Specific chemical and structural damage to proteins produced by synchrotron radiation. Proc Natl Acad Sci. 2000; 97:623–628. [PubMed: 10639129]

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, 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 Crystallogr D Biol Crystallogr. 2011; 67:235–242. [PubMed: 21460441]

Zou Y, Weiss WI, Kobilka BK. N-Terminal T4 Lysozyme Fusion Facilitates Crystallization of a G Protein Coupled Receptor. PloS One. 2012; 7:e46039. [PubMed: 23056231]

Zeke A, Lukac M, Lim WA, Remenyi A. Scaffolds: interaction platforms for cellular signalling circuits. Trends Cell Biol. 2009; 19:364–374. [PubMed: 19651513]

Highlights

- A large suite of GFP mutants have 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.
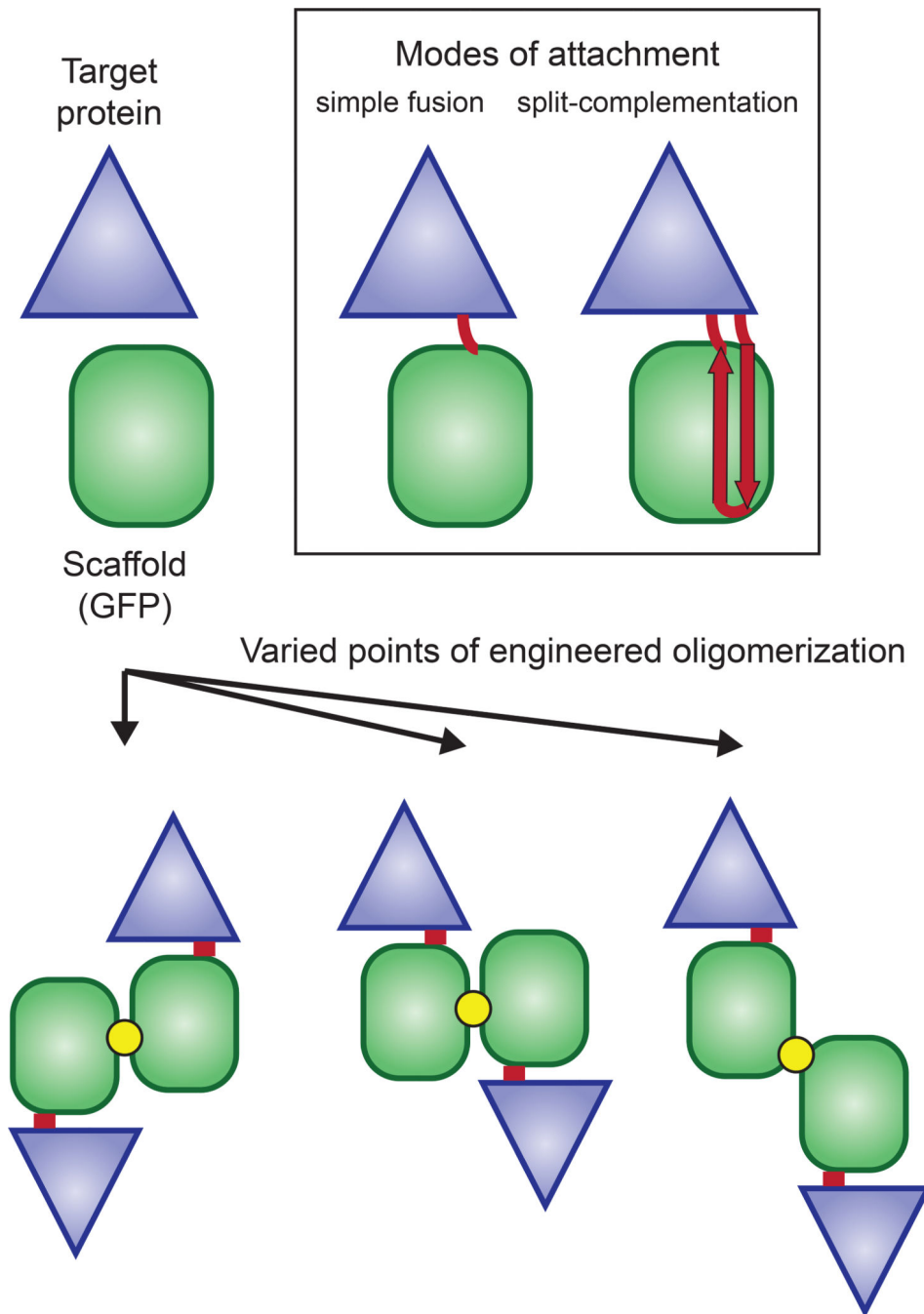
Figure 1. Concept of scaffold-mediated synthetic symmetrization
Here, GFP serves as a scaffold to induce synthetic symmetry. (top) Multiple modes for attaching a target protein to GFP are indicated, including simple fusion and split-form complementation where the target protein is fused to a fragment 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.
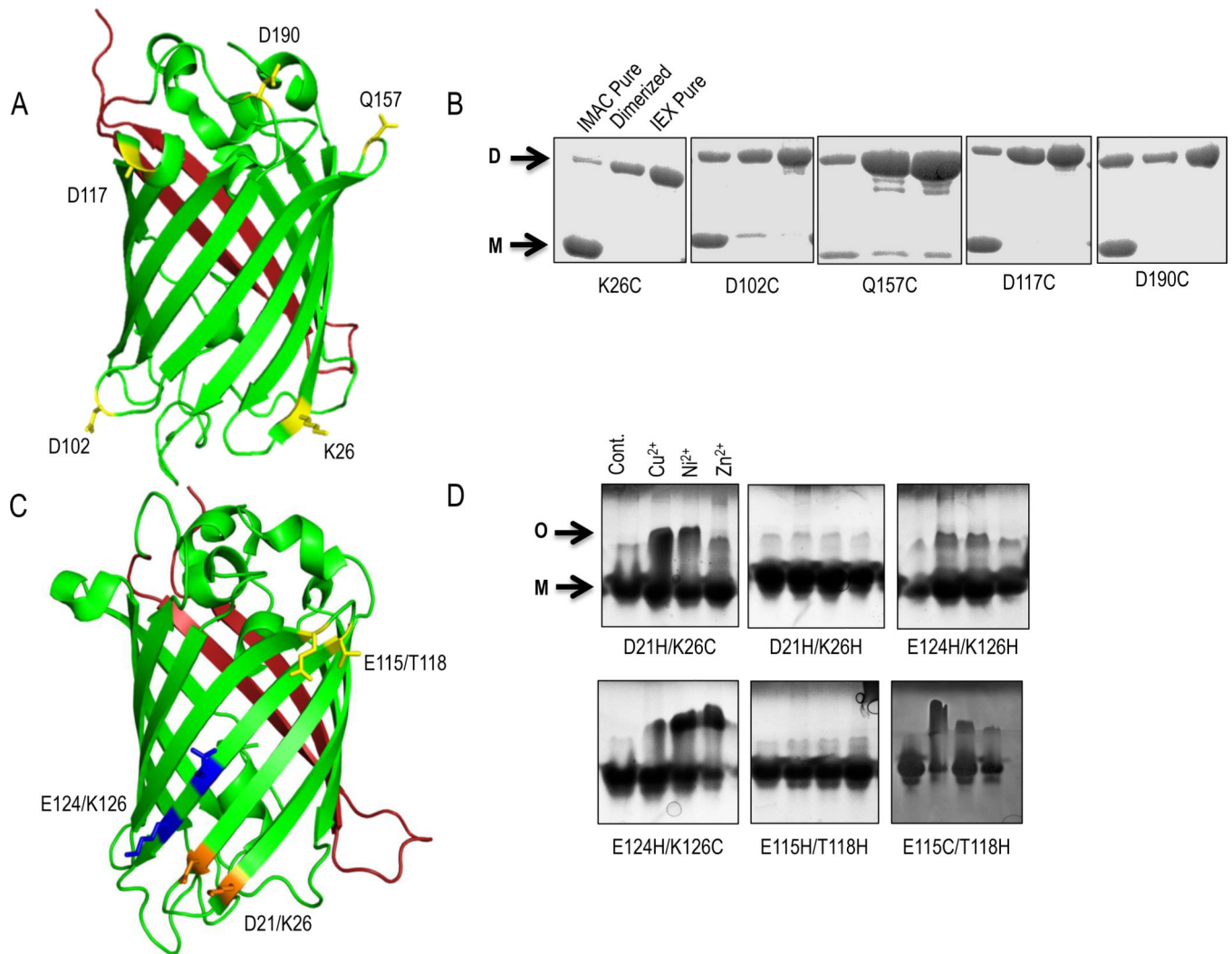
Figure 2. Locations of point mutations introduced on full-length split-GFP to induce oligomerization

A) Locations of the individual point mutations to cysteines on the GFP(1–9) core (green) on the opposite face of the beta-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; each site involves a pair of spatially proximal mutations (indicated). 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.
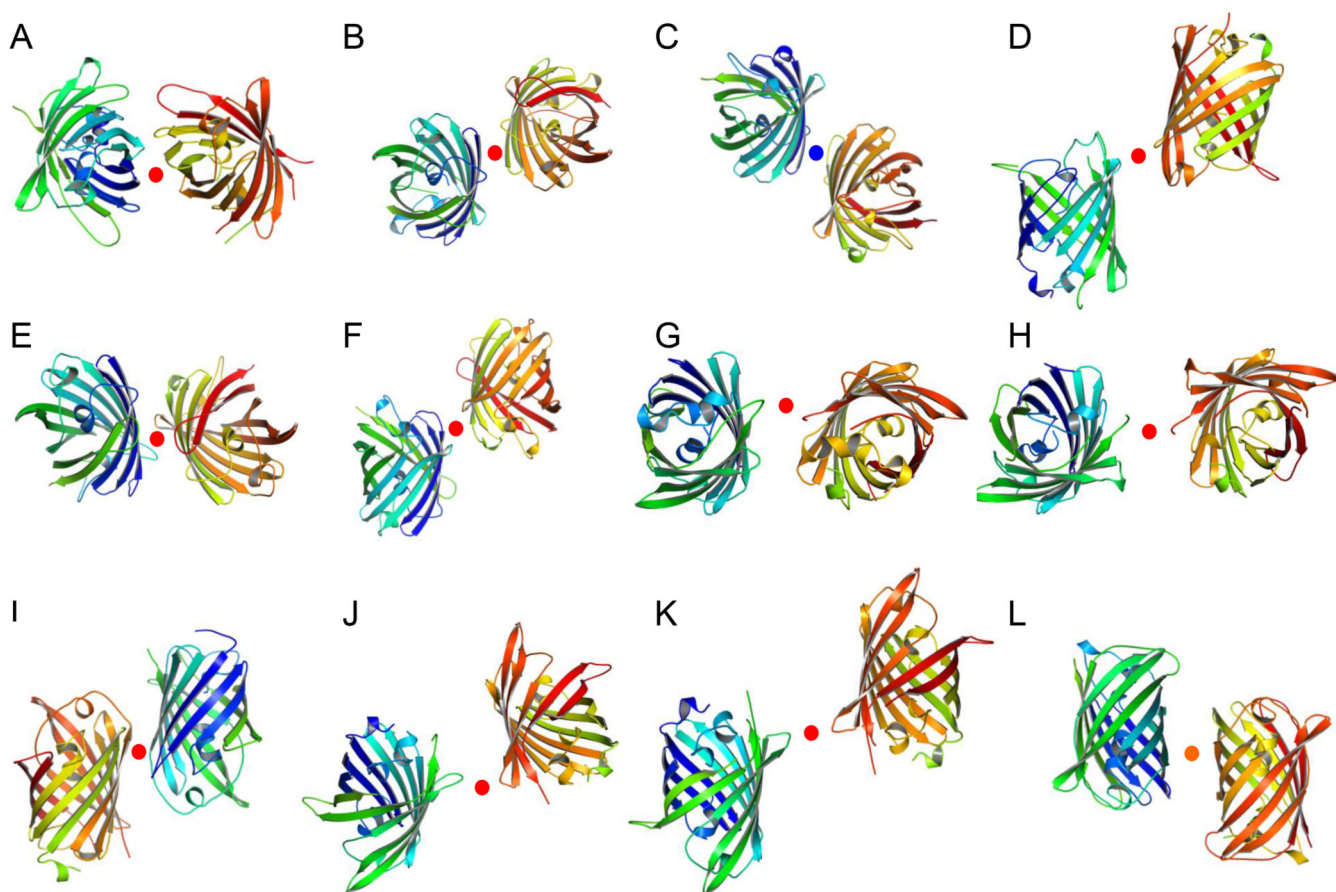
Figure 3. Examples of the GFP dimer observed

The internal rotation axis relating the subunits of each dimer is shown (red dot for disulfide dimers, blue for the mixed dimer and orange for the metal-mediate dimer). For each dimer the rotation axis corresponds to the location of the engineered disulfide bond, or metal-mediated crystal contact. The 12 dimers shown are from structures: A) 4W6B B) 4W6C C) 4W7C D) 4W6R E) 4W7X F) 4W6M G) 4W6G H) 4W6I I) 4W6S J) 4W69 K) 4W6K L) 4W7R. They are representative of the complete set of 43 total dimers visualized in this work.

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 visualiz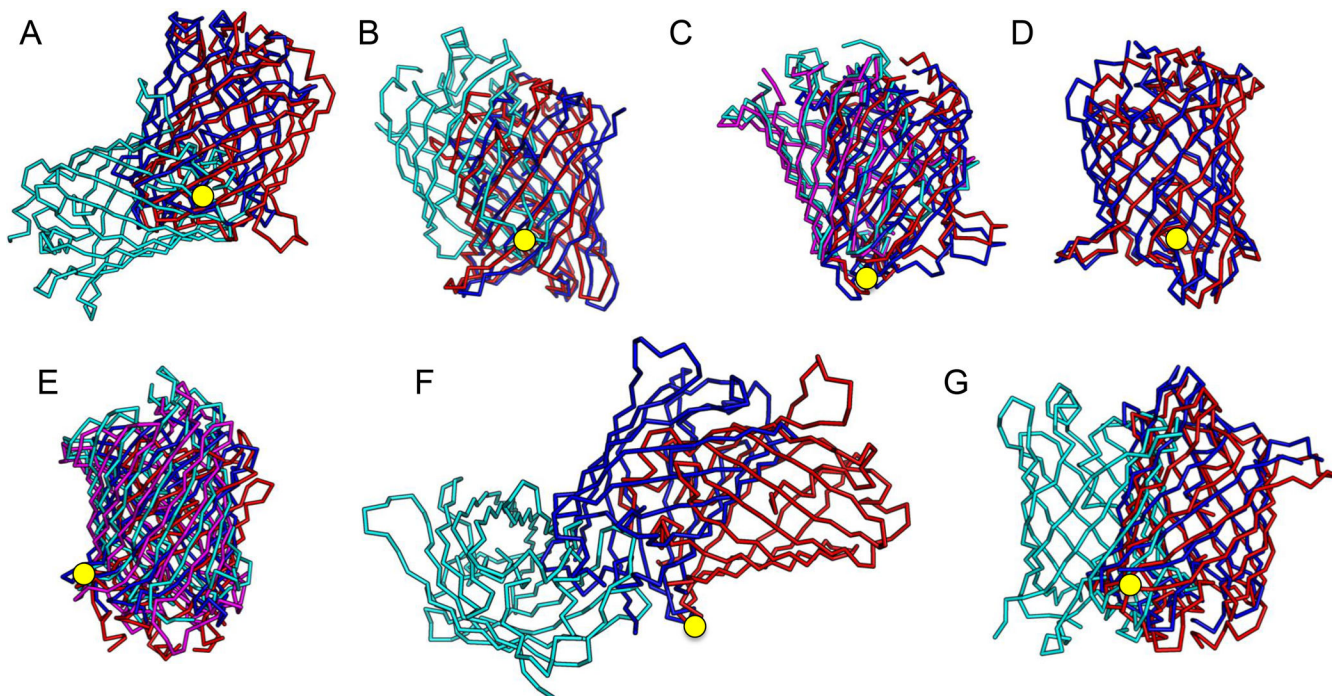e the difference in the orientation of the distinct versions of chain B. Only the chain B backbone traces are depicted. Each panel illustrates the multiple conformations observed for one specific cysteine mutant. The blue and red traces represent the range of orientations the chains adopted. When a single 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. The structures and dimer chains displayed are: A) K26C red: 4W6C, blue: 4W6F, cyan: 4W6B. B) D21H/K26C; red: 4W7A AB dimer, blue: 4W7A CD dimer, cyan: 4W75. C) D102C; red: 4W6P CD dimer, blue: 4W6P FG dimer, cyan: 4W6R AN dimer, magenta: 4W6R KL dimer. D) E115C; red: 4W72, blue: 4W73. E) D117C; red: 4W6O, blue: 4W6K, cyan: 4W6N BF dimer, magenta: 4W6J. F) Q157C; red: 4W69, blue: 4W6A A dimers, cyan: 4W6A B dimer. G) D190C; red: 4W6H, blue: 4W6I, cyan: 4W6G.
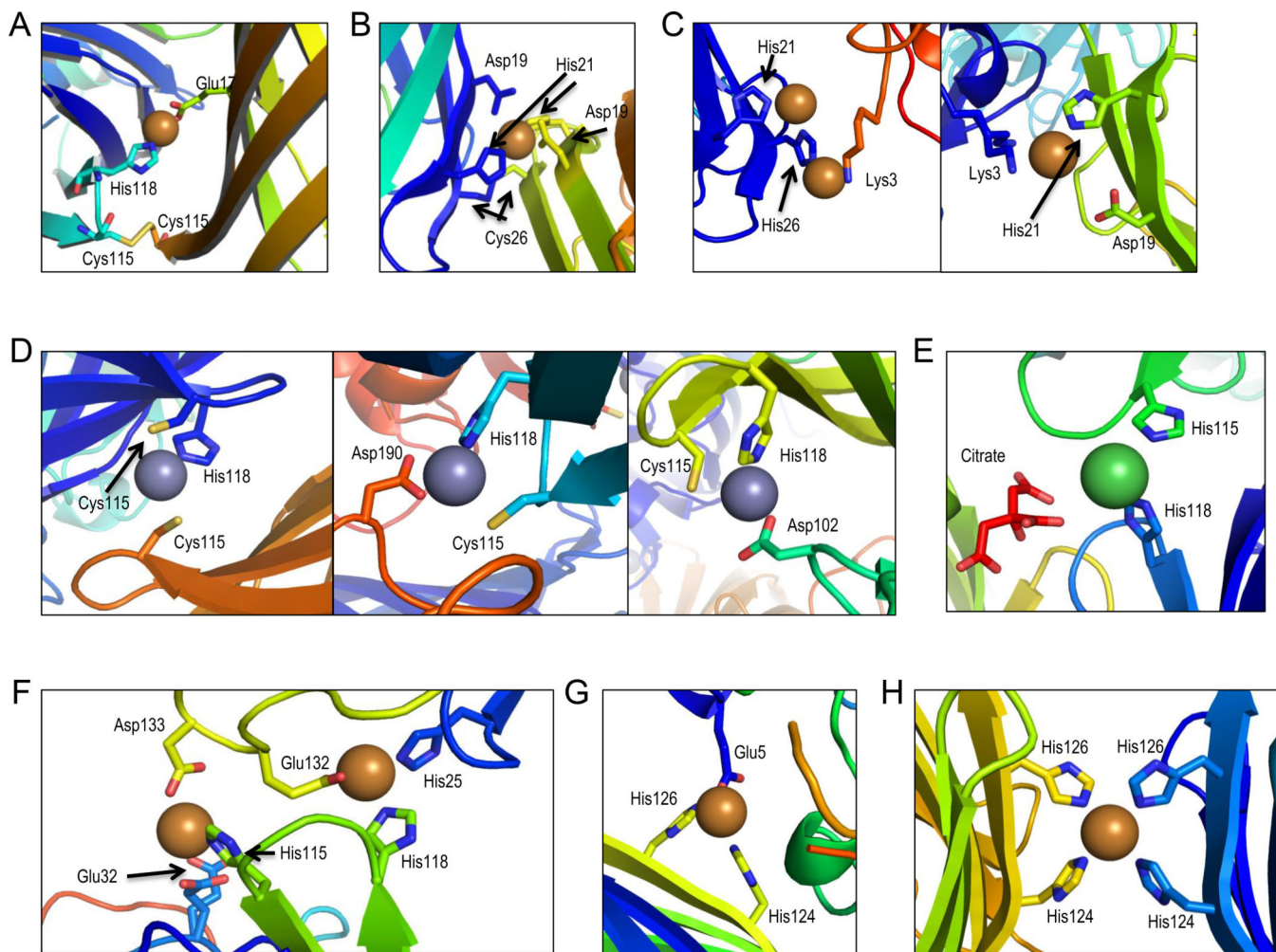
Figure 5. Observed metal-mediated crystal contacts

A) Structure 4W72: a disulfide bond is formed in addition to the copper binding site. B) Example of a mixed dimer from structure 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 4W7D. D) The three observed zinc-mediated contacts found in 4W74. (left) Cys115/His118 from one chain and Cys 115 from another chain chelate the zinc. (middle and right) Cys115/His118 from one chain chelate the zinc ion along with an aspartate (Asp190 or Asp 102) from another chain. E) A nickel mediated crystal contact in the structure of 4W6U involving hisitidines from the two proteins and a citrate molecule. F) A double copper-mediated crystal contact in the structure of 4W6T, both involving a combination of histidine and carboxylates. G) A copper-mediated crystal contact in the structure of 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 4W7R.

A

B
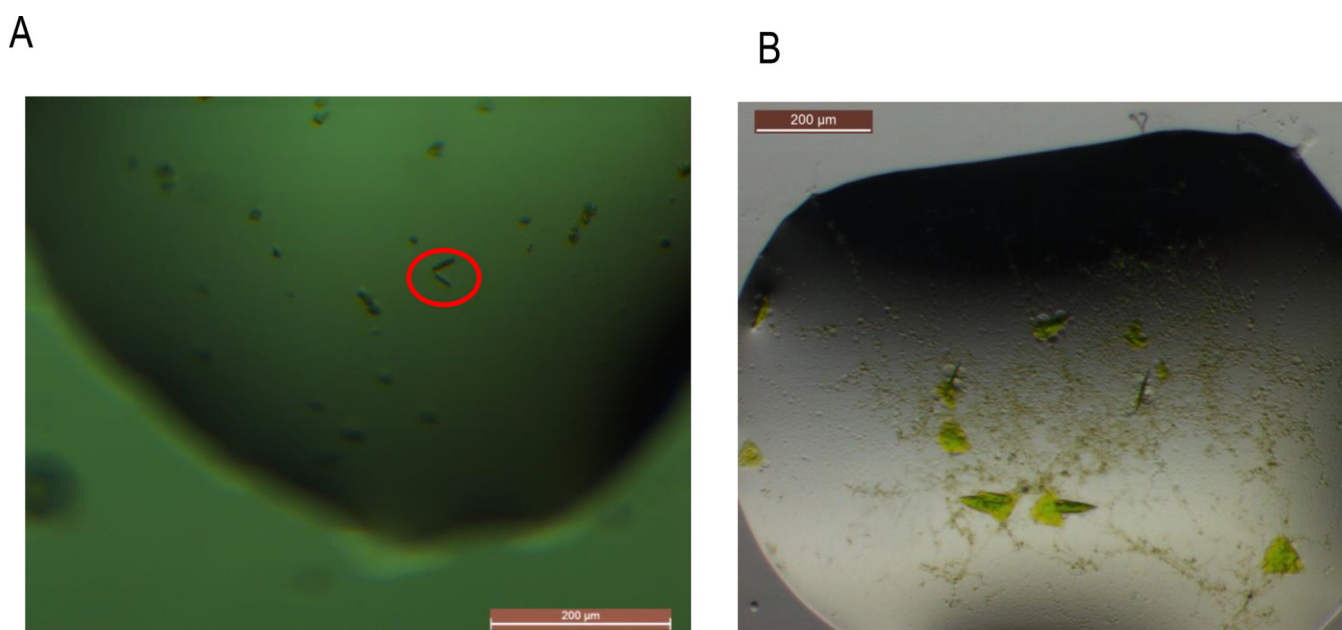
Figure 6. Crystals of split-GFP with a novel crystallization target

A) Crystals of the STARD9–10/11 – GFP1–9 (D21H/K26C) complex were obtained in a condition composed of 10%v/v 2-propanol, 0.1M MES pH 6.0 and 0.2M 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 ($\sim$20μM in the largest dimension) are highlighted by the red circle. B) Crystals of a designed protein with an engineered internal 10/11 hairpin in complex with GFP1–9 (D117C). The triangular plate crystals ($\sim$50–75μM) grew in a condition containing 0.1M SPG buffer pH 5.0 and 25% w/v PEG-1500.
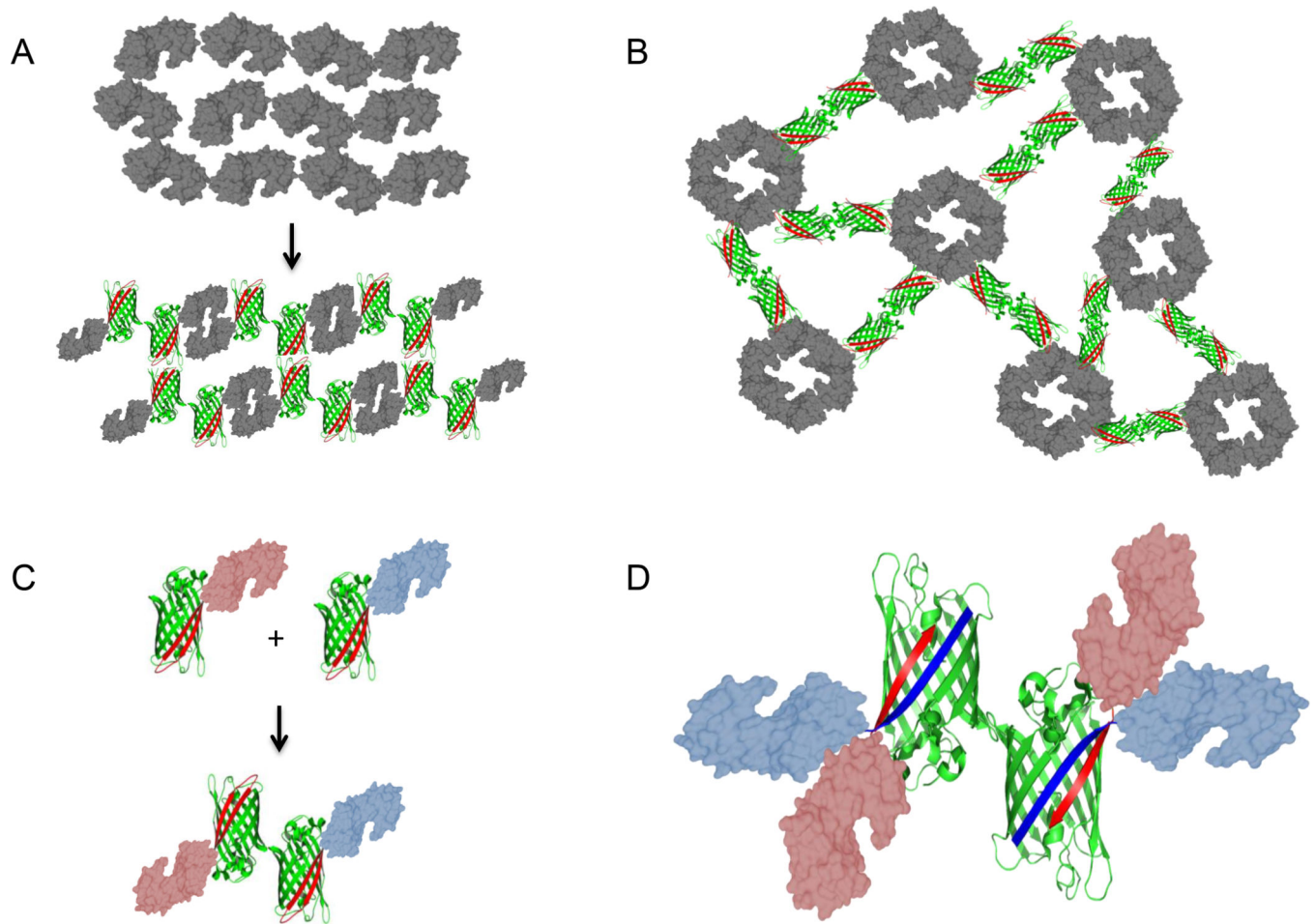
Figure 7. Alternative applications for engineered oligomeric GFPs
Beyond their proposed utility as carriers for the crystallization of novel proteins, other potentially useful application are possible. A) Fusion to GFP dimers could be used to change the crystal forms of existing proteins. Here a disordered 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.

Table 1

Summary of New GFP Crystal Forms

PDB	Mutation	Type	Space Group	Resolution (Å)	ASU#
4W69	Q157C	Disulfide	P 43 21 2	3.98	2
4W6A	Q157C	Disulfide	P 32 2 1	2.99	2
4W6B	K26C*	Disulfide	P 21 21 21	1.90	2
4W6C	D21H/K26C^	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^	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

PDB	Mutation	Type	Space Group	Resolution (Å)	ASU[#]
4W7X	E115C/T118H	Disulfide	P 1 21 1	2.80	4
4W75	D21H/K26C[^]	Disulfide + Metal Contacts	P 21 21 21	3.47	2
4W76	D21H/K26C[^]	Disulfide + Metal Contacts	P 21 21 21	2.35	2
4W77	D21H/K26C[^]	Disulfide + Metal Contacts	P 21 21 21	3.10	2
4W7A	D21H/K26C[^]	Disulfide + Metal Contacts	P 21 21 21	3.60	4
4W7C	D21H/K26C[^]	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

[*] Superfolder GFP C48A backbone mutation,

[^] Split-GFP C48A backbone mutation. All other sequences have the double mutations of C48A and C70A.

[#] Number of GFP chains in the asymmetric unit

Table 2

GFP Disulfide Dimer Characterizations

Mutant	PDB	Dimer	Disulfide Cα Distance (Å)	Dimer Angle (°)	Grouped PDBs	Chain "B" Variation Range (°)
K26C	4W6B	AB	6.4	151.66		
	4W6C	AB	6.2	175.55	Group: 4W6C, 4W6D, 4W6F Outlier: 4W6B	Group 4W6C – 4W6F = 33.3 Maximum Range: 4W6B – 4W6D = 140.4
	4W6D	AB	6.2	158.12		
	4W6F	AB	5.6	144.29		
D21H/K26C	4W7A	AB	5.8	169.72		
	4W7A	CD	6.2	177.95		
	4W7C	AB	5.9	173.38	Group: 4W7A, 4W7C, 4W76 4W77 Outlier: 4W75	Group: 4W7A AB – 4W7A CD = 6.3 Maximum Range: 4W7A CD – 4W75 = 32.1
	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		
	4W6P	CD	4.6	146.21		
	4W6P	EH	4.6	143.79		
	4W6P	FG	4.6	139.64		
	4W6R	AN	5.2	165.37	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
	4W6R	BI	4.7	165.15		
	4W6R	CD	4.1	170.66		
	4W6R	CD	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		

Mutant	PDB	Dimer	Disulfide Cα Distance (Å)	Dimer Angle (°)	Grouped PDBs	Chain "B" Variation Range (°)
	4W72	AB	5.9	159.85		
	4W73	AB	6.4	170.95		
D117C	4W6J	AB	5.7	154.89		
	4W6K	AB	5.7	166.82		
	4W6L	AB	5.5	180.0		
	4W6M	AC	5.6	178.44	Group1: 4W6K, 4W6L, 4W6M, 4W6O	Group 1: 4W6O - 4W6K = 16.4
	4W6M	BD	6.5	178.14	Group2: 4W6I, 4W6N	Group 2: 4W6N BF - 4W6J = 10.8
	4W6N	AD	6.1	148.41		Maximum Range: 4W6N BF - 4W6M AC = 34.8
	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		
	4W6A	A	5.78	180.0	--	4W96 – 4W6A B = 129
	4W6A	B	*11.7	180.0		
D190C	4W6G	AB	5.8	140.95	Group: 4W6G, 4W6H	Group: 4W6G – 4W6H = 6.3
	4W6H	AB	5.8	135.23	Outlier: 4W6I	Maximum Range: 4W6H – 4W6I =41.4
	4W6I	AB	6.4	171.21		

* Potential disulfide broken during crystallization.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

X-ray Diffraction Data and Refinement Statistics