JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

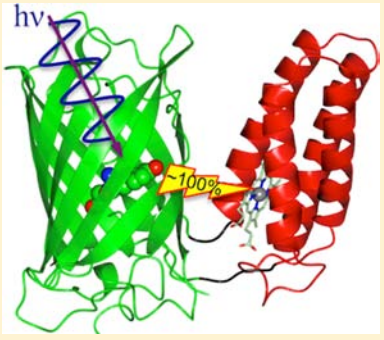
Structural Basis for Efficient Chromophore Communication and Energy Transfer in a Constructed Didomain Protein Scaffold

James A. J. Arpino,[†] Honorata Czapinska,[‡,#] Anna Piasecka,[§,▽] Wayne R. Edwards,[†] Paul Barker,[∥] Michal J. Gajda,[⊥,○] Matthias Bochtler,*[,†,‡,#] and D. Dafydd Jones*[,†]

[†]School of Biosciences and [‡]School of Chemistry, Main Building, Park Place, Cardiff University, Cardiff CF10 3AT, United Kingdom
[§]School of Medicine, Cardiff University, UHW Main Building, Heath Park, Cardiff CF14 4XN, United Kingdom
[∥]Department of Chemistry, University of Cambridge, Lensfield Road Cambridge, UK CB2 1EW, United Kingdom
[⊥]EMBL c/o DESY, 22603 Hamburg, Germany

Supporting Information

ABSTRACT: The construction of useful functional biomolecular components not currently part of the natural repertoire is central to synthetic biology. A new light-capturing ultra-high-efficiency energy transfer protein scaffold has been constructed by coupling the chromophore centers of two normally unrelated proteins: the autofluorescent protein enhanced green fluorescent protein (EGFP) and the heme-binding electron transfer protein cytochrome b_{562} (cyt b_{562}). Using a combinatorial domain insertion strategy, a variant was isolated in which resonance energy transfer from the donor EGFP to the acceptor cyt b_{562} was close to 100% as evident by virtually full fluorescence quenching on heme binding. The fluorescence signal of the variant was also sensitive to the reactive oxygen species H_2O_2, with high signal gain observed due to the release of heme. The structure of oxidized holoprotein, determined to 2.75 Å resolution, revealed that the two domains were arranged side-by-side in a V-shape conformation, generating an interchromophore distance of ∼17 Å (14 Å edge-to-edge). Critical to domain arrangement is the formation of a molecular pivot point between the two domains as a result of different linker sequence lengths at each domain junction and formation of a predominantly polar interdomain interaction surface. The retrospective structural analysis has provided an explanation for the basis of the observed highly efficient energy transfer through chromophore arrangement in the directly evolved protein scaffold and provides an insight into the molecular principles by which to design new proteins with coupled functions.



INTRODUCTION

Construction of new active biomolecular components not currently present in the natural repertoire is central to the emerging areas of synthetic biology and bionanotechnology.[1,2] The generation of relatively complex scaffolds from naturally disparate proteins is particularly attractive, as it allows normally unrelated but potentially compatible functions to be organized and coupled within a single protein.[3−6] This in turn opens up the possibility of constructing novel sensors, switches, transducers, and energy transfer components for use in biological systems or non-natural contexts.

Single polypeptide chains that house multiple functions are common in nature and generally consist of discrete domains.[7] The exact spatial arrangement of individual domains is essential for functional linkage. Most domains are arranged in a tandem "head-to-tail" manner but a significant minority (∼9%) are linked through domain insertion,[8] in which one protein is inserted within another. Domain insertion can generate intimate structural and spatial linkage between domains that facilitates functional coupling and is an emerging protein engineering strategy for linking unrelated proteins.[9−11] However, our molecular understanding of functional coupling in engineered domain insert proteins is hindered by the lack of detailed structural information. This in turn limits our ability to construct these potentially useful protein scaffolds. To date, most success has been achieved using combinatorial mutagenesis approaches to sample diverse domain insert positions and linking sequences followed by selection for coupled functionality.[12−15]

Chromophore communication is central to many processes in nature, such as photosynthesis,[16−18] and has been used as a principle for constructing biosensors.[19] Critical to the efficiency of chromophore communication is their orientation, distance, and tuning, which can be dictated through the use of suitable protein scaffolds. In this case we plan to link the properties of the redox active chromophore heme and the fluorescence output of GFP to generate a light-induced heme-dependent energy transfer system. Heme is a biologically important redox active small molecule that is commonly found associated with proteins.[20] Heme can also act as an acceptor during resonance energy transfer,[21,22] making it an excellent partner to a

Received: February 28, 2012
Published: July 23, 2012

   dx.doi.org/10.1021/ja301987h | J. Am. Chem. Soc. 2012, 134, 13632−13640

fluorescence donor. Cytochrome b_{562} (cyt b_{562}) is an example of one of the simplest heme-binding protein scaffolds. It is a helical bundle protein that binds heme noncovalently and is thought to be involved in electron transfer in the periplasm of *Escherichia coli*.[23,24] Cyt b_{562} has also been used as a model for generating simplified systems to address photoinduced resonance energy and electron transfer central to processes such as photosynthesis[25−27] and is currently being considered as an important long-range electron transfer device in bio/nanoelectronics.[28−30] Heme's affinity for cyt b_{562} is dependent on the iron oxidation state,[31] with the affinity of the reduced (ferrous/Fe^{2+}) form being much higher than that of the oxidized (ferric/Fe^{3+}) form. Cyt b_{562}-bound heme can be prone to induced oxidative modification.[32] Autofluorescent proteins such as green fluorescent protein (GFP) make excellent light-harvesting/reporter components. GFP and its variants (such as EGFP) are widely used cellular reporters[33,34] whose properties are sensitive to changes in protein conformation and proximity of Förster resonance energy transfer (FRET) partners.[34,35] Apart from fluorescence emission (and resonance energy transfer), EGFP can also photoreduce secondary molecules through electron transfer.[36,37]

Using a combinatorial domain insertion approach, we have generated variants of cyt b_{562} inserted in EGFP that exhibit nearly total fluorescence quenching in a heme-dependent manner. One variant displayed oxidant-dependent quenching. The structure of the holo form of this variant was determined by X-ray crystallography and revealed an unexpected side-by-side arrangement of the two domains with the EGFP chromophore and cyt b_{562} heme lying spatially close to each other. Critical to the juxtaposition of the two domains was the length difference of the linking sequences, which generated a molecular pivot at the linkage point, and interdomain interactions.

RESULTS

Construction and Selection of Cyt b_{562} Integral Fusions with EGFP. A recently developed random non-homologous recombination method[12,39,40] was used to generate a library of integral fusion proteins in which cyt b_{562} was randomly inserted within EGFP (see Figure S1 of the Supporting Information for details). The library was screened to identify integral fusion variants that retained EGFP fluorescence by selecting *E. coli* colonies that displayed a green color phenotype on excitation with UV light, which is indicative of a functional EGFP unit. Approximately 7% of the colonies exhibited a green phenotype (~2000 colonies in total). Sequencing of 36 randomly selected fluorescent cyt b_{562}−EGFP integral fusion variants revealed that 26 contained correctly orientated in-frame cyt b_{562} cassette inserts. Of these 26 variants, 23 had unique sequences sampling 15 different insertion positions. The disparity between the number of insertion sites and the unique sequences observed is due to the different linker sequences that can be sampled (see Supporting Information, Table S1). A detailed analysis of the cyt b_{562}−EGFP integral fusion variants will be provided elsewhere.

Screening of the 23 unique sequence variants for heme-dependent EGFP fluorescence identified three different variants, CG2, CG4, and CG6 (Figure 1), which displayed almost total quenching in the presence of excess heme (Figure 2A). In the three integral fusion variants, cyt b_{562} insertion sites were spatially clustered (Figure 1). CG2 and CG4 had cyt b_{562} inserted within five amino acids of each other toward the N-terminus of EGFP, while CG6 contained the insert in an adjacent turn between residues Tyr39 (mutated to Phe on domain insertion) and Gly40. A key difference between the three integral fusion proteins was the introduced linker sequences (Figure 1). Both CG2 and CG4 had similar length artificial linkers based on the GlyGlySer tripeptide sequence. CG6 had different length linkers: a tripeptide GlyGlySer linking EGFP to the N-terminus of cyt b_{562} but a single Gly linking the C-terminus of cyt b_{562} to EGFP (Figure 1). The single Gly linker was not originally programmed into this particular library (Supporting Information, Table S1); the origin of this important event remains unknown. In the absence of heme, all three variants exhibited excitation and emission peaks at ~488 and ~510 nm, respectively, similar to that observed for EGFP (Table 1 and Supporting Information, Figure S2). Quenching of CG6 was particularly marked, as fluorescence emission was essentially zero at a molar ratio of 1:1 of protein to heme (Figure 2). The binding curve suggested that CG6 retained high affinity for heme under reducing conditions, mirroring that of wild-type cyt b_{562} (K_D in the picomolar range). In contrast, full quenching of CG2 and CG4 required higher concentrations of heme (~7.5- and ~3-fold more, respectively).

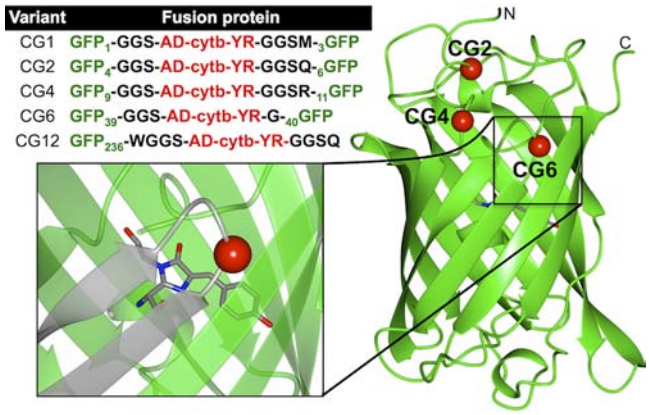


Figure 1. Insertion of cyt b_{562} within EGFP. The sequence of the cyt b_{562}−EGFP fusion proteins are shown in the table (top left). The EGFP section is colored green, with the subscript numbers representing the residues immediately upstream and downstream of the insertion; black letters are linker sequences; red letters represent the cyt b_{562} insert domain. The structure of GFP (right) with the cyt b_{562} domain insert positions for CG2, CG4, and CG6 shown as red spheres. A close up of the cyt b_{562} insertion position for CG6 is shown in the inset (bottom left), with the chromophore shown in the stick representation. All molecular structure diagrams were generated using CCP4mg[38] or PyMol.

In comparison, library variants CG1 (N-cyt b_{562}-EGFP-C) and CG12 (N-EGFP-cyt b_{562}-C) that equate to classical head-to-tail fusions (Figure 1) retained >35% fluorescence even under saturating concentrations of heme but maintained high affinity for heme (Table 1 and Figure 2A). The addition of heme to EGFP had no affect on fluorescence emission, confirming that quenching was induced by heme binding to the fusion proteins (Supporting Information, Figure S3).

Characterization of CG6. Initially, all variants were standardized to a set fluorescence emission intensity of 100 au in the absence of heme (equivalent to 20 nM of CG6). When compared to CG6, the requirement of higher heme concentrations to fully quench fluorescence of CG2 and CG4 suggests that domain insertion is having an unfavorable

dx.doi.org/10.1021/ja301987h | J. Am. Chem. Soc. 2012, 134, 13632−13640

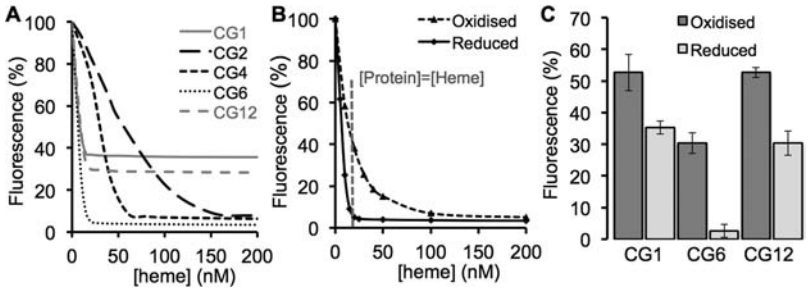


Figure 2. Heme-mediated fluorescence quenching of EGFP−cyt b_{562} chimeras. Oxidizing and reducing conditions were induced by the addition of 1 mM KNO_3 or 1 mM ascorbic acid, respectively. Excitation and emission wavelengths were 488 and 511 nm, respectively. The fluorescence intensity prior to heme addition was standardized to 100 au and equates to 20 nM protein. (A) Heme-dependent fluorescence quenching of the selected variants, as annotated in the figure, under reducing conditions. (B) Heme-dependent fluorescence quenching of CG6 under oxidizing and reducing conditions, as annotated in the figure. (C) Redox-dependent heme-mediated fluorescence quenching of selected variants. [Protein]:[heme] was 1:1.

Table 1. Fluorescence Properties of EGFP−Cyt b_{562} Variants

variant	λ_{Ex} (nm)	λ_{Em} (nm)	ε ($M^{-1}cm^{-1}$)[a]	quantum yield	brightness ($M^{-1}cm^{-1}$)	fluorescence lifetime (ns) τ_{apo}	fluorescence lifetime (ns) τ_{holo}	holo-chimera λ_{max} (Ox/Red)[b] (nm)	heme^{Ox} K_d (nM)
EGFP	487	511	55 000	0.63	34 650	2.47	2.48	NA	NA
CG1	487	509	52 330	0.63	32 960	2.38	1.64	418/428	11.5 ± 1.4
CG6	488	511	41 250	0.65	26 810	2.47	NA	422/428	11.0 ± 1.8
CG12	486	512	49 250	0.61	30 040	2.39	1.80	418/428	11.8 ± 1.0
cyt b_{562}	NA	NA	NA	NA	NA	NA	NA	418/427[c]	~10[c]

[a]Molar absorbance coefficient of EGFP measured at 488 nm. [b]The λ_{max} values correspond to absorbance peaks derived from the heme-bound species of the protein. [c]Reported by Robinson et al.[31] and Della Pia et al.[28]

influence on the fluorescence (e.g., reduced quantum yield) and/or heme binding (e.g., lower affinity) properties of these variants. Therefore, CG6 was considered the optimal scaffold in terms of maintaining high fluorescence emission intensity and heme affinity and thus was chosen for further characterization.

In the case of CG6, cyt b_{562} was inserted not within an extended loop region but between Tyr39(Phe) and Gly40, which lie in a helical-like turn (Figure 1). Insertion within the turn did not appear to dramatically change the fluorescence properties of the EGFP domain (Table 1). Excitation and emission maxima together with the quantum yield were similar to that observed for EGFP. Brightness (26 810 M^{-1} cm^{-1}) was slightly reduced compared to that of EGFP due to a lower molar absorbance coefficient (Table 1). The fluorescence lifetime measurements mirrored results of fluorescence quenching. The lifetime for apo-CG6 was 2.47 ns, similar to that of EGFP (Table 1), but no fluorescence was detectable for holo-CG6. The lifetimes for holo-CG1 and holo-CG12 were still detectable, albeit lower than for the apo forms, confirming that full quenching was not achieved.

Heme binding by CG6 was also similar to that of cyt b_{562}. The main Soret absorbance band under reducing conditions for holo-CG6 was 428 nm, with the oxidized form slightly red-shifted by 4 nm (Table 1) compared to cyt b_{562}. CG6 retained the high affinity for oxidized heme with an observed K_D (11 nM) similar to that of wild-type cyt b_{562} (Table 1). The higher affinity of reduced (ferrous/Fe^{2+}) over oxidized (ferric/Fe^{3+}) heme for cyt b_{562} was manifested in CG6, with maximal fluorescence quenching of CG6 requiring a 10-fold higher concentration of heme under oxidizing conditions compared to reducing conditions (Figure 2B). At equimolar concentrations of CG6 and heme (20 nM), fluorescence intensity was >20-fold higher under oxidizing conditions (Figure 2C), with negligible fluorescence intensity observed under reducing conditions. Only a ~1.5-fold difference in fluorescence intensity was observed for either CG1 or CG12, which is likely a result of high fluorescence signal even in the heme-bound state.

The real-time kinetics of redox-induced heme association to and disassociation from CG6 was also assessed. Addition of heme in the presence of the natural reducing agent ascorbate caused fluorescence to be quickly quenched (0.55 units min^{-1}) to near zero (Figure 3A). In comparison, the fluorescence intensity of CG1 dropped by only 55% (Supporting

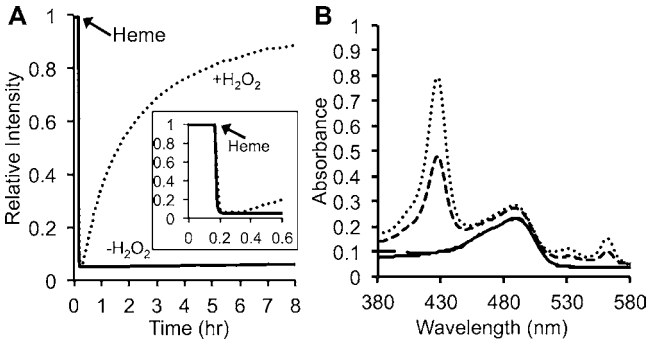


Figure 3. Oxidant-induced fluorescence switching of CG6. (A) Heme (30 nM) was added to 20 nM purified apo-CG6 in the presence of ascorbate (1 mM) to induce quenching. At 20 min, 0.02% (v/v) H_2O_2 was added to simulate a switch from reducing to oxidizing conditions (dashed line) or omitted as a control (black line). The inset shows the first 36 min of the curve. (B) Absorbance spectra of apo-CG6 before addition of heme (solid line) and CG6 on addition of heme under reducing conditions (dotted line). Spectra were then measured 2 h (short dashed line) and 20 h (long dashed line) after the addition of H_2O_2.

dx.doi.org/10.1021/ja301987h | J. Am. Chem. Soc. 2012, 134, 13632−13640

Information, Figure S4), but the rate of heme binding was ∼12-fold quicker (6.4 units min^{-1}). In the absence of any oxidizing agent, little heme dissociation was observed (Figure 3A). Various oxidants were then assessed for their ability to relieve quenching. On addition of H_2O_2, a common biologically important reactive oxygen species, relatively slow heme dissociation was observed, as evident by the rate of gain in the fluorescence signal (0.01 units min^{-1}) back to the original level (Figure 3A). Two other oxidizing agents, KNO_3 and NaOCl, did not result in a significant gain in fluorescence signal (Supporting Information, Figure S5). In the case of CG1, the signal rapidly reached the original level on addition of H_2O_2 (Supporting Information, Figure S4), with a rate ∼13-fold quicker (0.13 units min^{-1}) than that of CG6. Neither ascorbate nor H_2O_2 alone affected apo-CG6 fluorescence (Supporting Information Figure S6). Absorbance spectroscopy confirmed that heme was bound to CG6 under reducing conditions as the spectrum has characteristic peaks of both holo-cyt b_{562} and EGFP (Figure 3B). On addition of H_2O_2, the characteristic holo-CG6 spectra fades, leaving only the peak attributed to EGFP, confirming heme dissociation from CG6.

Domain Arrangement of Holo-CG6. To understand the molecular basis of the efficient fluorescence quenching in the CG6 EGFP−cyt b_{562} integral fusion variant, the holo form was crystallized in the oxidized form. The apo and holo forms of CG6 gave similar circular dichroism spectra and size exclusion chromatograms, suggesting that there were no gross structural changes between the two monomeric forms (Supporting Information, Figure S7). Crystals grew in space group $P2(1)$ and contained three molecules of the fusion protein in the asymmetric unit. Two of these were very well ordered. The third molecule (chain C) had high temperature factors, particularly for the EGFP domain. The crystallographic statistics are shown in Table 2. Routine molecular replacement was used to orient and position the EGFP and cyt b_{562} domains for the two well-ordered copies of the molecule in the asymmetric unit. The poorly ordered third molecule did not give significant molecular replacement signals and was found during refinement. In all three instances in the asymmetric unit, the two domains were juxtaposed in the crystal, so that the fusion protein adopted a "V" shape (Figure 4A). The V-shape of the EGFP−cyt b_{562} fusion brought the two domains into close proximity. The structures of the three molecules in the unit cell were similar to each other (Supporting Information, Figure S8), with the individual domains overlaying well with previously solved structures of EGFP and cyt b_{562} (Supporting Information, Figure S9). There was some variation in the hinge angle between the domains (up to 23° for different pairs; Supporting Information, Figure S8A), but the interface appeared largely unchanged. The slight differences in hinge angle between the domains observed in the three molecules in the asymmetric unit might suggest a degree of structural flexibility.

Table 2. Crystallographic Statistics

Data Collection	
space group	$P2(1)$
a (Å)	64.75
b (Å)	125.20
c (Å)	89.26
β (deg)	90.37
resolution range (Å)	30−2.75
total reflections	136 754
unique reflections	36 793
completeness (%) (last shell)	99.3 (100.0)
I/σ (last shell)	19.9 (2.4)
R(sym) (%) (last shell)	3.4 (32.4)
B(iso) from Wilson (Å^2)	80.5
Refinement	
protein atoms excluding H	7996
solvent molecules	98
R-factor (%)	23.4
R-free (%)	27.7
rmsd bond lengths (Å)	0.02
rmsd angles (deg)	1.8
Ramachandran core region (%)	92.7
Ramachandran allowed region (%)	6.6
Ramachandran additionally allowed region (%)	0.1
Ramachandran disallowed region (%)	0.6

The differential lengths of the introduced linker sequences appear to play a critical role in formation of the V-shape domain arrangement. The single glycine that links Arg148 (106^{cytb}) and Gly150 (40^{EGFP}) is located in a loop between the N-terminal end of β-strand three in EGFP and the C-terminal α-helix of cyt b_{562} that comprises the acute inner turn pivot point at the domain junction. The longer Gly-Gly-Ser linker sequence that connects Phe39 ($Y39F^{EGFP}$) and Ala43 ($A1^{cytb}$) contributes to the more elongated outer turn (Figure 4B). The structure suggests that linker residues, especially those comprising the outer turn, also play a role in the defining interdomain interactions (vide infra).

Small angle X-ray scattering (SAXS) was used to confirm the domain arrangement of cyt b_{562} and EGFP in solution. The experimental intensity data were compared with the predicted data for different domain arrangements: the two domains aligned side-by-side (Figure 5), at a right angle (Supporting Information, Figure S10A) and aligned to form a long rod (Supporting Information, Figure S10B). An excellent fit was observed for the side-by-side domain arrangement ($\chi = 0.98$) as observed in the crystal structure and clearly stands out compared to the other models (χ = 1.92 and 3.03). This confirms that the V-shaped domain arrangement appears to be at least predominant (or even exclusively present) in the investigated solution.

Chromophore Arrangement. The V-shape of CG6 positions the chromophores of EGFP and cyt b_{562} close together in all three molecules in the unit cell (Figure 6A and Supporting Information, Figure S8). The distance between the OH group of the EGFP chromophore and the heme iron is estimated to be 17 Å and thus much smaller than would be expected for a more open arrangement of the domains. This distance is similar in all three molecules (∼17.6 Å) in the asymmetric unit (Supporting Information, Figure S8B). The shortest measured edge-to-edge distance, between the OH of the EGFP chromophore and a propionate Cα of heme, is ∼13.7 Å (Figure 6A and Supporting Information, Figure S8B). The two chromophores lie in the same plane with the angle between them approximately 45° (Figure 6A).

EGFP Chromophore Environment. The observed residue arrangement suggests that an extensive hydrogen-bonding network between the EGFP chromophore and surrounding residues in the core of the protein, characteristic of fluorescent proteins, is maintained in CG6 (Figure 6B). Residue orientations and distances indicate that the phenolate group of the EGFP chromophore forms direct hydrogen bonds with

dx.doi.org/10.1021/ja301987h | J. Am. Chem. Soc. 2012, 134, 13632−13640

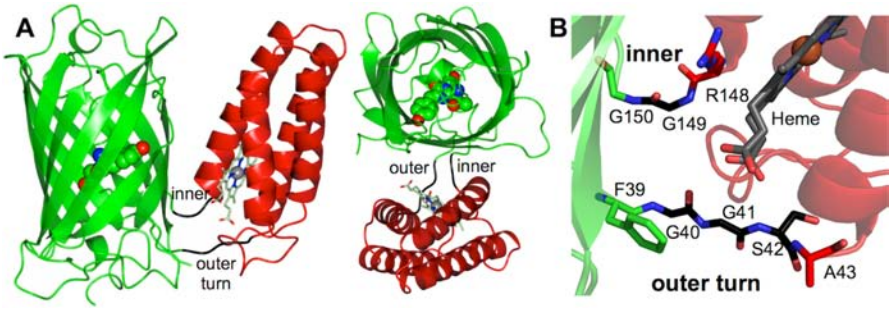


Figure 4. Crystal structure of oxidized holo-CG6. (A) Ribbon representation of CG6 (PDB 3U8P) viewed from the side (left) and top (right) with the EGFP domain colored green, the cyt b_{562} domain colored red, and the introduced linkers black. The inner and outer turns are labeled. Heme (gray) and the EGFP chromophore (green) are shown as sticks and spacefill model, respectively. (B) Linker sequence structure with domains colored as in panel A.

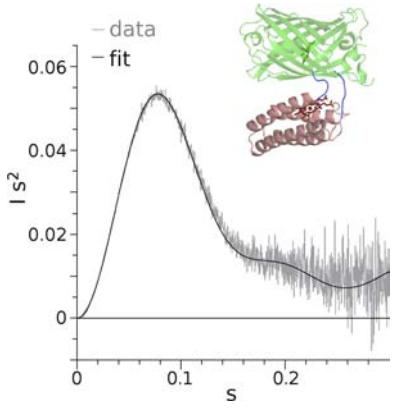


Figure 5. Domain arrangement of CG6 in solution as determined by SAXS. The fit of predicted small-angle X-ray scattering (SAXS) curve for the domain arrangement observed in the crystal (black line) versus the experimentally determined diffraction data for CG6 after appropriate scaling (gray line). Fits to alternative domain arrangements are shown in the Supporting Information (Figure S10).

the hydroxyl group of Thr311 (203^{EGFP}) and the Nδ of His256 (148^{EGFP}). The Nε of His256 is within hydrogen-bonding distance of the backbone nitrogen of Arg276 (168^{EGFP}). The position of residues that contribute to the hydrogen bond network promoting formation of the phenolate form of the chromophore that gives rise to the single peak excitation spectra in EGFP appear to be preserved. The apparent hydrogen-bonding chain observed in wild-type GFP linking Ser65 (Thr65 in EGFP) to Tyr66 through the side chains of Glu222, Ser205, and a conserved water molecule is interrupted between Glu330 (222^{EGFP}) and Thr175a (65^{EGFP}) in CG6 (Figure 6B). The nature of this hydrogen-bonding network suppresses deprotonation of Glu330 and in combination with the polar interactions with the chromophore phenolate promotes formation of the anionic form of the chromophore in the ground state (Supporting Information, Figure S2). Residues suggested as contributors to the hydrogen bond network around the EGFP chromophore are also in close proximity to the domain–domain interface (Figure 6B,C). Some adjacent residues that are normally surface exposed are suitably orientated and close to residues in cyt b_{562}, suggesting that defined interdomain contacts are made, most notably Gln312 (204^{EGFP}) and Phe331 (223^{EGFP}) in proximity to Arg148 (106^{cytb}).

Interdomain Interactions. While the structural resolution is limited to 2.75 Å, it is clear that the two domains form an interaction surface close to the molecular pivot point. The interface between cyt b_{562} and EGFP is remarkably hydrophilic (Figure 6C) and does not look like a typical naturally evolved protein interaction surface. The interface between the cytochrome and EGFP domains extends over an area of approximately 700–800 Å^2 and buries about 1400–1600 Å^2 of solvent accessible surface. Analysis of the structure suggests that a defined set of residues from both domains contribute to the formation of the domain interaction surface mostly via hydrogen bonding, albeit with the exact nature of some of the interactions differing slightly between the molecules in the unit cell. The structure suggests that favorable interactions are possible that stabilize the conformation of the outer linker (Figure 6C and Supporting Information, Figure S11). Depending on the molecule, a hydrogen bond is likely to form between

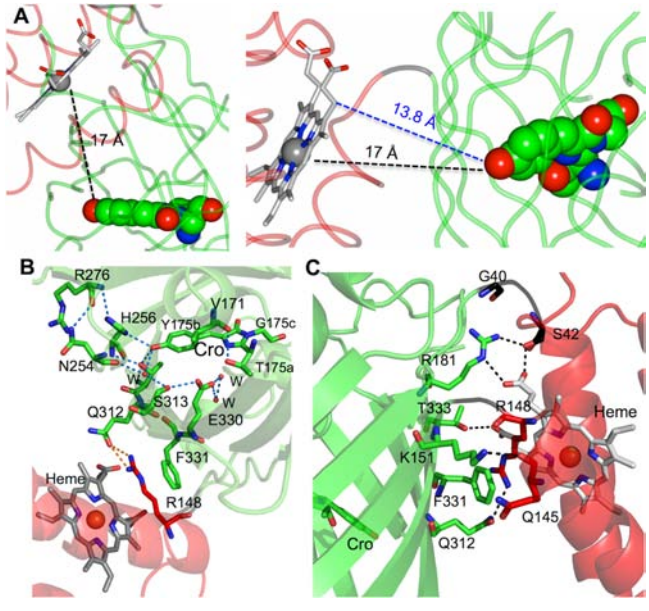


Figure 6. Interdomain communication. (A) Relative orientations and distance between the two chromophores (colored as in Figure 4A). The iron atom at the center of the porphyrin ring is shown as a sphere. Shown are the edge-to-edge (blue) and porphyrin iron to GFP chromophore hydroxyl group (black) distances. (B) The apparent hydrogen bond network associated with the EGFP chromophore. (C) Interdomain interface region for molecule A. Polar interactions between residues calculated on the basis of observed distances and orientations are shown as dashed lines. The interactions for molecule B are shown in the Supporting Information (Figure S11).

the guanidino group of Arg181 (73^{EGFP}) and the main chain O atom of Gly40 (linker) and/or from the main chain N atom of Ser42 (linker) to one of the heme carboxylate groups. A salt bridge is also possible that links the heme carboxylate and the guanidino group of Arg181 (73^{EGFP}). Again, depending on the molecule in the asymmetric unit, the other heme carboxylate may also form a salt bridge with Arg181 (73^{EGFP}) or alternatively could form a hydrogen-bonding interaction with Thr333 (225^{EGFP}). Another important contributor to the domain interface is thought to be Arg148 (106^{cytb}), which stacks favorably against Phe331 (223^{EGFP}) and forms a hydrogen bond with Gln312 (204^{EGFP}).

Heme Binding Environment. In wild-type oxidized holo-cyt b_{562}, the heme group is bound noncovalently, with the heme iron coordinated between the sulfur atom of Met7 and the Nε of His102. In CG6, the heme group is coordinated in the same manner with the main difference being placement of the ring C associated propionate (propionate 7) group (Figure 7). In comparison to wild-type cyt b_{562}, formation of the domain interface results in the heme group becoming largely solvent inaccessible; the accessible surface of the heme group drops approximately by half from ~160 Å^2 in wild-type cyt b_{562} to ~70 Å^2 in CG6. The ring D associated propionate group also makes cross-domain interactions. These additional heme interactions and reduced solvent exposure may be the cause of the observed small red shift in the Soret band λ_{max} of oxidized holo-CG6 (Table 1).

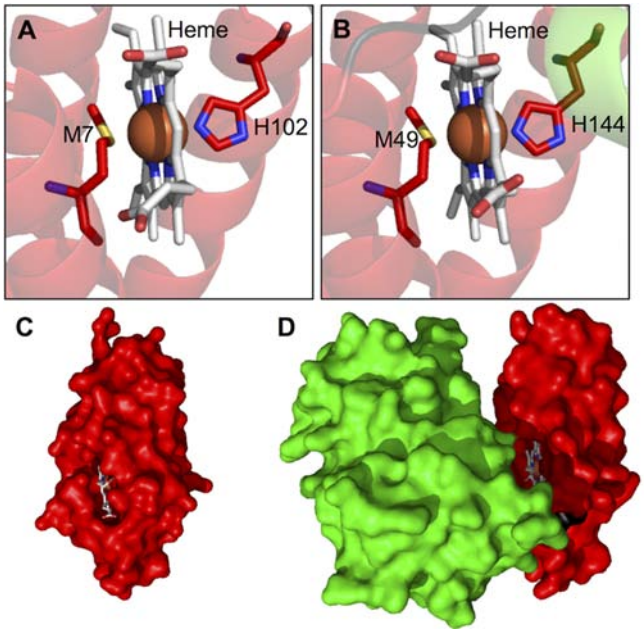


Figure 7. Heme binding environment in CG6. (A) wild-type cyt b_{562} (red, PDB code 256B) with heme and its axial ligands, Met7 and His102, shown as sticks. (B) CG6 with the EGFP domain in green, the cyt b_{562} domain in red, and the linkers in black. Heme and its axial ligands, Met49 and His144, are shown as sticks. Surface representation for wild-type cyt b_{562} (C) and CG6 (D).

DISCUSSION

Domain insertion provides a general approach for structurally and spatially linking normally disparate proteins so that the function of one protein is coupled to another; it provides a mechanism to artificially construct novel molecular switches for biosensing or even as energy transducers in bionanotechnology. From a protein engineering perspective, the challenge lies in predicting sites within the accepting protein that not only tolerate insertion of a whole domain but also couple the functions of the two proteins. The use of directed evolution to sample a diverse range of domain insertion sites and domain linking sequences coupled with screening for linked functionality has provided a useable approach that has met with some success.[12,14,15,35,41] However, in these constructed protein scaffolds little is known about how the key facet of functional coupling is achieved at the molecular level. This in turn hinders our ability to generate suitable scaffold models that could form the basis of designing optimal protein switches. Therefore, retrospective structural analysis of directly evolved protein constructs is crucial.

In our constructed system, EGFP and holo-cyt b_{562} act as classical fluorescing and sensitizing components, respectively. Heme is known to quench fluorescence, which is thought to occur via resonance energy transfer.[21,22] Critical to transfer efficiency is the positioning of the two chromophores within the protein scaffold; addition of heme alone to EGFP does not promote energy transfer (Supporting Information, Figure S3). Results shown here (Figure 2) and elsewhere[22] confirm that the generation of traditional head-to-tail cyt b_{562}–EGFP fusions (CG1 and CG12 in this study) is not sufficient to maximize energy transfer, presumably because the two domains are not fixed relative to one another and the distance between the chromophores is nonoptimal (~41 Å based on the level of quenching). Therefore, a more sophisticated protein scaffold is required to promote chromophore communication. The three variants that displayed near total quenching in the presence of heme, CG2, CG4, and CG6 are all clustered to the same spatial region of the EGFP β-barrel structure (Figure 1). Therefore, optimal coupling between the chromophore centers may require insertion of cyt b_{562} into specific structural regions or "hot spots" in EGFP. The importance of the site within EGFP where domain insertion takes place is further emphasized by variant CG2; cyt b_{562} is inserted just five residues from the EGFP N-terminus (Figure 1), but compared to CG1, total quenching was observed on binding heme (Figure 2). Thus, small shifts in insertion position can have a dramatic effect with regard to the organization and positioning of heme with respect to the EGFP chromophore.

The structure of CG6 provides an explanation for high-efficiency energy transfer between the protein-bound heme and the EGFP chromophore. The two domains lie side-by-side, forming a characteristic V-shape conformation (Figures 4–6). The domain arrangement results in an interchromophore distance of ~17 Å (~14 Å edge-to-edge) (Figure 6A). The calculated R_0 (the Förster radius at which energy transfer is 50% efficient) for the chromophores of cytochrome b_{562} and EGFP is 46 Å.[22] Energy transfer efficiency (E) is related to the interchromophore donor–acceptor distance (r) through the inverse sixth power by the equation $E = 1/[1 + (r/R_0)^6]$, which with r being 17 Å gives an E of 99.7%. This is in line with observed near total quenching of fluorescence of holo-CG6 (Figure 2). High energy transfer efficiency is confirmed through analysis of the fluorescence lifetime in the presence and absence of heme through the relationship $E = 1 - (\tau_{holo}/\tau_{apo})$, where τ_{holo} and τ_{apo} represent the lifetimes in the presence and absence of heme. As τ_{holo} is essentially 0 (Table 1), E is 100%. Close to total energy transfer is not normally achieved in constructed fluorescent protein-based scaffolds,[34,42] thus energy

dx.doi.org/10.1021/ja301987h | J. Am. Chem. Soc. 2012, 134, 13632–13640

transfer efficiency, interchromophore distance, and orientation in the presented new scaffold is comparable to natural energy transfer systems,[16,17] such as the light-harvesting complexes. This feature potentially makes the CG6 scaffold an excellent simple system to study the basis of resonance energy transfer mechanisms in biology and may act as a starting point for a simple biobased light-harvesting system. As EGFP can also undergo photoinduced electron transfer[36,37] and cyt b_{562} can act as an efficient electron shuttle,[28,29] CG6 may also have a potential application as an artificial mimic of a photosynthetic reaction center and may even act as a biobased photovoltaic device. The minimum observed edge-to-edge distance between the two chromophores in CG6 is 13−14 Å (Figure 6B), which is within the distance limit applicable for natural electron transfer systems.[18]

Critical to the side-by-side placement of the domains is the linker sequences and interaction of the two domains. The different lengths of the two introduced interdomain linkers (Figures 1, 4B, and 6C) contribute to the formation of a molecular pivot point. The single amino acid linker forms the acute inner turn, while the longer linker forms the outer turn. While the introduction of a single amino acid at one of the linking sites was a serendipitous event (Supporting Information), it underlies the importance of differential linker lengths for the two domains assuming their relative positions. Gly40 occupies a confined environment, and replacement with a longer linker may not be tolerated sterically in the current conformation and introduce additional flexibility between the two domains. Contributing to the domain arrangement is the formation of an interdomain interaction surface comprised of predominantly polar contacts (Figure 6B,C and Supporting Information, Figure S11), which is atypical of naturally evolved domain/subdomain interfaces. Residues outside the linking sequences are important contributors to the domain interface. One of these residues, Arg181 from the EGFP domain, forms a salt bridge to a heme carboxyl group. While some limited focus is given to arbitrary linking sequences when constructing linked domains, very little consideration is given to any potential domain interactions, as these are generally difficult to anticipate. From a design perspective, we have shown by retrospective analysis of a directly evolved domain insert scaffold that domain placement is essential for maximized functional coupling (in this case energy transfer); linker sequences together with domain interactions are essential in achieving this.

The structure also provides an insight into some of the observed heme-dependent fluorescence output features of CG6. Heme binding induces very little apparent structural change in CG6 (Supporting Information, Figure S7), so the general structure of the apo- and holoproteins could be considered largely equivalent. The oxidation-state-dependent affinity of heme for cyt b_{562} in CG6 is maintained with fluorescence quenching occurring at lower heme concentrations under reducing conditions (Figure 2). As heme binds to CG6 in a similar manner to cyt b_{562} (Table 1 and Figure 7), it is not unexpected that redox-dependent affinity is retained. The oxidation-state-dependent heme affinity coupled with the high signal gain on heme dissociation makes CG6 an attractive potential sensor of changes from the normally reducing conditions inside the cell to oxidizing that accompanies several important biological events. Redox-dependent fluorescence output was demonstrated by heme titration; at a 1:1 ratio of heme and CG6, fluorescence output was >20-fold higher under oxidizing conditions (Figure 2B,C). Real-time analysis of CG6 showed that initial heme binding and fluorescence quenching under reducing conditions were rapid but was still ∼12-fold slower than that observed for CG1. On addition of an oxidizing agent, the signal gain was slow or nonexistent (Figure 3 and Supporting Information, Figure S5). The only tested oxidant to elicit a response was H_2O_2, a common biological reactive oxygen species, making CG6 a potentially H_2O_2 specific sensor. H_2O_2 may be exerting its effect through oxidative chemical modification of heme rather than a change in the iron moiety redox state. Oxidative cleavage of the porphyrin ring of cyt b_{562}-bound heme has been observed previously.[32] The structure of CG6 provides a plausible explanation for both slower heme association and dissociation kinetics compared to the head-to-tail construct CG1. The heme binding pocket that is normally accessible to the solvent lies at the domain interface in CG6, thus restricting access (Figure 7C,D). In the holo form, heme also makes additional interactions with the protein via the carboxylate groups (Figure 6C and Supporting Information, Figure S11). In comparison, the head-to-tail fusion variant CG1 displays faster association and H_2O_2-dependent dissociation kinetics (Supporting Information, Figure S4) that are likely due to the heme retaining access to the solvent (Figure 7C) and the absence of contacts between the heme carboxyl groups and the protein.

Domain insertion could be considered a largely disruptive mutational event, given that continuity of the polypeptide chain of the protein accepting the domain is broken. Dogma dictates that insertion positions are likely restricted to inherently flexible regions such as loops; traditionally, rational sequence insertion approaches have focused on such loops.[43] CG2 and CG4 together with other so far unreported EGFP−cyt b_{562} integral fusion proteins conform to this dogma. CG6, on the other hand, has cyt b_{562} inserted within a more structurally constricted site, a helical-like turn linking strands 2 and 3 of EGFP (Figure 1). Insertion at this position does not appear to disrupt the structures (Figure 4 and Supporting Information, Figure S9), chromophore environments (Figures 6B and 7A,B), or functions of the individual protein domains significantly (Table 1). Functional insertion between residues 39 and 40 of EGFP has not been to our knowledge reported before. Thus, providing a tolerant insertion site and appropriate linker sequences are utilized, it is possible to generate unanticipated integral domain fusions with the desired coupled function.

In conclusion, we have provided a molecular explanation of how the functions of two normally disparate proteins are coupled through a directed evolution domain insertion process. The side-by-side domain arrangement results in high-efficiency energy transfer from the chromophore of EGFP to the heme group of cyt b_{562}, comparable to that of natural systems and beyond that of existing fluorescent protein-based systems. Critical to the domain organization are the linker sequences and domain interactions. Retrospective structural analysis has thus proved valuable in understanding the functional features of a directly evolved protein scaffold not present in nature and provides the basis by which to engineer future scaffolds for use as novel and useful biological components. With respect to CG6, this may include its adaption for use as a heme/redox sensor to nanoscale light-induced electron transfer devices.

METHODS

Library Construction and Preliminary Screening. A detailed description of library construction and screening is provided in the

Supporting Information (Methods section). Briefly, insertion of the engineered transposon MuDel into the *eGFP* gene encoding EGFP residing within the pNOM-XP3 plasmid was performed using an in vitro transposition and selection procedure described previously.[44] The transposon insertion library (termed EGFPΔ2504) was estimated to contain 2504 variants with MuDel inserted throughout the *eGFP* gene. Introduction of the DNA cassette containing cyt b_{562} was performed essentially as described previously.[12,41] The main difference was the use of a kanamycin selection marker within the cyt b_{562} cassette to facilitate identification of gene variants with a cassette insert from those that did not. The resulting library was screened to isolate variants that retained fluorescence by selection of colonies on plates that were fluorescent.

Protein Production and Analysis. Detailed description of all the applied methods is provided in the Supporting Information (Methods section). Cell lysates containing the selected variants were produced from cultures and variants purified. Excitation and emission spectra of cell lysates and purified protein were recorded in 50 mM Tris-HCl pH 8.0 supplemented with 150 mM NaCl at 25 °C using a Varian Cary Eclipse fluorescence spectrophotometer with cuvette dimensions of 5 × 5 mm, 10 nm band-pass, and a medium scan rate (600 nm/min). Quantum yields were determined using fluorescein as a reference. The heme binding affinity under oxidizing conditions was calculated from the heme-mediated fluorescence quenching data. UV−visible absorbance spectra were recorded with a Hewlett-Packard diode array spectrophotometer with a 1 cm path length. Size exclusion chromatography and apparent molecular weight calculation were performed using a calibrated Superdex 200 column. CD spectroscopy was performed using a Chirascan CD spectrometer (AppliedPhotophysics) between 190 and 250 nm at a scan rate of 1 nm/s in a 1 mm path length quartz cuvette.

CG6 Structure Determination. Holo-CG6 protein samples (10 mg/mL CG6 in 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, and 256 μM heme) were screened for crystal formation by the sitting drop vapor diffusion method with incubation at 4 °C. Drops were set up with equal volumes of protein and precipitant solutions. The crystal of holo-CG6 was obtained from 0.1 M MES/NaOH, pH 6.4, 200 mM magnesium acetate, and 20% (w/v) PEG 8000. A crystal was transferred to mother liquor supplemented with 16% (v/v) glycerol as a cryoprotectant and vitrified. Data were collected on the Diamond Light Source beamline I02. Holo-CG6 fusion protein crystallized in a nearly orthorhombic lattice. Reprocessing the data for a monoclinic lattice indicated cell constants of 64.75 Å × 125.20 Å × 89.26 Å and β = 90.37°. The extinctions on the 0l0 reciprocal space axis further predicted that the correct space group was *P*1 2(1) 1. Assuming that the insertion of cyt b_{562} into EGFP would leave the β-barrel structure of GFP intact, molecular replacement was attempted with a single GFP molecule (PDB code 2HQZ) as the search model, using the CCP4 program MOLREP.[45] Two molecular replacement solutions were initially found. Two cyt b_{562} models were then placed using MOLREP with fixed EGFP molecules already in place. To pair the correct EGFP and cyt b_{562} fragments, crystallographic symmetry mates were displayed. Pairings were supported by connecting density, as well as similar relative domain orientations. The structure with two fusion proteins was adjusted manually using COOT.[46] The presence of a third molecule was suggested by a gap in the packing arrangement and by the strong clustering of residual density, which was of approximately the right size and shape. Inspection indicated that the density for the EGFP domain was poorer than for the cyt b_{562} domain. Therefore, the cytochrome domain alone was used as a search model using the EPMR[47] and PHASER[48] programs. The resulting placement was identical and confirmed further by the anomalous signal for iron, with a peak at the expected site in the anomalous difference Fourier map. The EGFP domain of the third molecule was located using a difference Fourier map calculated with phases for the already placed models. Refinement of the completed molecule was carried out using the REFMAC program,[49] with an adjusted library for heme to enforce expected geometry and particularly planarity. Refinement was done with separate TLS parameters for the six domains in the asymmetric unit, with separate (loose) NCS restraints for GFP and cytochrome domains. Data collection and refinement statistics are summarized in Table 2. Small angle X-ray scattering was performed, and data were analyzed as outlined in the Supporting Information.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Methods, Table S1, and Figures S1−S11. This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

The structure of CG6 has been submitted to the PDB under the code 3U8P.

■ AUTHOR INFORMATION

Corresponding Author

jonesdd@cardiff.ac.uk; mbochtler@iimcb.gov.pl

Present Addresses

[#]International Institute of Molecular and Cell Biology, Trojdena 4, 02−109 Warsaw, Poland.

[∇]Department of Anatomy & Structural Biology, Gruss-Lipper Biophotonics Center, Albert Einstein College of Medicine, Bronx, NY 10461.

[○]Max-Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

J.A.J.A. was supported by a BBSRC CASE studentship in collaboration with Merck KGaA. This work was supported by BBSRC grants BB/E001084 and BB/FOF/263 to D.D.J. M.B. and H.C. acknowledge funding from the EC FP7 Grant (HEALTH-PROT, GA No 229676). H.C. also acknowledges financial support from the International Institute of Molecular and Cell Biology, Warsaw, Poland. We are grateful to Prof. Thomas Sorensen and Dr. James Sandy (Diamond Light Source, Didcot, UK) for granting beamtime and assisting during data collection at the IO2 beamline at the Diamond Light Source, Didcot, UK, and Dr. Dmitri Svergun at X33 of EMBL, DESY, Hamburg, Germany. We would like to thank Dr. Roger Chittock for advice on fluorescence lifetime measurements.

■ REFERENCES

(1) Astier, Y.; Bayley, H.; Howorka, S. *Curr. Opin. Chem. Biol.* **2005**, *9*, 576.

(2) Channon, K.; Bromley, E. H.; Woolfson, D. N. *Curr. Opin. Struct. Biol.* **2008**, *18*, 491.

(3) Ambroggio, X. I.; Kuhlman, B. *Curr. Opin. Struct. Biol.* **2006**, *16*, 525.

(4) Koide, S. *Curr. Opin. Biotechnol.* **2009**, *20*, 398.

(5) Lim, W. A. *Curr. Opin. Struct. Biol.* **2002**, *12*, 61.

(6) Ostermeier, M. *Curr. Opin. Struct. Biol.* **2009**, *19*, 442.

(7) Bornberg-Bauer, E.; Beaussart, F.; Kummerfeld, S. K.; Teichmann, S. A.; Weiner, J., 3rd *Cell. Mol. Life Sci.* **2005**, *62*, 435.

(8) Aroul-Selvam, R.; Hubbard, T.; Sasidharan, R. *J. Mol. Biol.* **2004**, *338*, 633.

(9) Fastrez, J. *ChemBioChem* **2009**, *10*, 2824.

(10) Ferraz, R. M.; Vera, A.; Aris, A.; Villaverde, A. *Microb. Cell Fact.* **2006**, *5*, 15.

(11) Ostermeier, M. *Protein Eng. Des. Sel.* **2005**, *18*, 359.

(12) Edwards, W. R.; Busse, K.; Allemann, R. K.; Jones, D. D. *Nucleic Acids Res.* **2008**, *36*, e78.

(13) Edwards, W. R.; Williams, A. J.; Morris, J. L.; Baldwin, A. J.; Allemann, R. K.; Jones, D. D. *Biochemistry* **2010**, *49*, 6541.

(14) Guntas, G.; Mansell, T. J.; Kim, J. R.; Ostermeier, M. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 11224.
(15) Guntas, G.; Mitchell, S. F.; Ostermeier, M. *Chem. Biol.* **2004**, *11*, 1483.
(16) Huber, R. *EMBO J.* **1989**, *8*, 2125.
(17) Scholes, G. D.; Fleming, G. R.; Olaya-Castro, A.; van Grondelle, R. *Nat. Chem.* **2011**, *3*, 763.
(18) Page, C. C.; Moser, C. C.; Dutton, P. L. *Curr. Opin. Chem. Biol.* **2003**, *7*, 551.
(19) Palmer, A. E.; Qin, Y.; Park, J. G.; McCombs, J. E. *Trends Biotechnol.* **2011**, *29*, 144.
(20) Ponka, P. *Am. J. Med. Sci.* **1999**, *318*, 241.
(21) Willis, K. J.; Szabo, A. G.; Zuker, M.; Ridgeway, J. M.; Alpert, B. *Biochemistry* **1990**, *29*, 5270.
(22) Takeda, S.; Kamiya, N.; Arai, R.; Nagamune, T. *Biochem. Biophys. Res. Commun.* **2001**, *289*, 299.
(23) Arnesano, F.; Banci, L.; Bertini, I.; Faraone-Mennella, J.; Rosato, A.; Barker, P. D.; Fersht, A. R. *Biochemistry* **1999**, *38*, 8657.
(24) Feng, Y.; Sligar, S. G.; Wand, A. J. *Nat. Struct. Biol.* **1994**, *1*, 30.
(25) Choi, J. W.; Nam, Y. S.; Lee, B. H.; Ahn, D. J.; Nagamune, T. *Curr. Appl. Phys.* **2006**, *6*, 760.
(26) Hay, S.; Wallace, B. B.; Smith, T. A.; Ghiggino, K. P.; Wydrzynski, T. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 17675.
(27) Takeda, S.; Kamiya, N.; Nagamune, T. *Anal. Biochem.* **2003**, *317*, 116.
(28) Della Pia, E. A.; Chi, Q.; Jones, D. D.; Macdonald, J. E.; Ulstrup, J.; Elliott, M. *Nano Lett* **2011**, *11*, 176.
(29) Della Pia, E. A.; Elliott, M.; Jones, D. D.; Macdonald, J. E. *ACS Nano* **2012**, *6*, 355.
(30) Zuo, P.; Albrecht, T.; Barker, P. D.; Murgida, D. H.; Hildebrandt, P. *Phys. Chem. Chem. Phys.* **2009**, *11*, 7430.
(31) Robinson, C. R.; Liu, Y.; Thomson, J. A.; Sturtevant, J. M.; Sligar, S. G. *Biochemistry* **1997**, *36*, 16141.
(32) Rice, J. K.; Fearnley, I. M.; Barker, P. D. *Biochemistry* **1999**, *38*, 16847.
(33) Zhang, J.; Campbell, R. E.; Ting, A. Y.; Tsien, R. Y. *Nat. Rev. Mol. Cell. Biol.* **2002**, *3*, 906.
(34) Van Engelenburg, S. B.; Palmer, A. E. *Curr. Opin. Chem. Biol.* **2008**, *12*, 60.
(35) Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11241.
(36) Choi, J.; Nam, Y.; Choi, H.; Lee, W. H.; Kim, D.; Fujihira, M. *Synth. Met.* **2002**, *126*, 159.
(37) Bogdanov, A. M.; Mishin, A. S.; Yampolsky, I. V.; Belousov, V. V.; Chudakov, D. M.; Subach, F. V.; Verkhusha, V. V.; Lukyanov, S.; Lukyanov, K. A. *Nat. Chem. Biol.* **2009**, *5*, 459.
(38) Potterton, L.; McNicholas, S.; Krissinel, E.; Gruber, J.; Cowtan, K.; Emsley, P.; Murshudov, G. N.; Cohen, S.; Perrakis, A.; Noble, M. *Acta Crystallogr. Sect. D* **2004**, *60*, 2288.
(39) Baldwin, A. J.; Busse, K.; Simm, A. M.; Jones, D. D. *Nucleic Acids Res.* **2008**, *36*, e77.
(40) Jones, D. D. *Nucleic Acids Res.* **2005**, *33*, e80.
(41) Edwards, W. R.; Williams, A. J.; Morris, J. L.; Baldwin, A. J.; Allemann, R. K.; Jones, D. D. *Biochemistry* **2010**, *49*, 6541.
(42) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509.
(43) Doi, N.; Yanagawa, H. *FEBS Lett.* **1999**, *457*, 1.
(44) Baldwin, A. J.; Arpino, J. A.; Edwards, W. R.; Tippmann, E. M.; Jones, D. D. *Mol. Biosyst.* **2009**, *5*, 764.
(45) Lebedev, A. A.; Vagin, A. A.; Murshudov, G. N. *Acta Crystallogr. Sect. D* **2008**, *64*, 33.
(46) Emsley, P.; Cowtan, K. *Acta Crystallogr. Sect. D* **2004**, *60*, 2126.
(47) Kissinger, C. R.; Gehlhaar, D. K.; Fogel, D. B. *Acta Crystallogr. Sect. D* **1999**, *55*, 484.
(48) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Crystallogr.* **2007**, *40*, 658.
(49) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr Sect. D* **1997**, *53*, 240.