Supporting Information

Structural basis for efficient chromophore communication and energy transfer in a constructed didomain protein scaffold.

James A. Arpino[1], Honorata Czapinska[2], Anna Piasecka[3]§, Wayne R. Edwards[1], Paul Barker[4], Michal J. Gajda[5], Matthias Bochtler[1,2]* & D. Dafydd Jones[1*].

[1] School of Biosciences, Main Building, Park Place, Cardiff University, Cardiff CF10 3AT, UK. [2] School of Chemistry, Main Building, Park Place, Cardiff University, Cardiff CF10 3AT, UK. [3] School of Medicine, Cardiff University, UHW Main Building, Heath Park, Cardiff CF14 4XN, UK. [4] Department of Chemistry, University of Cambridge, Lensfield Road Cambridge, UK CB2 1EW, UK. [5] EMBL c/o DESY, 22603 Hamburg, Germany.

Corresponding authors: Dr D. Dafydd Jones, School of Biosciences, Main Building, Park Place, Cardiff University, Cardiff CF10 3AT, UK. Email: jonesdd@cardiff.ac.uk. Tel. +44 29 20874290. Dr Matthias Bochtler, International Institute of Molecular and Cell Biology, Trojdena 4, 02-109 Warsaw, Poland. Email: mbochtler@iimcb.gov.pl. Tel. +48 22 597 07 32

Supporting Methods

Library construction

(1) Transposon-insertion library

Insertion of the engineered transposon MuDel into the *eGFP* gene encoding enhanced green fluorescent protein (EGFP) residing within the pNOM-XP3 plasmid was performed using an *in vitro* transposition and selection procedure described previously (1).

(2) Construction of domain insert cassette,

To allow the subsequent selection of chimeric genes, a DNA cassette containing the *cybC* gene fragment encoding the mature cytochrome b_{562} (cyt b_{562}) was engineered to incorporate the kanamycin resistance (*kan*^{r}) gene expression element as a selectable marker. *Kan*^{r} was inserted within the PstI site located in *cybC*. Any subsequent digestion with PstI would remove precisely the *kan*^{r} expression element and re-ligation would reconstitute *cybC*. The *kan*^{r} expression element from pEntranceposon F-759 (Finnzymes) was amplified by PCR using the primers WREkif030 (5'-GAAGCGCAGGCTGCTGCAGCCAGCCACCGTTTAAACGGATC-3', PstI site underlined) and WREkir032 (5'-GCGGGTCGTTTTCAGTTGCTCTGCAGTGTGTCGGCGGCC-3', PstI site underlined). The 1116 bp PCR product was mixed with pPB10 (2) and the fragments were amplified by splice-by-overlap-extension PCR using primers WREcbf011 (5'-GATCTTGAAGACAATATGGAAACCC-3') and WREkir032 to generate a 1375 bp PCR product comprising the 5' fragment of the *cybC* gene fused to the *Kan*^{r} expression element, or with primers WREkif030 and WREcbr012 (5'-ACGATACTTCTGGTGATAGGCGT-3') to generate a 1143 bp PCR product comprising the kanamycin-resistance cassette fused to the 3' fragment of *cybC*. The two fragments were mixed and amplified by PCR using primers WREcbf011 and WREcbr012 to generate the full-length (1402 bp) *cybC-kan*^{r} cassette. The PCR product was subsequently cloned into the pNOM plasmid (3) between the EcoRV site and subsequently sequenced to confirm its identity. Sequencing reactions were carried out at the Molecular Biology Support Unit, Cardiff University.

(3) Domain insertion

In preparation for the domain insertion step, the cloned *cybC-kan*^{r} cassette was amplified by PCR with different combinations of oligonucleotide primers outlined in Supporting Table S1. The different primers incorporated GlyGlySer encoding linkers and additional randomised bases to correct for out of frame insertions. The *cybC-kan*^{r} cassettes for each reading frame variation were ligated between the breaks introduced into *eGFP* by removal of randomly inserted MuDel from the EGFP library (step 1). The blunt end ligation was performed using the T4 Quick Ligase enzyme (NEB) with 111 ng of the cassette and 75 ng of the digested EGFP domain acceptance library in a 1:3 molar ratio of plasmid to insert. The products of the ligation reaction where subsequently used to transform *E. coli* DH5α with an estimated transformation efficiency of 1.1 x 10^6 colonies per μg of DNA in the ligation reaction. The transformed cells were grown in LB broth supplemented with 25 μg/ml kanamycin overnight at 37°C. Library DNA was isolated from the cultures and digested with PstI to remove the *kan*^{r} expression element. The resulting 3195 bp product was isolated after agarose gel electrophoresis and recircularised using the T4 Quick ligase system (NEB). The ligation reaction was used to transform *E. coli* BL21 (DE3) (transformation efficiency of 1 x 10^7 colonies per μg DNA in the ligation reaction). The transformed cells were plated on M9 minimal media agar supplemented with 100 μg/ml ampicillin and 150 μM isopropyl β-D-1-

thiogalactopyranoside (IPTG) and grown at 28°C for up to 48 hrs then stored at 4°C. The plates were irradiated with a UV transilluminator and fluorescent colonies marked for sequence analysis. Variant CG6 did not contain the GlyGlySer linker at the C-terminal end of the cyt b_{562} insert. The origin of this remains unknown but may have arisen during oligonucleotide synthesis.

Analysis of fusion proteins.

Protein production

Cell lysates containing CG1, CG6 and CG12 were produced as followed. The chimeric variants were expressed in the Tuner™ strain of *E. coli* BL21 (DE3) grown in M9 minimal media supplemented with 100 μg/ml ampicillin. Protein production was induced by the addition of 300 μM IPTG and the cells incubated at 20°C for a further 48 hr. After harvesting, the cells were resuspended in 20 ml 50 mM Tris-HCl pH 8.0 (buffer A) supplemented with 1 mM phenylmethanesulfonylfluoride (PMSF) and 1 mM ethylenediaminetetraacetic acid (EDTA) and then lysed using a French pressure cell press. The lysate was then centrifuged (20000 rpm in Beckman JA20 rotor for 30 mins) to pellet cell debris. The supernatant was decanted and stored at 4°C prior to analysis.

The production and subsequent purification of EGFP, CG1, CG6 and CG12 was performed as follows. LB Broth (15 ml) supplemented with 100 μg/ml ampicillin was inoculated with a single colony of *E. coli* BL21 (DE3) Tuner™ containing the relevant plasmid (pNOM-EGFP, pNOM-CG1, pNOM-CG6 or pNOM-CG12) to generate a starter culture and incubated for 6 hrs at 37°C. A 1/100 dilution of the starter culture was used to inoculate 1l M9 minimal media supplemented with 100μg/ml ampicillin and grown at 20°C overnight until an optical density of $A_{600} = 0.4$ was achieved. The overnight culture was supplemented with 22 mM glycerol and protein expression induced by the addition of 300 μM IPTG. The culture was incubated for a further 48 hrs at 20°C. The 1l culture was then harvested by centrifugation (3000 x g for 20 mins) and the pellet resuspended in 20 ml buffer A supplemented with 1 mM PMSF and 1 mM EDTA. The cells were lysed by French press using a chilled pressure cell. The lysate was then centrifuged (20000 rpm in Beckman JA20 rotor for 30 mins) to pellet any cell debris and the supernatant was decanted and stored at 4°C. The cell lysate was applied to a Q-Sepharose (GE healthcare) ion exchange column and elution monitored at 280 nm and 488 nm. Pooled fractions were then subjected to ammonium sulphate precipitation to further purify and concentrate the protein sample. An initial ammonium sulphate concentration of 45% (w/v) was used to precipitate unwanted proteins from solution. Further addition of ammonium sulphate to a final concentration of 75% (w/v) was carried out to precipitate EGFP or CG variants from solution. The precipitate was resuspended in buffer A (5ml) and the protein solution was then applied to a SP Superdex 200 gel filtration column (GE Healthcare) with elution monitored at 280 nm and 488 nm. The purified protein sample was finally stored in buffer A supplemented with 150 mM NaCl.

Fluorescence spectroscopy.

Excitation and emission spectra of cell lysates and purified protein were recorded 50 mM Tris-HCl, pH 8.0 25°C using a Varian Cary Eclipse fluorimeter with cuvette dimensions of 5 x 5 mm, 10 nm slit widths and a medium scan rate. For EGFP and each of the chimeric proteins, emission spectra were recorded after excitation at 488 nm and excitation spectra recorded by monitoring emission at 511 nm. Samples of each protein were prepared by either diluting the cell lysates into buffer A supplemented with 150 mM NaCl to give a final emission intensity of 100 a.u. at 511 nm or using 20 nM purified protein (equivalent to a fluorescence emission intensity of 100 a.u.). The influence of heme on fluorescence was

determined by titration of heme (0 – 1 μM) into the cuvette containing the protein sample and monitoring the fluorescence emission intensity at 511 nm after excitation at 488 nm. Oxidising conditions were maintained by the addition of 1 mM KNO_3 and reducing conditions induced by the addition of 1 mM ascorbic acid. For variant CG6, the effect of switching from reducing to oxidising conditions was determined by diluting the purified protein in buffer A supplemented with 1 mM ascorbic acid to give a final concentration of 20 nM. Heme was then added to the solution at approximately equimolar concentration to the protein variant. After a brief delay (10 min), 0.02% (v/v) H_2O_2 was added to the sample. The fluorescence emission at 511 nm was monitored throughout the experiment.

Absorbance spectroscopy.

UV-visible absorbance spectra were recorded with a Hewlett Packard diode array spectrophotometer with a 1 cm path length. Purified protein was diluted into buffer A supplemented with 150 mM NaCl. Absorbance spectrum of 8 μM CG6 was recorded after the addition of 1 mM ascorbic acid (to induce reducing conditions) and in the presence of a slight excess of heme (10 μM). Absorbance spectra were also recorded at 2 hrs, 4 hrs and 20 hrs after the addition of 0.02% (v/v) H_2O_2.

Quantum yield determination.

Quantum yields for EGFP and the apo forms of CG1, CG6 and CG12 were calculated using fluorescein as a reference. Fluorescein (in 0.1 M NaOH) and the chimeras (in Buffer A) were prepared with A_{488} of 0.02 in a 10 mm pathlength cuvette. Emission spectra were measured after excitation at 488 nm using a 5mm pathlength cuvette with excitation and emission slits of 2.5 nm. Integrated emission intensity between 500 and 650 nm was calculated and used in the following formula to generate quantum yield values

$$\phi_x = \phi_{st} . (Area_x/Area_{st}).(\eta^2_x/\eta^2_{st})$$

where the ϕ_x and ϕ_{st} refer to the fluorescence quantum yield of the sample and fluorescein standard, respectively. $Area_x$ and $Area_{st}$ are the integrated emission intensities for the sample and fluorescein standard, respectively. η_x and η_{st} is the refractive index of the solvent for the sample and fluorescein standard, respectively. The refractive index correction here was negligible as 0.1 M NaOH and aqueous buffers differ in refractive index by <1%.

Heme Binding Affinity

The heme binding affinity under oxidising conditions was calculated from the heme-mediated fluorescence quenching data. The value in the absence of heme (initial data point) was equal to 0 heme bound (apo-protein); the data point at the highest saturating heme concentration were no further decrease in fluorescence was observed was scaled to 1 (all binding sites occupied). The scaled data was plotted against heme concentration and least squares non-linear regression was used to fit the data to the following binding equation:

$$Holo_n = (Holo_{max}[Heme])/(K_D + [Heme])$$

where $Holo_n$ is the proportion of holo-protein derived from fluorescence quenching measurement (0-1), $Holo_{max}$ holo-protein fully occupied with heme.

Circular dichroism spectroscopy.

Circular dichroism spectroscopy was performed using a Chirascan CD spectrometer (Applied Photophysics) between 190 and 250 nm at a scan rate of 1nm/sec in a 1 mm pathlength quartz cuvette. Protein samples (5 μM) were prepared in 10 mM sodium phosphate buffer (pH 8.0) with fresh ascorbic acid (1 mM). CD spectra were measured in the absence and presence (5 μM) of heme.

Size exclusion chromatography. Gel filtration standards (Biorad) were applied to a Superdex™ 200 column (20 ml bed volume, 0.5 ml/min flow rate), as per the manufacturers guidelines, with protein elution monitored at 280 nm. A standard curve was generated from the plot of LogMw against K_{av}, where $K_{av} = (V_e-V_o)/(V_t-V_o)$, V_e is the elution volume, V_o is the void volume and V_t is the total volume. Protein samples (EGFP, CG1 or CG6) were prepared in 50 mM Tris-HCl, pH 8.0, 150 mM NaCl in final concentrations of 25, 50 or 100 μM and applied to the Superdex™ 200 column with protein elution monitored by absorbance at 488 nm. Elution volumes (V_e) were determined for each sample and K_{av} values calculated. Using the standard curve estimated molecular weights could be determined for each protein sample.

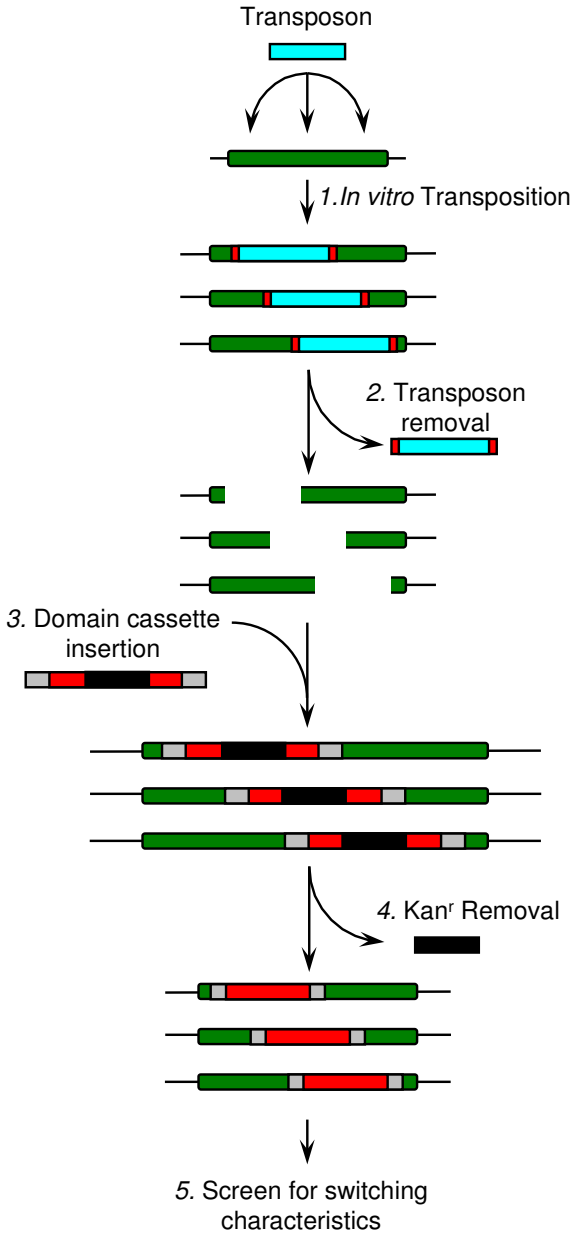
Small angle X-ray scattering: Small angle X-ray scattering data were collected at beamline X33 at the DORIS ring, DESY, Hamburg in scattering length range of 0.067 nm^{-1} Up to 6.024 nm^{-1} by 8 exposures of 15 seconds each with Pilatus 1M detector in cell temperature of 25C. Data were radially averaged to 2176 data points, background corrected and processed in the usual manner, checking for radiation damage between different 15 second long frames, using AutoPilatus, Beamline Metaserver. Small angle X-ray scattering data were recorded at various protein concentrations between 1 and 5mg/ml. As the dependence of relative intensities on the scattering vector was independent of protein concentration, all further analysis was done with the scattering data for the highest concentration. Scattering for a given atomic coordinates was calculated and fitted to the experimental data using the CRYSOL program (4). Ab initio shape determination was done with the DAMMIF program (5).

Supporting Table S1. Oligonucleotide sequences for generating the cyt b_{562} insertion cassette

Primer name	Sequence (5' to 3')[a]	Primer pairs[b]
DDJdi023	GGCGGTAGCGCAGATCTTGAAGACAATATGGA	ORF1-Sen
DDJdi024	GCTGCCACCCCTATACTTCTGGTGATAGGCGT	ORF1-As
JAJA003	NGGTGGGAGCGCAGATCTTGAAGACAATATGGA	ORF2-Sen
JAJA004	NNGCTCCCACCCCTATACTTCTGGTGATAGGCGT	ORF2-As
JAJA005	NNGGTGGGAGCGCAGATCTTGAAGACAATATGGA	ORF3-Sen
JAJA006	NGCTCCCACCCCTATACTTCTGGTGATAGGCGT	ORF3-As

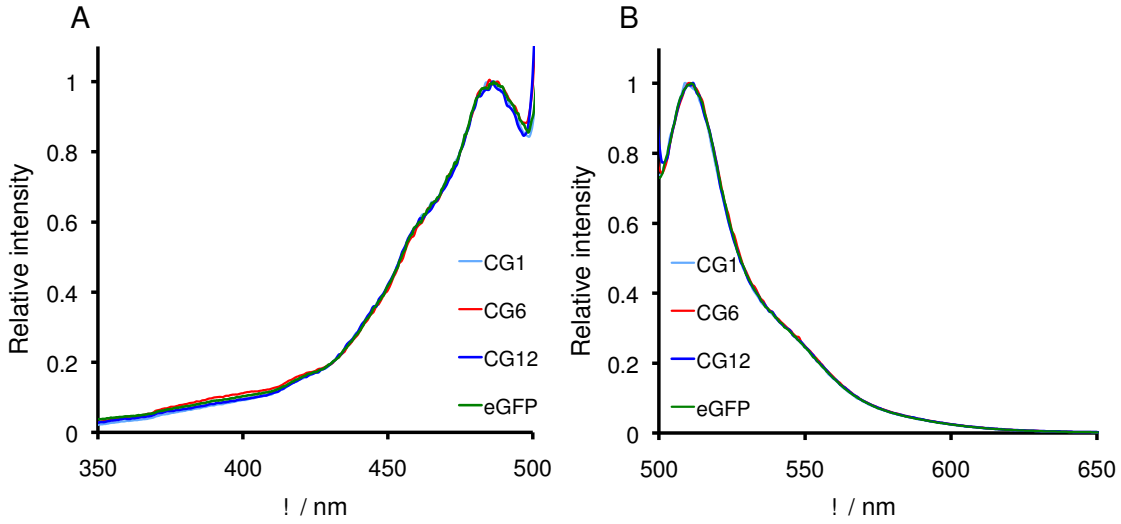
[a] Nucleotides shown in red encode linker sequences that join cytochrome to EGFP and encodes the GlyGlySer amino acid sequence. Nucleotides underlined are randomised additional bases to correct for insertions that would otherwise lead to frameshifts. N represents an equimolar mixture of the four nucleotides A,G,C and T. The presence of the N nucleotides results in an additional amino acid comprising the linker.

[b] ORFx refers to open reading frame primer pair, with Sen and As referring to the sense and antisense primer, respectively, that comprise the pair.

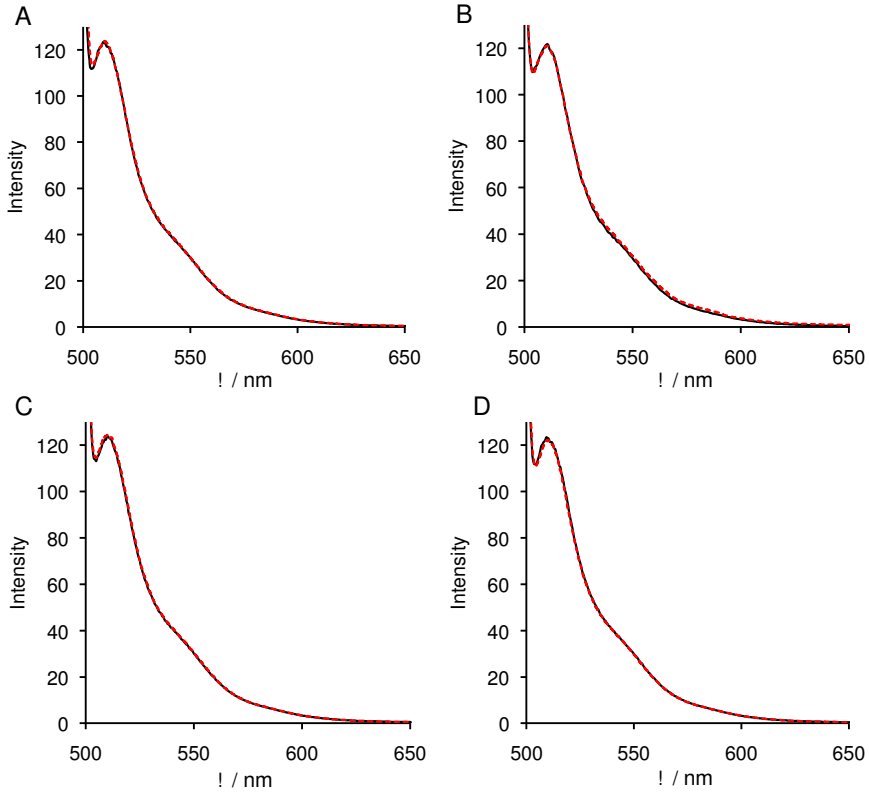


Supporting Figure S1. Transposon-based directed evolution technique for the creation of integral fusion proteins. (1) *In vitro* transposition of an engineered transposon (MuDel; cyan block) with low target site specificity allows insertion events to take place randomly throughout the target gene, *EGFP* (green block). (2) Subsequent removal of the transposon from the library by restriction digestion generates single random breaks within *eGFP*. (3) A DNA cassette encoding cyt b_{562} (red block) is then inserted into the single breaks within the library DNA with the linker encoded as part of the cassette (grey blocks). Also present in the cassette is the *Kan*^{r} gene that confers resistance towards kanamycin (black block) and allows selection of domain insert chimeric gene variants (4) Removal of the *Kan*^{r} section followed by self ligation reconstitutes the full length uninterrupted chimeric gene encoding the cyt

b_{562}-EGFP variants. (5) The library of integral fusion proteins is then screened for variants displaying switching characteristics.

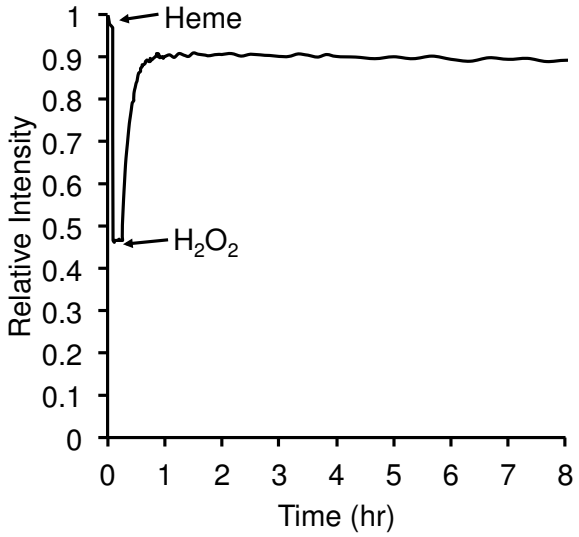


Supporting Figure S2. Excitation (A) and emission (B) spectra of selected cyt b_{562}-EGFP chimeras CG1, CG6 and CG12, together with EGFP in the absence of heme. The spectrum for each variant is coloured as indicated in the figure. Excitation and emission spectra were recorded using cell lysates as outlined above in the Supporting Methods.

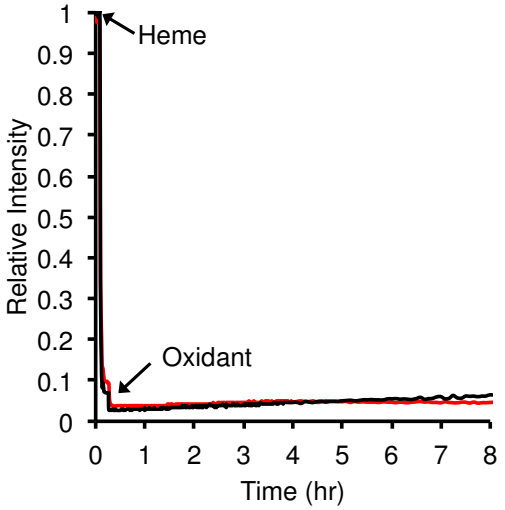


Supporting Figure S3. The fluorescence emission properties of EGFP under various conditions. (A) EGFP (solid black line), EGFP plus 1 mM ascorbic acid (dashed red line), EGFP plus 1 mM KNO_3 (black dotted line). (B) EGFP (solid black line); EGFP plus 0.02%

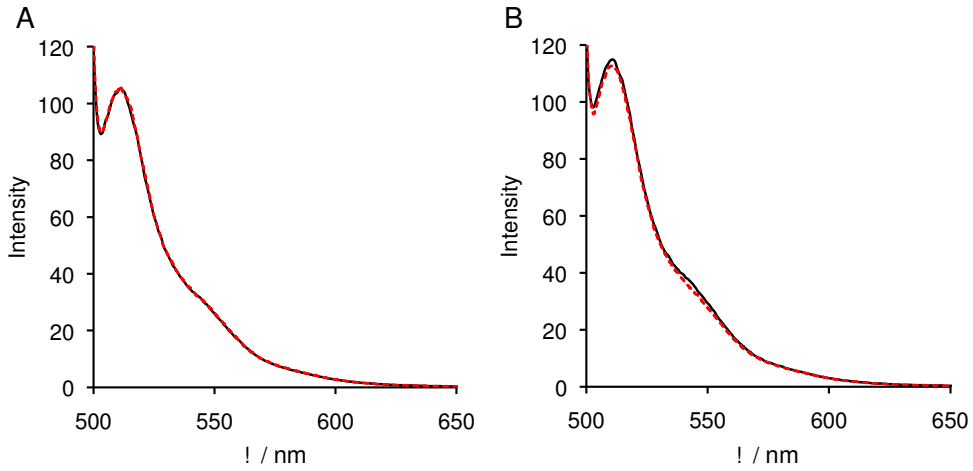
(v/v) H_2O_2 (red dashed line). (C) EGFP (solid black line); EGFP plus 1 μM heme (reducing conditions) (dashed red line). (D) EGFP (black line); EGFP plus 1 μM heme (oxidised conditions) (dashed red line). In all cases, 20 nM purified EGFP in buffer A supplemented with 150 mM NaCl was used. Emission spectra were recorded after excitation at 488 nm.



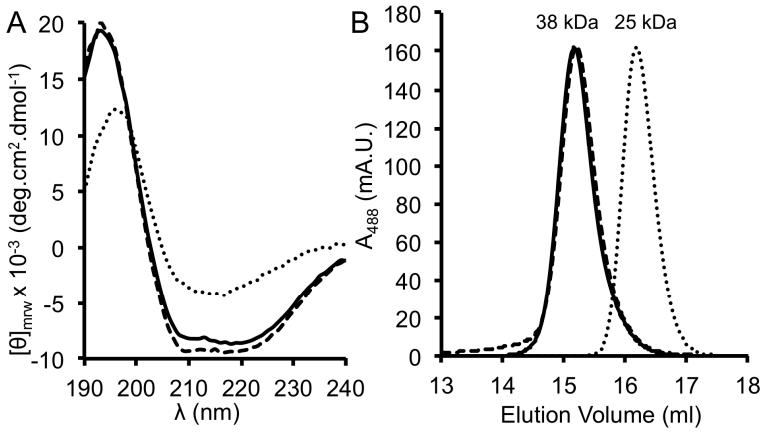
Supporting Figure S4. Oxidant-induced fluorescence switching of CG1. Heme (30 nM) was added to 20 nM purified apo-CG1 in the presence of ascorbate (1 mM) to induce quenching. At 15 min, 0.02% (v/v) H_2O_2 was added to stimulate switching from reducing to oxidising conditions.



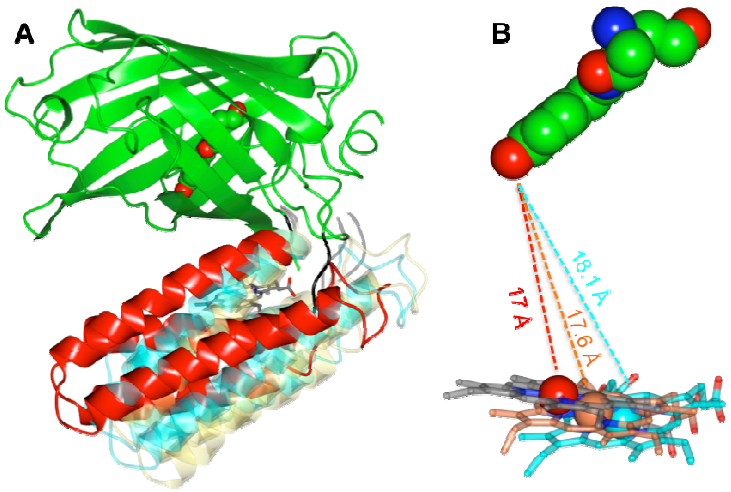
Supporting Figure S5. The effect of different oxidants on fluorescence signal gain for CG6. Heme (30 nM) was added to 20 nM purified protein in the presence of ascorbate (1 mM) to induce quenching. At 15 min 10 mM KNO_3 (black line) or 0.02% (v/v) NaOCl (red line) was added to stimulate a switch from reducing to oxidising conditions.



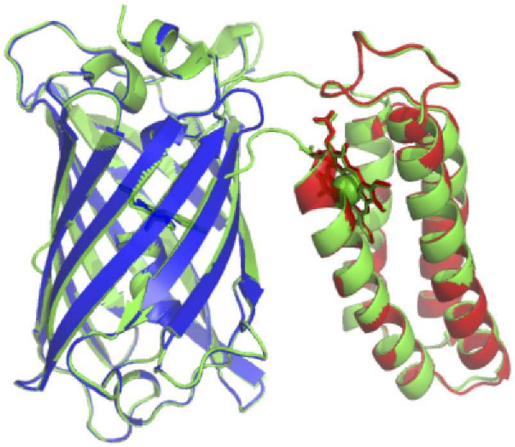
Supporting Figure S6. The effect of ascorbic acid and H_2O_2 on the fluorescence emission of cyt b_{562}-GFP variant CG6. (A) Emission spectra for apo-CG6 in the absence (black solid line) or presence (red dashed line) of 1 mM ascorbic acid. (B) Emission spectra for apo-CG6 in the absence (black solid line) or presence (red dashed line) of 0.02% (v/v) H_2O_2. Emission spectra were measured after excitation at 488 nm.



Supporting Figure S7. Structural changes on heme binding to CG6. (A) CD spectra and (B) SEC of EGFP (dotted lines), apo-CG6 (solid line) and holo-CG6 (dashed line).

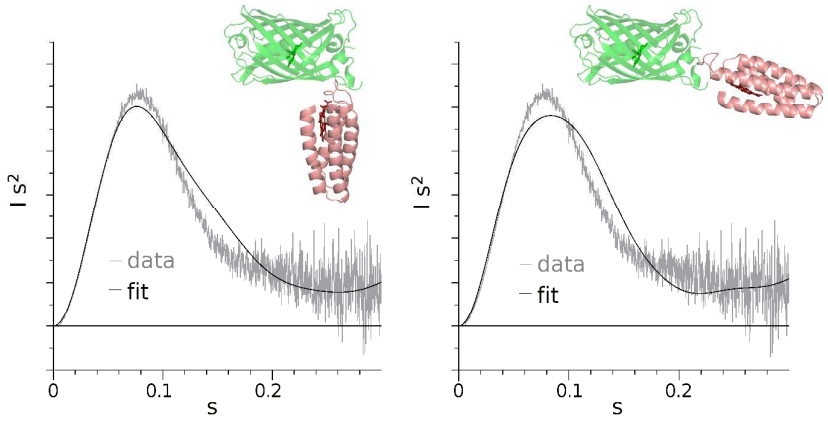


Supporting Figure S8. Overlay of the 3 molecules making up the asymmetric unit. (A) Overlay of whole molecules. The proteins were aligned based on the EGFP domain (coloured green) with the chromophore shown in spacefill. The structure used in the main text as a representation of CG6 has the cyt b_{562} domain coloured red, with the other two coloured blue and orange. (B) Overlay of the chromophores. The iron atoms are shown as spheres and coloured according the scheme for the cyt b_{562} domain in A. The distances between the heme iron and the hydroxyl group of the EGFP chromophore in each molecule are shown on the diagram.

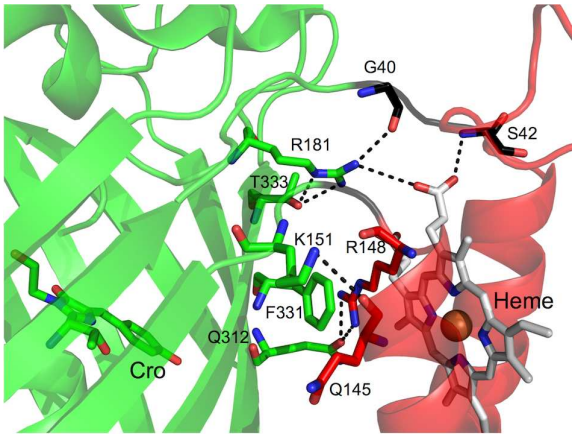


RMSD	EGFP	cyt b_{562}
CG6 Ch A	0.31Å	0.78Å
CG6 Ch B	0.32Å	0.82Å
CG6 Ch C	0.38Å	0.78Å

Supporting Figure 9. Superimposition of CG6 (green) with EGFP (blue) and cyt b_{562} (red). The RSMD for each of the three CG6 molecules in the unit cell against EGFP and cyt b_{562} is also shown.



Supporting Figure S10. Alternative fits of predicted curves (black lines) to measured SAXS data (grey lines) corresponding to other possible domain arrangements in solution, depicted in the insets. The poor agreement is reflected by chi of 1.92 for the "right-angle" model (left panel) and chi of 3.03 for the "elongated" model (right panel).



Supporting Figure S11. Interdomain and linker interactions for molecule B. Polar interactions between residues are shown as dashed lines and residues annotated as outlined in the figure.

Supporting References

1. Baldwin AJ, Arpino JA, Edwards WR, Tippmann EM, & Jones DD (2009) Expanded chemical diversity sampling through whole protein evolution. Mol Biosyst 5:764-766.
2. Barker PD, Nerou EP, Freund SM, & Fearnley IM (1995) Conversion of cytochrome b_{562} to c-type cytochromes. Biochemistry 34:15191-15203.
3. Jones DD (2005) Triplet nucleotide removal at random positions in a target gene: the tolerance of TEM-1 β-lactamase to an amino acid deletion. Nucleic Acids Res 33:e80.
4. Svergun DI, Barberato C, & Koch MHJ (1995) CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. J. Appl. Cryst. 28:768-773.
5. Franke D & Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J. Appl. Cryst. 42:342-346.