Journal Name

Dynamic Article Links ►

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Constructing a man-made, oxygen binding c-type cytochrome maquette *in vivo*.

Anderson, J. L. Ross[a,b,*]; Armstrong, Craig T.[a,*]; Kodali, Goutham[b]; Lichtenstein, Bruce R.[b]; Watkins, Daniel W.[a]; Mancini, Joshua A.[b]; Boyle, Aimee L.[c]; Farid, Tammer, A.[b]; Crump, Matthew P.[c]; Moser, Christopher C.[b]; Dutton, P. Leslie[b].

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
DOI: 10.1039/b000000x

The successful use of man-made proteins to advance synthetic biology requires both the fabrication of functional artificial proteins in a living environment, and the ability of these proteins to interact productively with other proteins and substrates in that environment. Proteins made by the maquette method integrate sophisticated oxidoreductase function into evolutionarily naive, non-computationally designed protein constructs with sequences that are entirely unrelated to any natural protein. Nevertheless, we show here that we can efficiently interface with the natural cellular machinery that covalently incorporates heme into natural cytochromes *c* to produce *in vivo* an artificial *c*-type cytochrome maquette. Furthermore, this *c*-type cytochrome maquette is designed with a displaceable histidine heme ligand that opens to allow functional oxygen binding, the primary event in more sophisticated functions ranging from oxygen storage and transport to catalytic hydroxylation. To exploit the range of functions that comes from the freedom to bind a variety of redox cofactors within a single maquette framework, this *c*-type cytochrome maquette is designed with a second, non-heme C, tetrapyrrole binding site, enabling the construction of an elementary electron transport chain, and when the heme C iron is replaced with zinc to create a Zn porphyrin, a light-activatable artificial redox protein. The work we describe here represents a major advance in *de novo* protein design, offering a robust platform for new *c*-type heme based oxidoreductase designs and an equally important proof-of-principle that cofactor-equipped man-made proteins can be expressed in living cells, paving the way for constructing functionally useful man-made proteins *in vivo*.

Introduction

Man-made proteins have an important role to play in the synthetic biology goal of constructing functional parts and devices for incorporation into either explicitly biological or artificial organisms or systems.[1] The program to integrate sophisticated oxidoreductase function into evolutionarily naive, non-computationally designed protein constructs is called the maquette approach.[2,3] By beginning with a generic protein sequence designed only to adopt a simple helical bundle fold and iteratively adding engineering elements onto this stripped-down protein chassis, the maquette approach allows the designer to avoid the layers of complexity that hinder the redesign of natural protein scaffolds. Man-made protein maquettes self-assemble *in vitro* with a wide range of the redox cofactors seen in nature, including hemes, chlorins, metal ions, flavins and quinones.[2] However synthetic biology requires that functional artificial proteins and enzymes interact productively with natural proteins and substrates. They must also fully and functionally assemble *in vivo*, meaning that cofactor binding sites must be occupied by the target cofactors to carry out the desired function.

Here we test the ability of a novel man-made protein to interface with natural proteins *in vivo* in order to assemble a functional redox protein. Despite the completely unnatural protein sequence, the natural post-translational machinery of *E. coli* (Fig. 1A)[4] successfully inserts heme B and forms two covalent links between the heme vinyls and protein cysteines to create a synthetic heme C cytochrome with exceptional efficiency. This man-made cytochrome *c* successfully forms a heme oxy-ferrous state with a stability akin to natural oxygen transport proteins containing heme B,[5] but with entirely unrelated sequence or structure. As part of a program to design increasingly sophisticated man-made oxidoreductases, this protein is equipped with an intraprotein electron-transfer chain by including a second non-heme C binding site that self-assembles *in vitro* with heme B. Light activated function is added to this dyad by replacing the Fe of the original heme C with Zn to create a Zn-porphyrin photo-center.

Results

Protein and vector design

We have previously designed a functional, man-made maquette containing heme B that is capable of reversibly binding molecular oxygen (Sequence **1** in Fig. 1B).[6] This maquette not only matches the diatomic ligand exchange kinetics and spectroscopy of natural heme containing globins, but also preferentially binds O_2 over CO. For a more versatile protein capable of taking advantage of the functionally diverse option of placing a range of different cofactors at two distinct sites, we broke the original dimeric symmetry and united the helices with a long, simple connecting loop composed of just glycine and serine residues. A short, stabilizing N-cap sequence was added to the N-terminus of the protein to increase thermal stability by restricting protein motion (sequence referred herein as **1.5**).[7]

We wished to include a site amenable to covalent heme C attachment for the dual purpose of establishing the interaction of this man-made protein with evolved natural redox proteins *in vivo*, and secondly to gain a securely attached cofactor that could

be exploited under harsh *in vitro* conditions. The majority of natural *c*-type cytochrome sequences contain a consensus CX_1X_2CH motif necessary for *in vivo* heme incorporation.[4] We surveyed the non-redundant PDB for structures with *c*-type hemes attached to helices (150 structures, Supplementary information, Table S1) and noted the prevalence of small (A/G) residues at X_2 and a general preference for hydrophobic residues at X_1 (Supplementary information, Fig. S1). We selected CIACH as the *c*-type incorporation motif (Sequence **2** in Fig. 1B) to reflect a balance between maintaining the helicity and structure of the protein, and satisfying the very broad substrate specificity of the promiscuous *E. coli c*-type heme maturation system (Ccm).[4] Furthermore, this selection is consistent with a previous analysis of helical porphyrin-binding sites in heme-containing proteins, where the idealized sequence for the most commonly observed histidine rotamer in helical *c*-type heme sites was identified as CX_1ACH.[8] To eliminate interference with the added N-cap sequence on helix 1, helix 2 was chosen as the site for the consensus motif, with the distal ligand provided by the histidine on helix 4. The histidines on helices 1 and 3 form an auxiliary binding site for tetrapyrroles such as heme B (Fig. 1B).

In addition to the consensus motif, *c*-type cytochromes must be exported to the periplasm for covalent heme incorporation by the Ccm machinery.[4] We cloned the sequence into the periplasmic expression vector pMal-p4x (NEB, USA), replacing the maltose binding protein (MBP) N-terminal fusion partner with a tobacco etch virus protease (TEV) cleavable hexahistidine tag, leaving the N-terminal signal peptide attached *via* the cleavable purification tag in the vector pSHTctmaq (Fig. 1C). The signal peptide facilitates periplasmic export through the Sec translocon[9,10] and is designed for *in vivo* cleavage prior to *c*-type heme incorporation.

In vivo cofactor incorporation

To ensure efficient heme incorporation when expressing *c*-type cytochromes, the *E. coli* Ccm machinery must be co-expressed.[4,10] This can be achieved at low efficiency through anaerobic growth under nitrate, though the highest levels of heme C incorporation in **2** were observed after aerobic expression in *E. coli* BL21(DE3) co-transformed with the pEC86 vector harboring *ccmA-H*.[11] In the absence of the pEC86 plasmid, the low levels of Ccm proteins expressed under aerobic conditions lead to very poor heme incorporation (< 6 %), producing large amounts of apoprotein (Supplementary information, Fig. S2) and reflecting the low levels of CcmE constitutively expressed under aerobic or micro-aerobic conditions. To further probe the mechanism of heme incorporation, pSHTctmaq was co-transformed with pEC864 - a variant of pEC86 harbouring the catalytically inactive H130A mutant of CcmE[12] - into *E. coli* strain EC65 (Δ*ccmE*).[13] The absence of any catalytically active CcmE under these conditions results in a small quantity of *b*-type heme binding (but no *c*-type heme binding) to **2**, highlighting the necessity of CcmE for efficient *c*-type heme incorporation into the maquette. Since heme C can be covalently incorporated into some natural *c*-type cytochromes under reducing conditions *in vitro*,[14,15] we attempted to do so with apo-**2** and both hemin and zinc protoporphyrin IX. We were unable to detect appreciable levels of covalent heme incorporation *in vitro* with either porphyrin, even after a week of incubation, further indicating the necessity of the CcmE protein for efficient maturation. With concomitant pEC86 expression of CcmA-H, the highest levels of expression and heme incorporation (where ~ 90 % of proteins contain heme C) occur during 15 hour aerobic expression in Luria-Bertani broth at 37 ^{o}C, the red color of the cell paste serving as a useful indicator of expression levels. Following cell lysis by sonication and Ni-NTA affinity chromatography, the His-tag was cleaved by TEV protease. To remove residual apo-protein and a small quantity of proteolysed protein, the crude protein was further purified by reverse-phase HPLC. MALDI-TOF mass spectrometry confirmed covalent incorporation of heme C (calculated mass = 15758 Da; observed mass = 15780 Da, **2** + Na^+, Supplementary information, Fig. S3). Treatment of **2** with acidified 2-butanone[16] does not lead to heme partitioning into the organic solvent (Supplementary information, Fig. S4), a characteristic that differs from proteins containing heme B but one that is shared with covalently bound heme C containing proteins. The UV/vis spectrum of the ferrous protein **2** (Fig. 2A) exhibits the expected blue shift in the Soret/alpha/beta absorption bands compared to ferrous protein **1**, consistent with a modification of the vinyl groups on the conjugated porphyrin.[17,18] There is a greater degree of spectral shift in comparison with the single thioether-linked heme in the synthetic *de novo* cytochrome *c* reported by Isogai[17] but a shift similar to that observed by Barker in converting cytochrome b_{562} to a *c*-type cytochrome.[18] Further confirming the nature of the heme species, the reduced pyridine hemochrome spectrum is identical to that of horse heart cytochrome *c* and those reported for natural *c*-type cytochromes[19] (Fig. 2B), and significantly shifted with respect to *b*-type heme containing proteins and the single thioether-linked heme in the synthetic *de novo* cytochrome *c* of Isogai.[17] Ferric and ferrous UV/visible (Fig. 2A) and EPR spectra (Supplementary information, Fig. S5) of **2** are indicative of 6-coordinate bis-histidine ligated heme C.[20,21]

Heme C containing maquette binds O_2

Despite the modifications to the original design (**1**), **2** retains the ability to undergo ligand exchange between the histidine sidechain occupying the 6th coordination site on heme C and either CO or O_2 (Fig. 3A,B). The spectral changes observed on binding CO and O_2 are similar to those observed for the corresponding heme B-bound species of **1**[6] and the oxyferrous absorbance spectrum is consistent with the oxyferrous spectra of the few natural and engineered *c*-type cytochromes that can undergo reversible oxygen binding (Supplementary information, Table S2).[22-24] The lifetime of the oxyferrous species of **2** is similar to those observed for **1**[6] and **1.5** (1, $t_{1/2}$ = 9.4 seconds; 1.5, $t_{1/2}$ = 10 seconds; Supplementary information, Table S3), though autoxidation at 15 ^{o}C occurs less rapidly ($t_{1/2}$ = 15 seconds), indicating a slight stabilization of the oxyferrous species on switching between hemes B and C. This rate is significantly more rapid than the autoxidation rates of the natural globins[25,26] and the natural and engineered *c*-type cytochromes,[27,28] most likely due to the relatively low reduction potential of the hemes in **2** and the other oxygen-binding maquettes,[6] but it is within the range of autoxidation rates exhibited by the oxygen binding and activating cytochromes P450 (Supplementary information, Table S3).[29-31]

Physical chemistry of mixed Heme C/B maquettes.

Titration of one equivalent of heme B into **2** creates a mixed heme C/B maquette (Fig. 4). Analytical ultracentrifugation of the C and C/B maquette showed that both sedimented as monomeric 4-helix bundles with the data fitting to species with molecular weights of 18,340 Da and 17,210 Da (Fig. 5A,B - residuals plotted in Supplementary information, Fig. S6), corresponding to 1.16x and 1.05x the monomer mass respectively (15,758 and 16,374 Da). Circular dichroism spectroscopy indicated both to be highly α-helical with an increase in helicity on heme B ligation (Fig. 5C). Visible CD spectra indicated that the hemes reside in chiral environments in both C and C/B forms. However, ligation of heme B in the auxilliary site confers significantly more thermal stability than does heme C to the heme-free maquette (Fig. 5D; second derivatives of the CD melt data are plotted in Supplementary information, Fig. S7). Using thermal denaturation to probe the protein stability (monitored at 222 nm in the circular dichroism spectrum), the heme C form melting transition at 34 ^{o}C

indicates the heme C does not confer significant stability compared to the apo-protein lacking the CXXCH motif (**1.5**, T_m 42 ^oC, Supplementary information, Fig. S8). Given the thermal stability generally imparted on maquettes by the ligation of heme B,[2] this is a surprising observation and may reflect a non-idealized and strained binding geometry between heme C and the proximal histidine-containing helix. We were unable to determine the T_m of apo-**2** due to the formation of oligomers in the absence of heme C. Binding heme B to the empty site of the heme C form leads to a dramatic 32 ^oC increase in T_m and a significantly more cooperative melting transition. More support for the stabilizing effect of heme B, as opposed to heme C, is found in the broad 1D ^1H NMR spectrum of **2**, an indicator of conformational averaging of the protein scaffold and thus a protein with a mobile, liquid interior (Supplementary information, Fig. S9), compared to the distinct sharpening of the NMR signals on adding heme B.

The additions of a connecting inter-helix loop and thioether linkages to the heme macrocycle raise the heme C redox midpoint in **2** by 54 mV compared to the B heme **1** (-252 to -198 mV vs NHE, Supplementary information, Fig. S10), possibly accounting for the increase in the lifetime of the oxyferrous state. The C and B heme redox potentials in the mixed C/B maquette are -188 mV and -250 mV respectively (Fig. 6). The greater than 60 mV split in the redox potentials means that electron transfer from B to C will be more than 10 times more favorable than the reverse reaction, providing direction to the nascent electron transfer chain.[32]

Conversion to a photoactive maquette

By covalently appending the cofactor to the protein backbone, diverse functionality can be incorporated into the heme C site of the maquette whilst maintaining a robust scaffold with the capacity for a secondary or complementary function at the auxiliary tetrapyrrole binding site. Since *c*-type cytochromes undergo facile demetallation in the presence of hydrofluoric acid and remetallation by zinc to produce light sensitive porphyrins, the heme C in **2** can be modified to create a maquette capable of simple light-activated electron/energy transfer. Using a protocol first devised by Vanderkooi,[33] lyophilized heme C maquette is dissolved in 70% hydrogen fluoride in pyridine, the almost instantaneous color change to purple indicates the formation of the metal-free form. Following removal of the excess hydrogen fluoride and pyridine, reloading with zinc by stirring with an excess of Zn^{2+} at pH 2 produces Zn cytochrome *c* form (Zn(II)(**2**)), which retains the auxiliary tetrapyrrole binding site. The UV/visible spectra and fluorescence excitation and emission spectra of metal-free and Zn porphyrin forms (Fig. 7) are consistent with those obtained for the metal free and zinc horse heart cytochrome *c*.[33] Adding 1 equivalent of heme B resulting in the mixed ZnC/B form (Supplementary information, Fig. S11). This binding of heme B to the auxiliary site quenched the fluorescence of the covalently appended zinc porphyrin by approximately 40 %, indicating the presence of an efficient energy transfer mechanism between chromophores with overlapping emission/absorbance spectra, a feature common to natural light harvesting complexes.[34]

Discussion

We have shown that it is plausible to practice synthetic biology with completely artificial redox proteins that plug in to natural biochemical pathways and are functionalized *in vivo*. Furthermore, with maquettes we can potentially customize inter-protein interactions by adjusting the patterning of the maquette external charges to be complementary with localized regions of natural protein surface charge, independent of the incorporation of hemes in the interior. We have also described a tractable procedure whereby a designer 4-helix bundle capable of non-covalently binding heme cofactors can be converted into a functional artificial *c*-type cytochrome. The design criteria are simple: take an existing heme B binding protein maquette and convert to a single chain 4-helix bundle to break the bundle symmetry and satisfy the required heme C ligation, in this case 6-coordinate bis-histidine ligation; incorporate the consensus CX_1X_2CH motif derived from natural *c*-type cytochrome sequences and co-express in the *E. coli* periplasm with the natural cytochrome *c* maturation genes. Efficient translocation to the periplasm is achieved by appending the natural N-terminal signal sequence from MBP, demonstrating the compatibility of artificial protein sequences with complex cellular machinery, resulting in a very high percentage of the translocated protein. Efficient heme incorporation into the maquette is only achieved in the *E. coli* periplasm in the presence of active, co-expressed heme lyase, CcmE (Supplementary information, Fig. S2).[4]

We caution that working with *in vivo* Ccm maturation proteins may require that our artificial proteins designs are not overly robust and thermally stable. Apo-horse heart cytochrome *c* is largely unfolded under physiological conditions,[35] only attaining the native state on heme B binding or heme C incorporation.[36] A similar situation is likely to exist for the double cysteine mutant of cytochrome b_{562} engineered for *c*-type heme incorporation (cytochrome cb_{562}): wild type apocytochrome b_{562} exists in a molten globule-like state,[37] while the mature cytochrome cb_{562} adopts a native state almost identical to that of the *b*-type heme containing version.[38,39] Such behavior is indicative of a preference of the maturation system for protein substrates with heme binding sites that exhibit disordered or at least molten globular characteristics when unoccupied, thus decreasing the kinetic barrier of chaperone-presented heme binding to the protein (*via* the CcmE heme lyase in the *E. coli*), the first step in covalent heme incorporation.[4] Since our *c*-type cytochrome maquette exhibits a relatively low T_m, we surmise that the high efficiency of heme incorporation can be attributed to the dynamic, molten state of the maquette.

The prokaryotic *E. coli* Ccm proteins are capable of incorporating heme C into a wider variety of *c*-type cytochromes compared to the eukaryotic system, which display marked specificity for eukaryotic cytochromes *c* and c_1 (the only known eukaryotic *c*-type cytochromes).[4] A recent study by Kranz[40] has demonstrated that the *S. cerevisiae* maturation apparatus can process prokaryotic homologues of eukaryotic cytochrome *c*, provided that a sequence recognition motif[41] is included preceding the CXXCH motif at the N-terminus of the apoprotein substrate. Furthermore, the complex of holocytochrome *c* synthase (HCCS), a eukaryotic cytochrome *c* maturation protein responsible for terminal heme transfer and attachment, and apocytochrome *c* is primarily mediated by interactions with heme at the protein:protein interface.[40] Since the heme ligands of the nascent heme C binding site are provided by a histidine from the CXXCH motif of the apocytochrome *c* substrate and a histidine from HCCS, it would appear that effective interfacing with HCCS and thus maturation with this system would require an unfolded apocytochrome substrate. We therefore postulate that the inclusion of the additional recognition sequence to our *c*-type maquette coupled with the partially unfolded nature of the apo-maquette under physiological conditions may enable efficient heme incorporation through this maturation system, thereby facilitating the *in vivo* assembly of artificial *c*-type cytochromes in eukaryotic organisms.

There is only one known example of a natural, oxygen binding *c*-type cytochrome that forms the oxyferrous state in the wild-type protein: Sphaeroides Heme Protein (SHP), a class 1 cytochrome *c*

from *Rhodobacter sphaeroides* and related proteobacteria.[24,28] SHP adopts the fold of a typical cytochrome *c*, but with either the unusual His/Asp or the more common bis-His ligation scheme. Both ligation schemes free up the sixth heme coordination site for diatomic ligand binding when reduced, favoring mono-His ligation in the ferrous state.[42] Two examples also exist of oxyferrous states in engineered variants of cytochrome *c* (horse heart cytochrome *c, S. cerevisiae* iso-1-cytochrome *c*)[22,23,27] where the distal heme-ligating methionine ligands are replaced with alanine, freeing the sixth coordination site for ligand binding. The oxyferrous state is exceptionally long-lived in the engineered iso-1-cytochrome *c,*[27] exhibiting an autoxidation rate that is significantly lower than that of myoglobin.[26] Taken with the oxygen-binding capabilities of **2**, we therefore conclude that the covalent linkages of heme C do not interfere with or limit the oxygen binding capabilities of *c*-type cytochromes, and that the lack of characterized, natural oxygen-binding *c*-type cytochromes is not due to an intrinsic issue with the biophysical properties of heme C *per se*. We also believe our heme C maquettes have the capacity for further incorporation of functional elements that may support the activation of bound molecular oxygen for catalytic hydroxylation. Indeed our cytochrome *c* maquette displays an oxyferrous autoxidation rate that falls well within the range of the oxygen binding and activating cytochromes P450,[29-31] thus potentially providing a robust scaffold for studying oxygen activation in the context of an artificial protein.

In engineering artificial redox proteins, there is a great advantage to covalently securing the tetrapyrrole redox cofactor. The covalent modification confers on the protein an effectively infinite affinity for heme. Heme C redox and biophysical properties can thus be modulated by selecting alternative ligation schemes such as 6-coordinate His/Met and 5-coordinate mono-His, facilitating site differentiation in a maquette over hundreds of millivolts scale and the formation of 5-coordinate catalytic heme sites for small molecule binding and subsequent redox catalysis. Such modulation of cofactor reduction potential is a key factor in the functional engineering of natural heme-containing oxidoreductases, largely dictating the enzymatic function of the cofactor,[2] or, when incorporated into a chain of redox cofactors, the driving force for and direction of electron transfer.[32]

Furthermore, the covalent incorporation of heme should enable dense packing of hemes within an artificial cytochrome. Natural *b*-type cytochromes rarely contain more than 2 hemes per polypeptide chain or discrete functional unit, while there are many examples of *c*-type cytochromes with between 4 and 34 hemes per polypeptide chain.[43] These multiheme *c*-type cytochromes often arrange their heme cofactors in close proximity to support very rapid electron transfer rates across long distances, functioning very much like a protein-insulated wire. The protein maquettes reported here might be redesigned for multi-electron reduction and oxidation of small molecules or expressible, self-assembling molecular wires for extracellular metal reduction.

Since maquettes have a demonstrated ability to accommodate a wide range of cofactors in a relatively malleable interior, further development of these artificial multi-cofactor oxidoreductases can exploit a rich combinatorial diversity of cofactor moderated functions within a single protein,[44] often with only minor modification of the protein framework, ranging from reversible oxygen binding (an obligate intermediate in cytochrome P450 and heme dioxygenase catalysis[3]) to light-activated electron transfer.

Materials and methods

Molecular Biology, Expression and purification

Synthetic genes were obtained from DNA2.0 (USA), and expressed from a modified variant of pMal-p4x (pSHT) and the proteins were purified using a combination of standard Nickel-affinity chromatography and high performance liquid chromatography (as described in the Supplementary information). The preparation of metal-free and Zn-substituted **2** is described in the Supplementary information.

Spectroscopy and biophysical characterization

Detailed procedures for maquette characterization by UV/visible, fluorescence, EPR, NMR and CD spectroscopies, stopped-flow spectrophotometry, AUC, MALDI-TOF mass spectrometry and spectroelectrochemistry are contained within the Materials and Methods section of the Supplementary information.

Acknowledgements

This work was supported at the University of Bristol by the BBSRC (grant no: BBI014063/1), the Royal Society, through a University Research Fellowship to JLRA. Research on the development of genes, molecular biology supplies for cloning and mutagenesis, protein and heme C expression and purification from *E. coli*, and the work on the biophysical and spectroscopic characterization of the heme C maquette as an oxygen transporter was supported at the University of Pennsylvania by the US National Institutes of Health, General Medical Institutes [RO1 GM 41048]. Research on converting the C-heme maquette into a light-active Zn-porphyrin maquette and its biophysical characterization was supported by the US Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering [DE-FG02-05ER46223]. The Authors wish to thank Prof. Dek Woolfson (Bristol) for kindly providing access to his equipment, Prof. Stuart Ferguson (Oxford) for his kind gift of the pEC864 vector and *E. coli* strain EC65, and Prof. Graeme Reid (Edinburgh) for his kind gift of the pEC86 vector.

Notes and references

[a] *School of Biochemistry, University of Bristol, University Walk, Bristol, BS8 1TD, UK. Fax: +44 117 331 2168; Tel: +44 117 331 2151; E-mail: ross.anderson@bristol.ac.uk (to whom correspondence should be addressed).*
[b] *The Johnson Research Foundation, Dept. of Biochemistry and Biophysics, University of Pennsylvania, PA19104-6059, USA.*
[c] *School of Chemistry, University of Bristol, Bristol, BS8 1TS, UK.*
[*] *These authors contributed equally to the work.*

† Electronic Supplementary Information (ESI) available: [Experimental methods, supplementary tables S1-3 and supplementary figures S1-11]. See DOI: 10.1039/b000000x/

1. K. Channon, E. H. Bromley, D. N. Woolfson, *Curr. Opin. Struct. Biol.* 2008, **18**, 491-8.
2. B. R. Lichtenstein, T. A. Farid, G. Kodali, L. A. Solomon, J. L. R. Anderson, M. M. Sheehan, N. M. Ennist, B. A. Fry, S. E. Chobot, C. Bialas, J. A. Mancini, C. T. Armstrong, Z. Zhao, T. V. Esipova, D. Snell, S. A. Vinogradov, B. M. Discher, C. C. Moser, P. L. Dutton, *Biochem. Soc. Trans.* 2012, **40**, 561-566.
3. C. T. Armstrong, D. W. Watkins, J. L. R. Anderson, *Dalton Trans.* 2013, **42**, 3136-3150.
4. J. M. Stevens, O. Daltrop, J. W. A. Allen, S. J. Ferguson, *Acc. Chem. Res.* 2004, **37**, 999-1007.
5. S. Kakar, F. G. Hoffman, J. F. Storz, M. Fabian, M. S. Hargrove, *Biophys. Chem.* 2010, **152**, 1-14.
6. R. L. Koder, J. L. R. Anderson, L. A. Solomon, K. S. Reddy, C. C. Moser, P. L. Dutton, *Nature* 2009, **458**, 305-309.
7. A. Chakrabartty, A. J. Doig, R. J. Baldwin, *Proc. Natl. Acad. Sci. USA* 1993, **90**, 11332-11336.
8. C. Negron, C. Fufezan, R. L. Koder, *Proteins* 2009, **74**, 400-416.
9. C. di Guan, P. Li, P. D. Riggs, H. Inouye, *Gene* 1988, **67**, 21-30.

10. M. Braun, L. Thony-Meyer, *Proc. Natl. Acad. Sci. USA* 2004, **101**, 12830-12835.
11. E. Arslan, H. Schulz, R. Zuffery, P. Kunzler, L. Thony-Meyer, *Biochem. Biophys. Res. Commun.* 1998, **251**, 744-747.
12. D. A. I. Mavridou, J. M. Stevens, L. Monkemeyer, O. Daltrop, K. di Gleria, B. M. Kessler, S. J. Ferguson, J. W. A. Allen, *J. Biol. Chem.* 2011, **287**, 2342-2352.
13. H. Schulz, L. Thony-Meyer, *J. Bacteriol.* 2000, **182**, 6831-6833.
14. O. Daltrop, J. W. Allen, A. C. Willis, S. J. Ferguson, *Proc. Natl. Acad. Sci. USA* 2002, **99**, 7872-7876.
15. O. Daltrop, S. J. Ferguson, *J. Biol. Chem.* 2004, **279**, 45347-45353.
16. F. W. Teale, *Biochim. Biophys. Acta* 1959, **35**, 543.
17. M. Ishida, N. Dohmae, Y. Shiro, T. Oku, T. Iizuka, Y. Isogai, *Biochemistry* 2004, **43**, 9823-9833.
18. P. D. Barker, E. P. Nerou, S. M. V. Freund, I. M. Fearnley, *Biochemistry* 1995, **34**, 15191-15203.
19. E. A. Berry, B. L. Trumpower, *Anal Biochem* 1987, **161**, 1-15.
20. T. Yamanaka, The Biochemistry of Bacterial Cytochromes. *Japan Scientific Societies Press and Springer-Verlag,* 1992.
21. O. Iakovleva, M. Reiner, H. Rau, W. Haehnel, F. Parak, *Phys. Chem. Chem. Phys.* 2002, **4**, 655-660.
22. Y. Lu, D. R. Casimiro, K. L. Bren, J. H. Richards, H. B. Gray, *Proc. Natl. Acad. Sci. USA* 1993, **90**, 11456-11459.
23. C. J. A. Wallace, I. Clark-Lewis, *J. Biol. Chem.* 1992, **267**, 3852-3861.
24. T. E. Meyer, M. A. Cusanovich, *Biochim. Biophys. Acta* 1985, **80**, 308-319.
25. S. Dewilde, L. Kiger, T. Burmester, T. Hankeln, V. Baudin-Creuza, T. Aerts, M. C. Marden, R. Caubergs, L. Moens, *J. Biol. Chem.* 2001, **276**, 38949-38955.
26. T. Shibata, D. Matsumoto, R. Nishimura, H. Tai, A. Matsuoka, S. Nagao, T. Matsuo, S. Hirota, K. Imai, S. Neya, A. Suzuki, Y. Yamamoto, *Inorg. Chem.* 2012, **51**, 11955-11960.
27. K. Bren, H. B. Gray, *J. Inorg. Biochem.* 1993, **51**, 111.
28. K. Klarskov, G. Van Driessche, K. Backers, C. Dumortier, T. E. Meyer, G. Tollin, M. A. Cusanovich, J. J. Van Beeumen, *Biochemistry* 1998, **37**, 5995-6002.
29. L. Eisenstein, P. Debey, D. Douzou, *Biochem. Biophys. Res. Commun.* 1977*,* **77**, 1377-1383.
30. I. G. Denisov, Y. V. Grinkova, M. A. McLean, S. G. Sligar, *J. Biol. Chem.* 2007, **282**, 26865-26873.
31. I. F. Sevrioukova, J. A. Peterson, *Arch. Biochem. Biophys.* 1995, **317**, 397-404.
32. C. C. Page, C. C. Moser, X. Chen, P. L. Dutton, *Nature* 1999, **402**, 47-52.
33. J. M. Vanderkooi, R. Landesberg, G. W. Hayden, C. S. Owen, *Eur. J. Biochem.* 1977, **81**, 339-347.
34. M. Sener, J. Strumpfer, J. Hsin, D. Chandler, S. Scheuring, C. N. Hunter, K. Schulten, *Chemphyschem.* 2011, **12**, 518-531.
35. G. A. Elove, A. K. Bhuyan, H. Roder, *Biochemistry* 1994, **33**, 6925-6935.
36. M. E., Dumont, A. F. Corin, G. A. Campbell, *Biochemistry* 1994, **33**, 7368-7378.
37. Y. Feng, S. G. Sligar, A. J. Wand, *Nat. Struct. Biol.* 1994, **1**, 30-5.
38. F. Arnesano, L. Banci, I. Bertini, S. Ciofi-Baffoni, T. de L. Woodyear, C. M. Johnson, P. D. Barker, *Biochemistry* 2000, **39**, 1499-1514.
39. J. Faraone-Menella, F. A. Tezcan, H. B. Gray, J. R. Winkler, *Biochemistry* 2006, **45**, 10504-10511.
40. B. San Francisco, E. C., Bretsynder, R. G. Kranz, *Proc. Natl. Acad. Sci. USA* 2012, **110**, E788-797.
41. W. B., Asher, K. L. Bren, *Chem. Commun.* 2012, **48**, 8344-8346.
42. P. O. Quintas, T. Catarino, S. Todorovic, D. L. Turner, *Biochemistry* 2011, **50**, 5624-5632.
43. C. G. Mowat, S. K. Chapman, *Dalton Trans.* 2005, **7**, 3381-3389.
44. T. A. Farid, G. Kodali, L. A. Solomon, B. R. Lichtenstein, M. M. Sheehan, B. A. Fry, C. Bialas, N. M. Ennist, J. A. Siedlecki, Z. Zhao, M. A. Stetz, K. G. Valentine, J. L. R. Anderson, A. J. Wand, B. M. Discher, C. C. Moser, P. L. Dutton, *Nat. Chem. Biol.*, 2013, DOI: 10.1038/NCHEMBIO.1362.

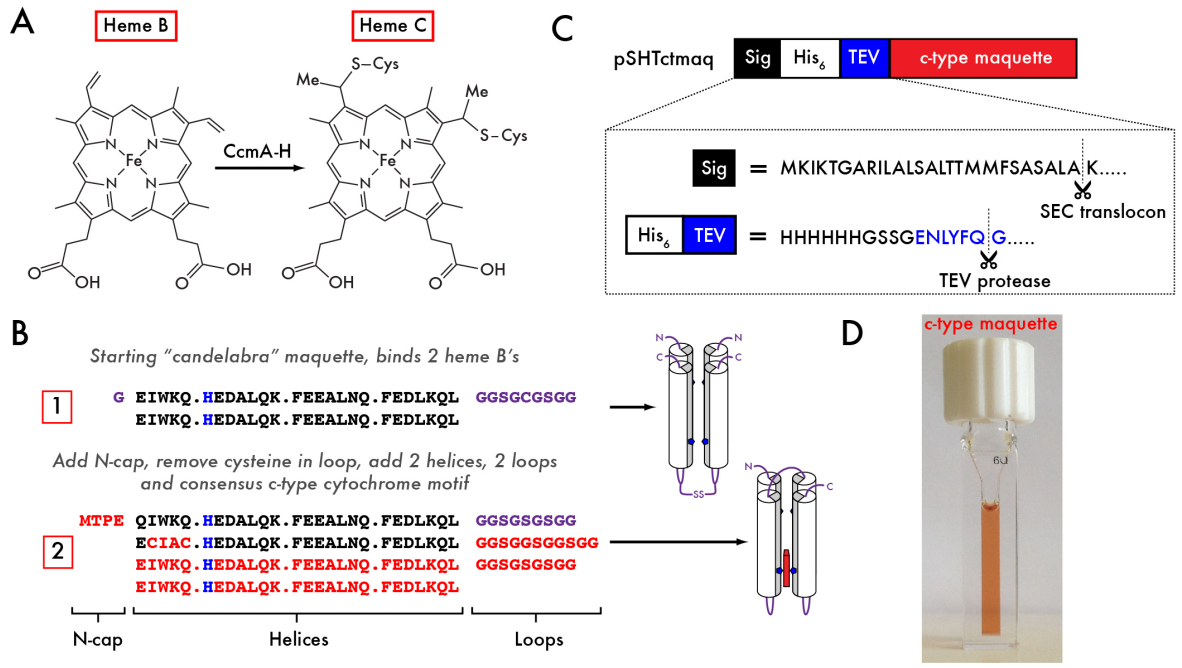

Fig. 1. Design and expression of a single-chain, artificial *c*-type cytochrome. A, Heme B is covalently attached to the substrate protein backbone *via* thioether linkages between the peripheral vinyl substituents on the porphyrin and the cysteine sidechains within the consensus CXXCH motif. This is achieved in *E. coli* with the cytochrome *c* maturation proteins (CcmA-H) located in the periplasmic space. **B**, Starting with the designed oxygen transport maquette, **1**, 2 loops (comprising glycine and serine) and a 4-residue N-cap (MTPE) are added to allow symmetry breaking across the 4-helix bundle and to add rigidity at the N-terminus. The CXXCH motif is then added into the sequence for covalent heme incorporation by the *E. coli* cytochrome *c* maturation apparatus (**2**). Changes to the amino acid sequence between iterations are highlighted in red, heme ligating histidine residues in blue and loops and N-termini in purple. Cartoon representations of the proteins are indicated to the right of the sequences: heme-ligating histidines are represented by blue pentagons, loops and N-/C-termini by purple lines, heme C positioning by the red rectangular box. **C,** The N-terminal signal sequence from the periplasmic *E. coli* maltose binding protein (MBP, black) is cleaved between A and K on transport across the inner membrane. A hexa-histidine tag (white) facilitates purification on Ni-NTA, and can be subsequently cleaved through the addition of a TEV protease cleavage site (blue - ENLYFQ-G). **D,** This expression system yields highly efficient heme incorporation into the expressed **2**.

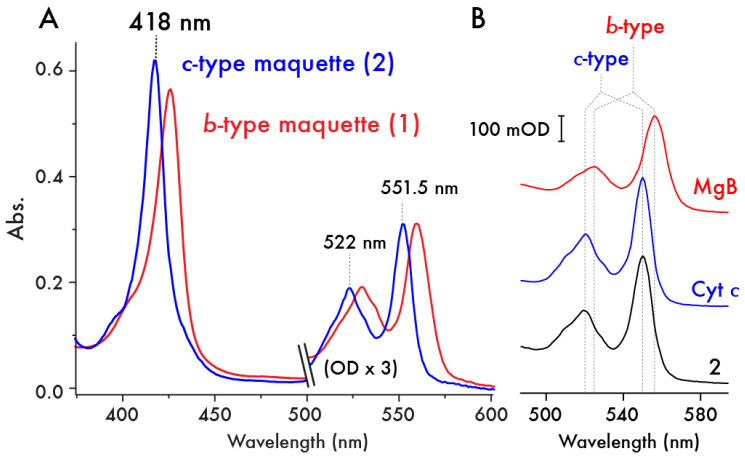

Fig. 2. Spectroscopic characterization of the *c*-type cytochrome maquette. A, UV/visible spectra of ferrous **1** with bound heme B (red) and purified **2** with covalently incorporated heme C (blue) reveals the expected blue-shift in the absorption spectrum of the *c*-type heme. **B**, UV/visible spectra of the pyridine hemochrome (α/β region) preparations of horse heart myoglobin (red, MgB), bovine cytochrome *c* (blue, Cyt c) and **2** (black). Stereotypical peak positions for *b*-type heme and *c*-type heme are indicated by dashed lines, highlighting the spectroscopic fingerprint for *c*-type heme in **2**.

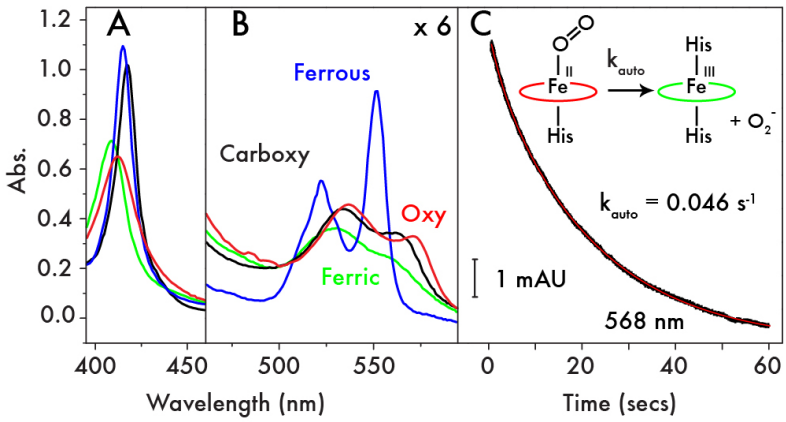

Fig. 3. Exogenous ligand binding by 2. UV/visible spectra in the **A**, Soret (γ), **B**, and Q-band (α/β) regions of ferrous (blue), ferric (green), oxyferrous (red) and carbonmonoxyferrous (black) **2**. **C**, Oxyferrous decay kinetics measured at 568 nm indicating the rate of autoxidation from oxyferrous to ferric **2** + superoxide (O_2^-). Raw data are displayed in black with a single exponential fit in red.

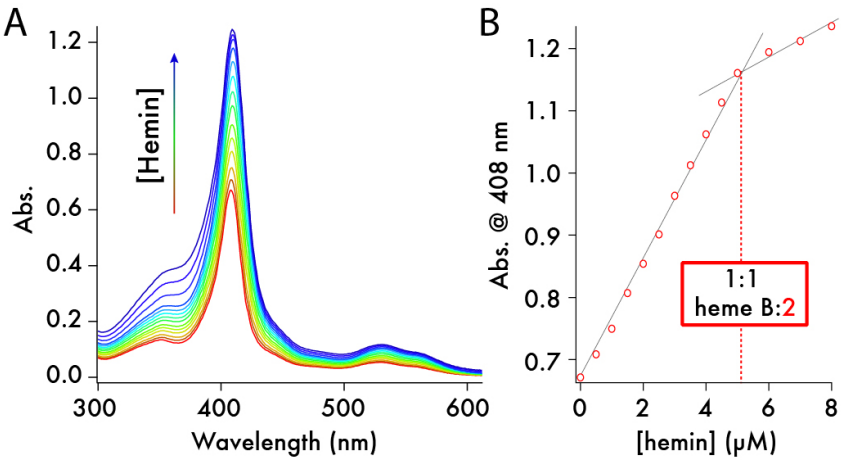

Fig. 4. 2 binds 1 equivalent of heme B per 4-helix bundle. A, UV/visible spectra from the titration of 0.5 μM hemin aliquots (0.5 mM hemin in DMSO) into 5 μM **2** (1 cm pathlength cuvette, 25 °C, 20 mM phosphate, 100 mM KCl, pH 7.5). **B**, heme binding isotherm indicating a stoichiometry of 1 heme B per molecule of **2** (dashed line).

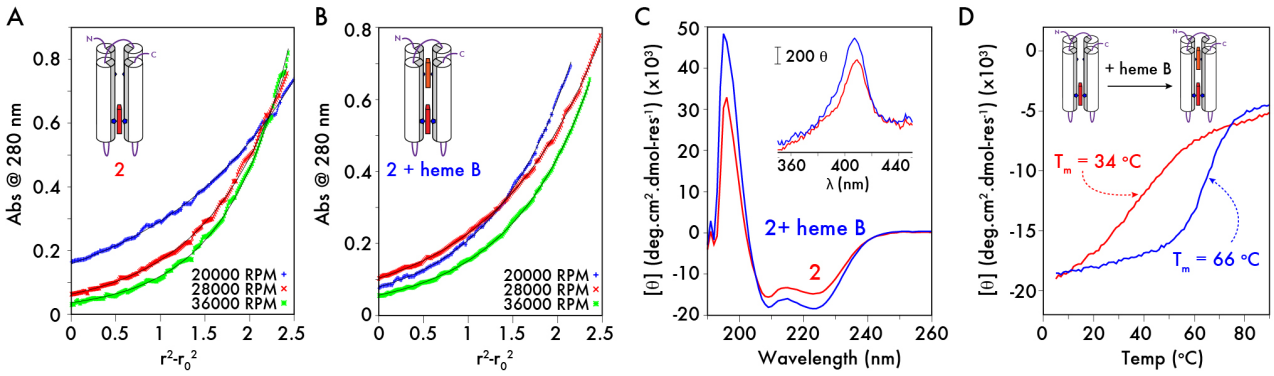

Fig. 5. Structural characterization of 2. A, B. Analytical ultracentrifugation of 15 μM **2** (**A**) and 15 μM **2** + heme B (**B**) in redox buffer (20 mM phosphate, 100 mM KCl, pH 7.5). Data were fitted to models of 1.16x the calculated MW for **2** and 1.05x the calculated MW for **2** + heme B. Residuals for the fits are plotted in SI Appendix, Fig. S6. **C**, Far-UV circular dichroism spectra of **2** (red) and **2** + heme B (blue) at 5 μM in redox buffer (20 mM phosphate, 100 mM KCl, pH 7.5). *Inset*, Near-UV/visible CD of **2** (red) and **2** + heme B (blue). (conditions as for far-UV CD measurements). **D**, Temperature dependence of the CD signal @ 222 nm for **2** (red) and **2** + heme B (blue), highlighting the increase in thermal stability (ΔT_m = +32 °C) on occupying the second heme binding site (conditions as for far-UV CD measurements).

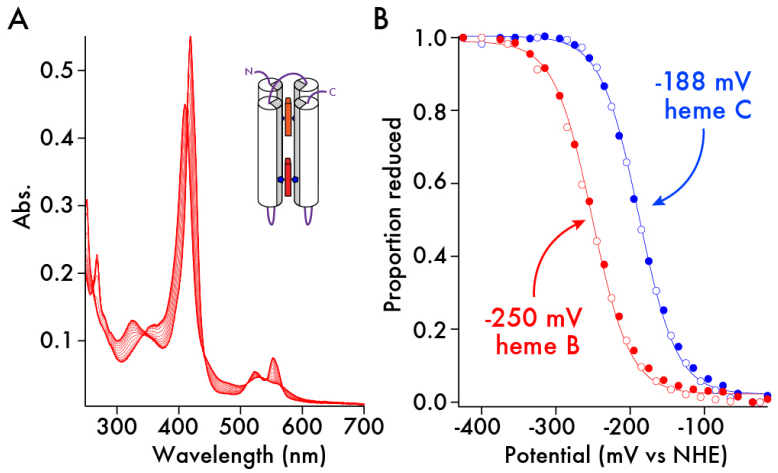

Fig. 6. Redox potentials of *c*-type and *b*-type heme in 2 + heme B. A. UV/visible spectra of the potentiometric titration of **2** + heme B (61 μM, 50 mM phosphate, 500 mM KCl, 10% glycerol, pH 7.5). **B.** Determination of redox potentials of c- and *b*-type hemes in **2** + heme B. Reductive (closed circles) and oxidative (open circles) data are fitted with 1-electron Nernst curves and quoted *vs* the Nernst hydrogen electrode (NHE). The proportion of reduced heme (*y-axis)* was determined for each sample at wavelengths where the contribution to the absorbance change was exclusive to the type of heme (*c*-type - 418 nm; *b*-type - 432 nm).

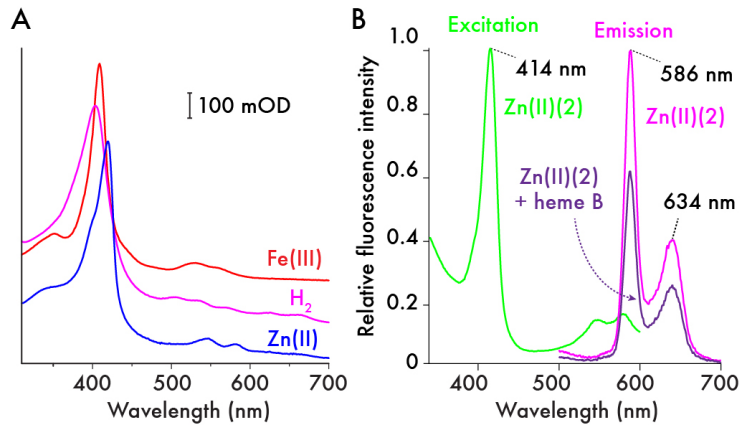

Fig. 7. Generation of a covalently-appended, light-sensitive cofactor in 2. A. UV/visible spectra of ferric (Fe(III)(**2**), red), demetallated (H_2(**2**), pink) and zinc (II) (Zn(II)(**2**), blue) **2** (5 μM protein, 20 mM phosphate, 100 mM KCl, pH 7.5). Iron is removed from the porphyrin with HF:pyridine solution, then after purification of demetallated **2** (H_2(**2**)), Zn^{2+} can be inserted to create a covalently bound, photosensitive cofactor. **B.** Fluorescence excitation (green) and emission (magenta) spectra of Zn(II)(**2**) (900 nM, 20 mM CHES, 150 mM KCl. pH 9.0). Emission spectra were recorded with an excitation wavelength of 420 nm in both heme-free and heme-bound Zn(II)(**2**) and normalized to the Zn(II)(**2**) maximum at 586 nm.

SUPPLEMENTARY INFORMATION

Supplementary Methods and Materials

Chemicals. Hemin was purchased from Frontier Scientific. All other chemicals were purchased from Sigma or Fisher Scientific.

Molecular biology. All cloning and mutagenesis was performed using the ligase-independent cloning/mutagenesis method of Tillett (1). The heme B-binding precursor of **2.0** (**1.5**) synthetic genes was obtained from DNA2.0 (USA). **1.5** was subcloned into pMal-p4x (NEB, USA) including the *E. coli* periplasmic maltose binding protein (MBP) as a fusion tag. **1.5** in pMal-p4x was mutated to insert the consensus CIACH motif required for *c*-type heme incorporation and the vector was altered to remove the MBP sequence, with the exception of the 26 amino acid periplasmic signal sequence. A cleavable N-terminal hexahistidine tag with a C-terminal tobacco etch virus N1a (TEV) protease-cleavable sequence were added after the periplasmic signal sequence to facilitate purification, producing the vector pSHT**2**.

Expression and purification. All proteins were expressed in BL21(DE3) cells. **2** was co-expressed with the *E. coli* cytochrome *c* maturation genes (*ccmA-H*) encoded on the pEC86 plasmid (2). All proteins were expressed and purified using the same protocol: transformed bacteria were grown in Luria-Bertani broth at 37 ^{o}C until the OD_{600} of the media reached 0.6—0.8. Protein expression was induced by the addition of IPTG (0.5 mM). We found that expressing **2** overnight (~15 hrs) before centrifugation yielded the best *c*-type heme incorporation. Cells were resuspended in 40 ml lysis buffer (50 mM phosphate, 100 mM KCl, 10 mM Imidazole, pH 8) per liter of culture, and sonicated on ice for 5 x 10 s pulses, with 10 s pauses. 1 ml of 100 mM PMSF per ml of lysis buffer was added before and after sonication to inhibit any protease activity in the lysate. Lysed cells were centrifuged at 20,000 g for 45 min and the resulting supernatant was loaded on to a Nickel-Sepharose His-Trap column (GE, USA) that had previously been equilibrated with lysis buffer, using an AKTA FPLC system (GE, USA). The column was washed with lysis buffer until the absorbance at 280 nm of the flow-through became stable. The bound protein was eluted using a linear gradient of imidazole, and the resulting fractions that contained protein of the correct molecular weight (as determined by gel electrophoresis) were pooled and dialyzed in to TEV cleavage buffer (50 mM Tris-HCl, 0.5 mM EDTA, 1 mM DTT, pH 8.0) and cleaved over night using TEV protease. Prior to HPLC purification, heme C is incorporated into 85 - 95 % of the *c*-type cytochrome maquette **2**. The final purification step was high performance liquid chromatography (HPLC) performed on a Jasco PU980 HPLC system in conjunction with a reverse-phase C18 column. Proteins were eluted using a linear gradient of acetonitrile (Fisher) in water and 0.1 % trifluoroacetic acid (Sigma Aldrich). The identity of the purified proteins were confirmed by matrix assisted laser desorption ionization time of flight (MALDI-TOF) mass spectrometry, performed on a 4700 TofTof mass spectrometer (Applied Biosystems).

Electronic spectroscopy and stopped flow spectrophotometry. UV/visible spectra were recorded on an Agilent Cary 60 spectrometer (Agilent, USA). Heme binding stoichiometries were determined by titrating solutions of hemin (0.1 - 2 mM in DMSO) into **2** at varying concentrations of protein (1 - 10 μM) and plotting ΔAbs vs [hemin] added. Pyridine hemochrome spectra of **2**, myoglobin and horse heart cytochrome *c* were recorded using the method of Berry and Trumpower (3). Extinction coefficients for **2** were calculated using the results of the pyridine hemochrome and were $\varepsilon_{280\ nm}$ =

25780 $M^{-1}cm^{-1}$ and $\varepsilon_{408\ nm}$ = 121400 $M^{-1}cm^{-1}$ for ferric **2** and $\varepsilon_{418\ nm}$ = 172700 $M^{-1}cm^{-1}$ for ferrous **2**. Stopped-flow UV/vis spectra of oxyferrous and carbonmonoxy **2** were collected either with an OLIS RSM 1000 spectrometer (OLIS, USA) or an Applied Photophysics SX20 spectrometer (Applied Photophysics, UK) housed in a Belle Technology anaerobic glove box (Belle Technology, UK). Solutions of ferrous **2** in redox buffer (20 mM phosphate, 100 mM KCl, pH 7.5) were prepared by carefully titrating sodium dithionite into degassed, anaerobic ferric **2** solutions and either transferred directly into the loading syringe of the anaerobically-housed SX20 spectrometer or *via* cannula into that of the RSM 1000 spectrometer. Full near-UV/visible spectra (300 - 800 nm) were recorded and the data analyzed using IgorPro (Wavemetrics, USA). Fluorescence excitation and emission spectra were recorded on a Fluorolog 3 fluorimeter (Horiba, Japan) with entrance and exit slits of 2 nm width. Emission spectra were recorded with an excitation wavelength of 420 nm in both heme-free and heme-bound Zn(II)(**2**). Electron paramagnetic resonance spectra were recorded using a Bruker Elexsys E500 spectrometer at X-band (9.4 GHz) using an Oxford Instrument ESR900 helium flow cryostat.

Heme extraction. The heme extraction test was performed according to the method of Teale (4). 2 ml of heme protein (50 μM) was added to ice-cold, acidified 2-butanone (50 μl concentrated HCl in 2 ml 2-butanone chilled in an ice/salt bath). The mixture was briefly shaken and then allowed to phase separate. Non-covalently bound heme B is extracted into the upper organic layer and covalently incorporated heme remains bound to the protein in the aqueous layer.

Circular dichroism. CD spectra were obtained using a JASCO J-810 spectropolarimeter fitted with a Peltier temperature controller (Jasco, UK). Thermal melts were performed with heating rates of 60 ^{o}C/hour with data points taken at 222 nm (bandwidth 3 nm, 16 s integration time) every 1 ^{o}C, and full spectra recorded every 20 ^{o}C (100 nm/min, bandwidth 1 nm).

Analytical ultracentrifugation. Analytical ultracentrifugation of **2** was performed using a Beckman-Optima XL-I analytical ultracentrifuge fitted with an An-60 Ti rotor. Peptide solutions (15 μM) were prepared in a buffer solution comprising 20 mM potassium phosphate and 100 mM potassium chloride at pH 7.5. Sedimentation equilibrium data were fitted simultaneously with Ultrascan (5) using a single ideal species model; protein partial specific volumes (0.7192 and 0.7186 without and with the *b*-type heme bound, respectively) were estimated using Sednterp (http://www.jphilo.mailway.com/), taking the partial specific volume of heme to be 0.881 ccm/g (6). The calculated molecular weight of **2** was 18300 Da (1.16x monomer mass), and that of **2** + heme B was 17210 Da (1.05x monomer weight).

Nuclear magnetic resonance. Solutions of 0.5 mM **2** and **2** + heme B were prepared in redox buffer containing 10% v/v D_2O (Sigma Aldrich, UK). 1D spectra were acquired using a Shigemi tube (Sigma Aldrich, UK) on a Varian INOVA 600 (University of Bristol) using double-pulsed field gradient spin-echo excitation for water suppression. Spectra were processed using NMRPipe (7) and visualized using the CCPNMR software suite (8).

Redox potentiometry. Anaerobic potentiometric titrations were performed using a modified quartz EPR OTTLE cell—equipped with platinum working and counter electrodes and a Ag/AgCl reference electrode (BASi, USA)—in concert with a Biologic SP-150 potentiostat (Bio-logic, France) and Cary 60 UV/visible spectrometer (Agilent, USA). Titrations were performed in OTTLE buffer (50 mM phosphate, 500 mM KCl, 10 % glycerol, pH 7.5) using a protein concentration of between 50 - 150 μM, and included the following redox mediators at 20 μM concentration: phenazine, anthroquinone-2-sulfonate, 2-hydroxy-1,4-naphthoquinone, benzyl viologen. Heme reduction potentials were determined by fitting the data to the Nernst equation (**Equation 1**) using IgorPro software

(Wavemetrics, USA). Reported reduction potentials are quoted vs the Nernst hydrogen electrode (NHE).

Eqn. 1. $E_{applied} = E_m + RT/nF.In([Ox]/[Red])$

Preparation of metal-free, zinc and zinc/heme B maquette 2. 3 mg of **2** was thoroughly lyophilized to ensure no water was present in the sample. 1.5 ml of 70% hydrogen fluoride in pyridine was added and the solution stirred for 10 minutes in the dark. The HF was evaporated under an argon stream and the metal-free **2** (H_2(**2**)) was dissolved in 2 ml of 50 mM ammonium acetate, pH 2.0. Residual pyridine was removed using a PD-10 desalting column (GE, USA), and 5 equivalents of zinc acetate were added to H_2(**2**) while the pH was maintained at pH 2.0. Zinc incorporation was monitored by UV/visible spectroscopy. After incorporation was complete, the protein was lyophilized and redissolved in 20 mM CHES, 150 mM KCl, pH 9.0 for characterization and heme B incorporation. Titrations were performed on 1 ml of 900 nM Zn(II)(**2**) calculated using the extinction coefficient of 234,000 $M^{-1}cm^{-1}$ at 420 nm.

References

1. Chiu, J., March, P.E., Lee, R., Tillett, D. (2004) Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Res* 32(21):e174.
2. Arslan, E., Schulz, H., Zuffery, R., Kunzler, P., Thony-Meyer, L. (1998) Overproduction of the *Bradyrhizobium japonicum c*-type cytochrome subunits of the *cbb*$_3$ oxidase in *Escherichia coli*. *Biochem Biophys Res Commun* 251(3):744-747.
3. Berry, E.A., Trumpower, B.L. (1987) Simultaneous determination of hemes a, b and c from pyridine hemochrome spectra. *Anal Biochem* 161(1):1-15.
4. Teale, F.W. (1959) Cleavage of the haem-protein link by acid methylethylketone. *Biochim Biophys Acta* 35:543.
5. Demeler, B. (2005) UltraScan—a comprehensive data analysis software package for analytical ultracentrifugation experiments. in *Modern Analytical Ultracentrifugation: Techniques And Methods* (eds. Scott, D.J. & Harding, S.E.) 210–229 (Royal Society of Chemistry).
6. Demoll, E., Cox, D.J., Daniel, E., Riggs, A.F. (2007) Apparent specific volume of human hemoglobin: effect of ligand state and contribution of heme. *Anal Biochem* 363(2):196-203.
7. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277-293.
8. Franken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., Laue, E.D. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* 59(4):687-696.

Supplementary Tables

3M4CA	2YYWA	3OA8A	2B7SA	1SP3A	1FGJA	2J7AA	3M15A
1S05A	2YYXA	2C1DA	1M64A	1SP3A	1FGJA	2J7AC	1M1QA
1CGOA	2J8WA	2C1DA	2E84A	1SP3A	1QO8A	1KSUA	3BNJA
1CGNA	2ZZS1	1CZJA	2E84A	1GU2A	1Q9IA	1KSSA	1W7OA
2RDZA	1GYOA	1EYSC	2E84A	2XLWA	2DVHA	2BQ4A	2VR0A
2OT4A	2XM4A	1EYSC	2E84A	2XLEA	1GQAA	2BQ4A	1VRNC
1J0PA	1OFWA	3PMQA	2E84A	2XL8A	2CCYA	1FT5A	1MDVA
1J0OA	1OFWA	1AQEA	2E84A	2XL6A	1MQVA	1FT5A	1H32A
1H29A	1WADA	3L1MA	1P2HA	2XLMA	1A7VA	1LJ1A	1H32A
1H29A	2WJNC	2CTHA	1P2EA	1DUWA	1DT1A	2CDVA	3H34A
1H29A	2WJNC	2CY3A	2BC5A	1DUWA	1Y0PA	3HNKA	
1H29A	1BBHA	3CYRA	2YXCA	3FOOA	1Z1NX	3HNJA	
1H29A	1CPQA	1D4DA	3H4NA	3FO3A	1Z1NX	1JRXA	
1H29A	1CPRA	1D4CA	2A3MA	1QDBA	1Z1NX	1JRYA	
1H21A	1I77A	3QW0A	2CVCA	2EWKA	1Z1NX	1JRZA	
2P0BA	3ML1B	1WR5A	2CVCA	2EWIA	1Z1NX	19HCA	
2JBLC	1E39A	1C52A	2CVCA	2EWUA	1Z1NX	19HCA	
2JBLC	3DE8A	3C62A	2CVCA	2Z47A	1Z1NX	1FS7A	
3IQ5A	3CAOA	3C63A	2CVCA	3NMIA	3RKHA	1FS9A	
2RF7A	1OAHA	2B7RA	2CVCA	1FGJA	1UP9A	3SCEA	

Table S1. PDB codes of *c*-type cytochromes with heme C binding sites located on a helix. Multiple entries for single PDB ID's indicate multiple helix-bound *c*-type heme sites are located within a single protein.

	Oxyferrous l_{max} (nm)			
c-type cytochromes	***g***	***b***	***a***	***Ref***
2	**410**	**535**	**571**	**This work**
Iso-1-cytochrome *c* M80A	408	537	570	30
Semi-synthetic horse heart cytochrome *c* M80A	n/a	535	571	31
SHP (*Sphaeroides* heme protein)	408	538	570	32

Table S2. UV/visible absorbance data for *c*-type cytochrome oxyferrous species, comparing the maquette, **2,** with methionine-to-alanine mutants of cytochrome *c* and wild type SHP (*Sphaeroides* heme protein).

Heme protein	k_{auto} (h^{-1})	$t_{1/2}$ (s)	*Temp* (°C)	*Ligation*	*Heme*	*Ref*
2	**165**	**15**	**15**	**His/His***	***c*-type**	**This work**
1.5	240	10	15	His/His*	*b*-type	This work
1	266	9.4	15	His/His*	*b*-type	This work
Murine neuroglobin	19	130	37	His/His*	*b*-type	33
Human neuroglobin	5.4	460	37	His/His*	*b*-type	33
Sperm whale myoglobin	0.083	3×10^4	35	His	*b*-type	34
P450cam - camphor	51	50	25	Cys	*b*-type	37
P450 3A4 + testosterone	1300	1.9	4	Cys	*b*-type	38
P450 BM3 + arachidonic acid	792	3.15	20	Cys	*b*-type	39
Iso-1 cytochrome *c* M80A	0.01	2×10^5	22	His	*c*-type	35
SHP	14	180	25	His	*c*-type	36

Table S3. Comparison of oxyferrous state autoxidation rates (k_{auto}) and half lives ($t_{1/2}$) of the maquette, **2**, with natural oxygen binding proteins. Though less stable to autoxidation with respect to the globins, **2** exhibits autoxidation kinetics in the range of the oxygen binding and activating cytochromes P450. Ligation refers to the iron-coordinating ligands in the ferrous heme state. * - The neuroglobins and oxygen binding maquettes undergo rapid ligand exchange at the distal histidine site, and the heme ligation can be considered as an equilibrium between 6-coordinate bis-histidine and 5-coordinate mono-histidine states.

Supplementary Figures

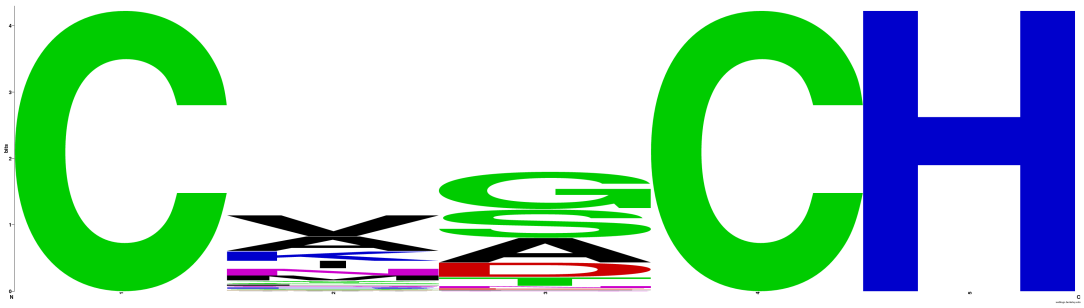

Fig. S1. Weblogo of 150 helical *c*-type heme CX_1X_2CH motifs extracted from the Protein Data Bank (PDB, http://www.rcsb.org/pdb/home/home.do). PDB files of natural cytochromes were downloaded from the non-redundant structural database (from the PISCES server) where the following conditions were satisfied: any CXXCH motif found on a helix of a non-redundant structure containing HEM or HEC. PDB files were hand-curated, the *c*-type binding motif recorded then visualized using WebLogo. A full list of downloaded PDB files can be found in **Table S1**.

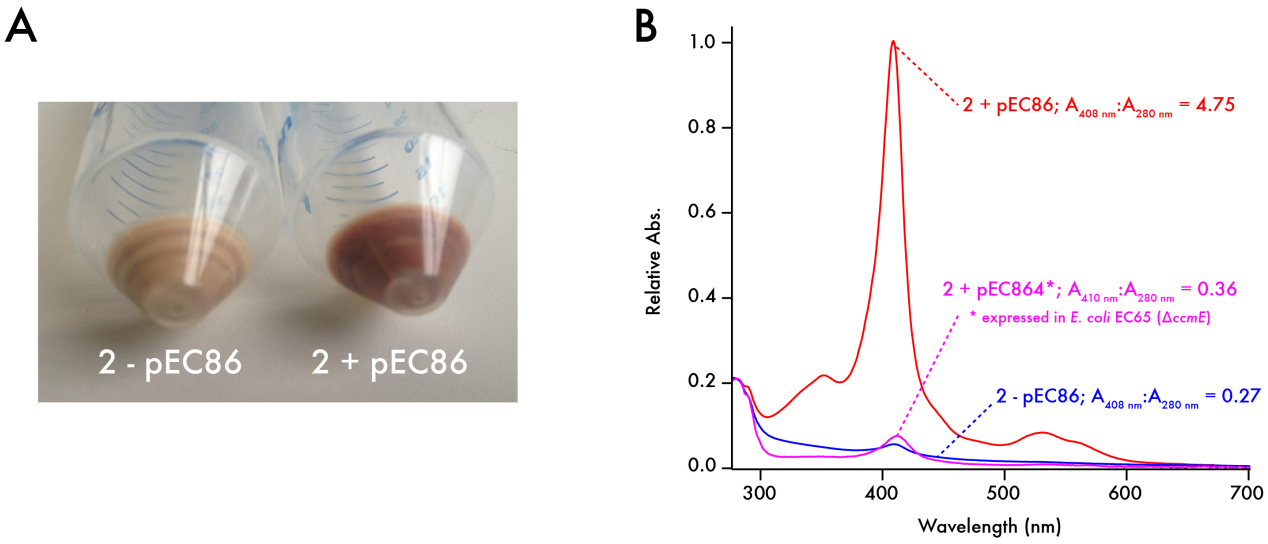

Fig. S2. Covalent heme incorporation is dependent on the presence of the heme lyase CcmE. **A,** The harvested pellet of *E. coli* expressing **2** in the absence (left) and presence (right) of the *ccmA-H* containing pEC86 vector indicates a significant increase in heme concentration *in vivo* when the vector is present. **B,** UV/visible spectroscopic comparison of Ni-sepharose purified **2** expressed in the absence (blue) and presence (red) of the pEC86 vector harboring the fully functional *c*-type cytochrome maturation apparatus. When co-expressed in *E. coli* strain EC65 (Δ*ccmE*) with pEC864 – a variant of pEC86 with catalytically inactive CcmE (H130A CcmE) – no *c*-type heme is observed and the protein instead is purified with *b*-type heme bound (magenta). All traces are normalized to the protein absorbance at 280 nm, and ratios of the absorbances at 408:280 nm (for *c*-type heme) or 410:280 nm (for *b*-type heme) to compare levels of heme incorporation.

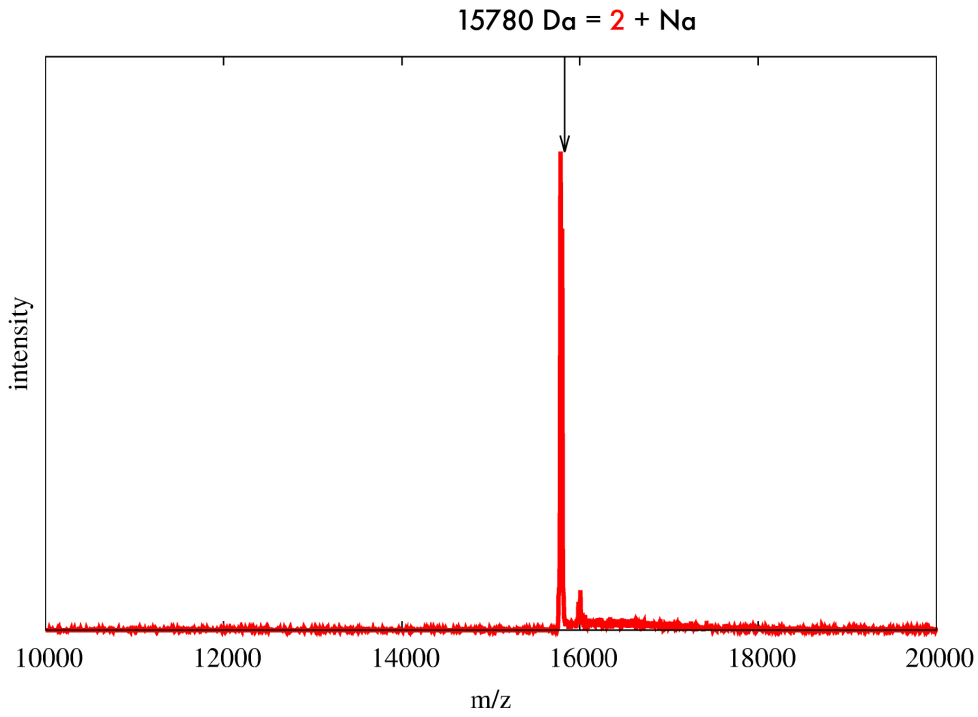

Fig. S3. MALDI-TOF mass spectrum of **2** (MW = 15758 Da) collected on a 4700 TofTof mass spectrometer (Applied Biosystems).

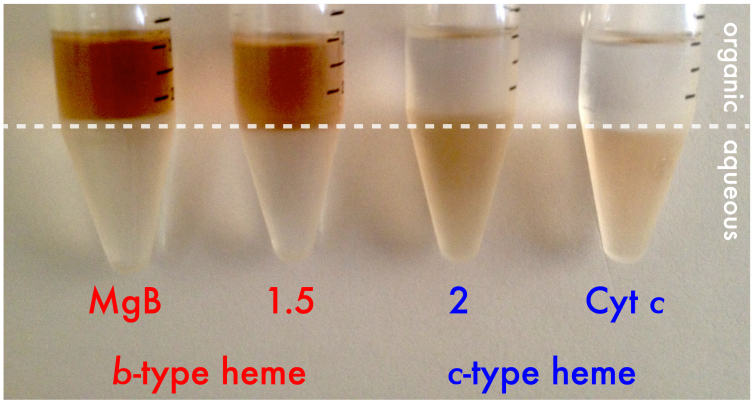

Fig. S4. The *c*-type heme-containing cytochrome *c* (Cyt *c*) and **2** retain heme and remain in the aqueous buffer-containing layer following treatment with acidified 2-butanone (2 ml of protein at 50 µM, 20 mM phosphate, 100 mM KCl, pH 7.5 added to 2 ml acidified 2-butanone at -10 ^{o}C, stirred for 5 minutes then allowed to partition) indicating covalent attachment of the heme to the protein backbone, while the *b*-type heme-containing myoglobin (MgB) and **1.5** experience heme loss to the organic layer due to cleavage of the histidine-heme linkages.

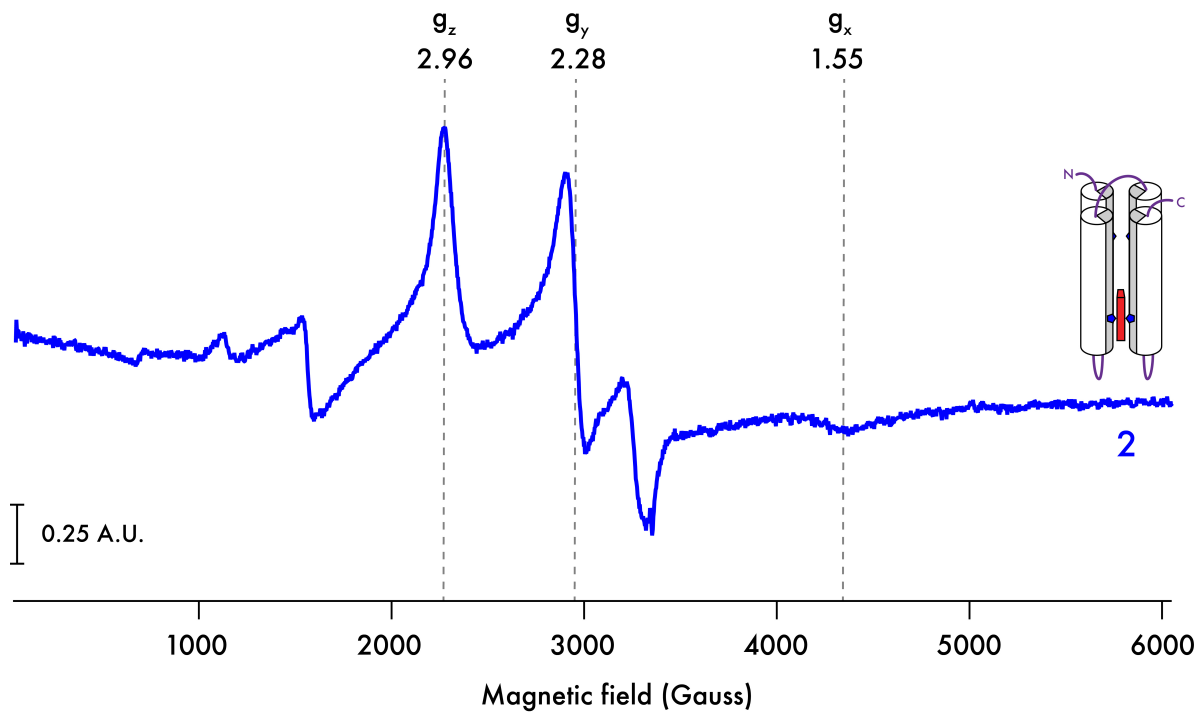

Fig. S5. EPR spectrum of ferric **2** (50 μM protein in 20 mM phosphate, 100 mM KCl), recorded at 10 K, 2 mW power.

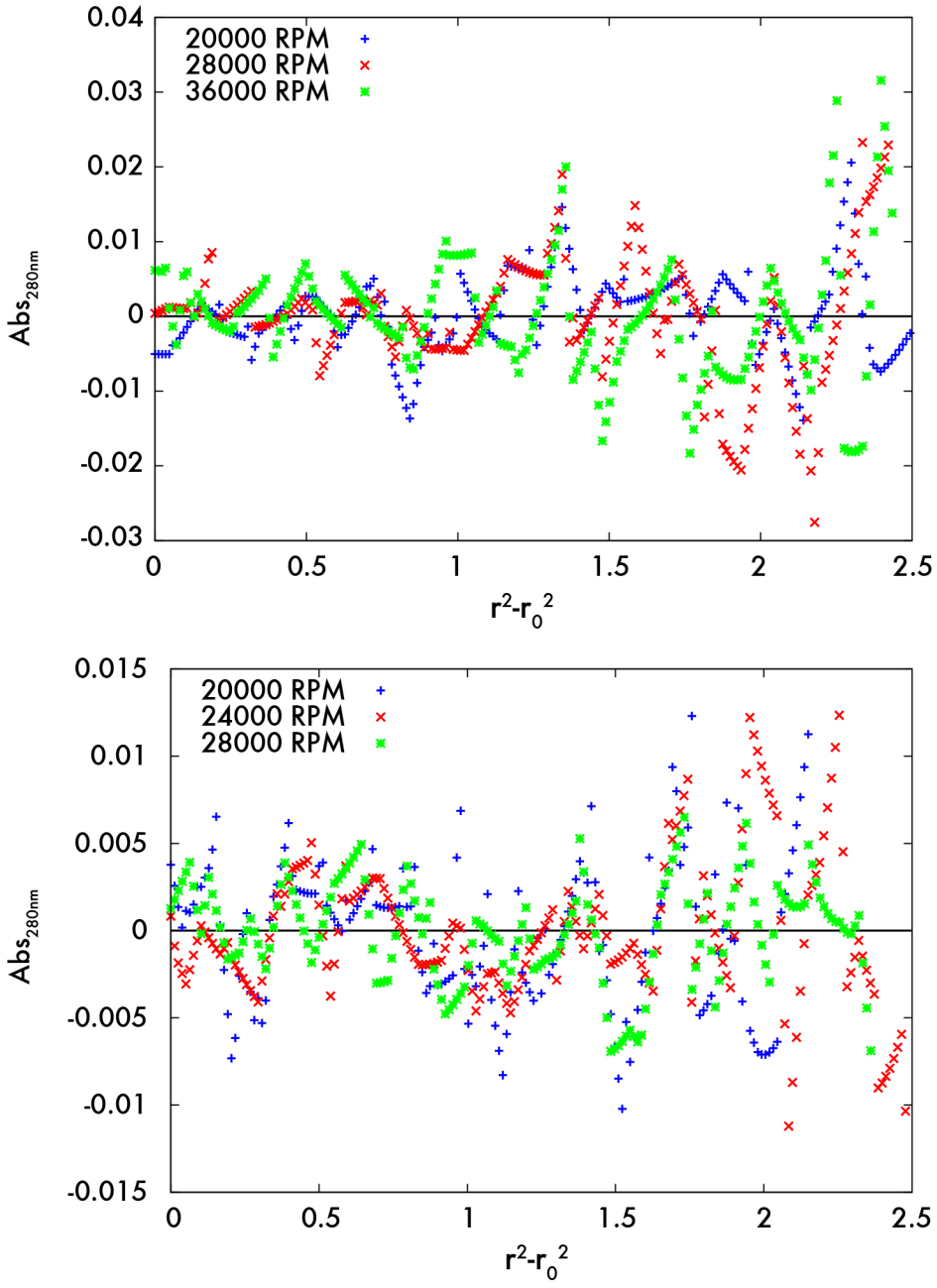

Fig. S6. Residuals calculated from the fitted Analytical Ultracentrifugation data. **Upper panel - 2**; **Lower panel - 2** + Heme B.

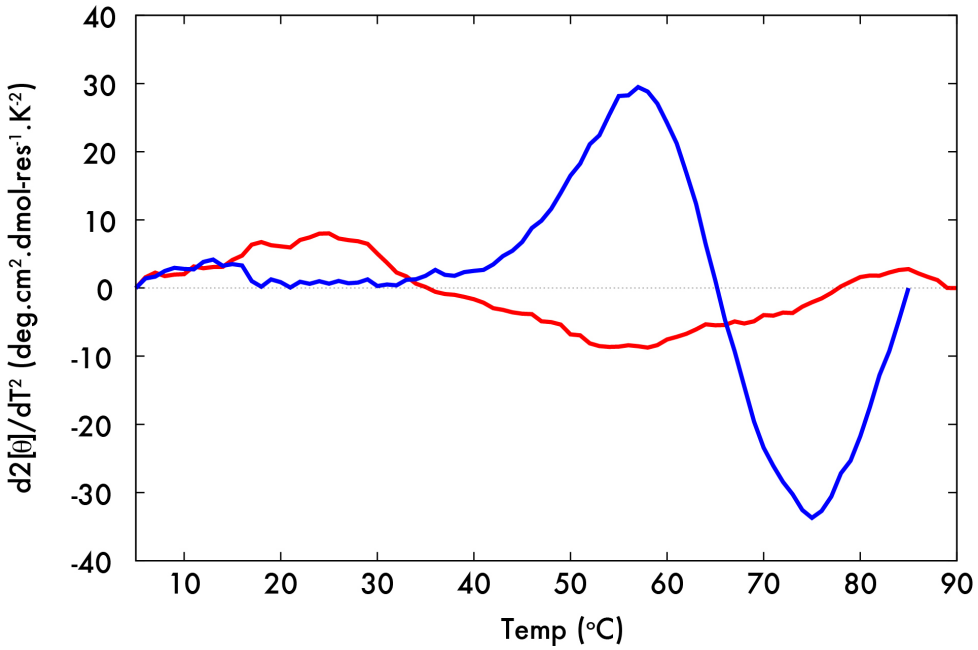

Fig. S7. Second derivatives of the CD melt data for **2** (red) and **2** + heme B (blue) indicating melting temperatures of 34 ^{o}C and 66 ^{o}C respectively. Experimental conditions as described in **Fig. 6**.

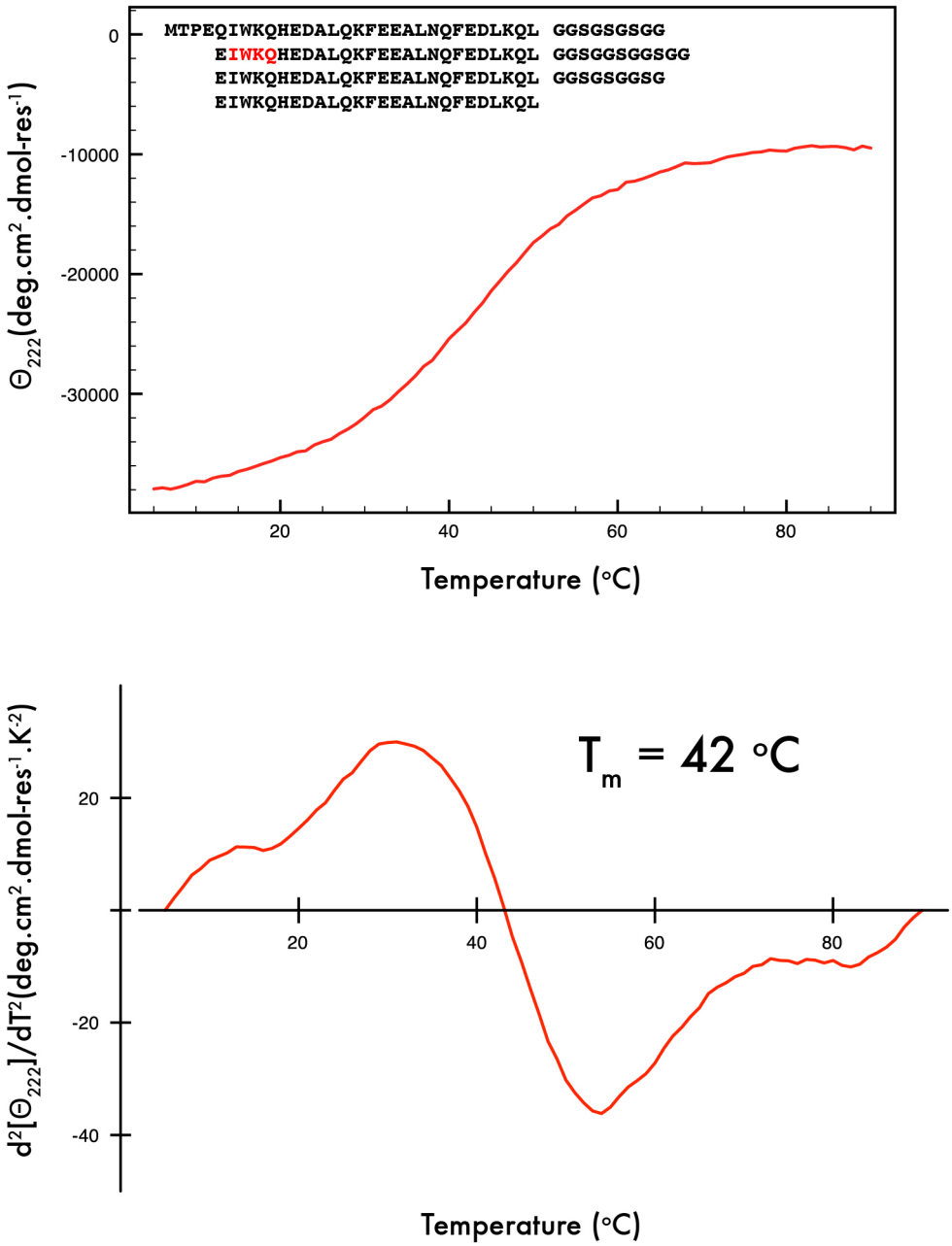

Fig. S8. Temperature dependence of the CD signal @ 222 nm for apo-**1.5** (upper panel) – the single chain maquette lacking the CXXCH motif - and the second derivative of the temperature dependence data (lower panel, Apo-**1.5**, T_m = 42 ^oC). The amino acid sequence of 1.5 is displayed in the upper panel. Differences with respect to **2** are highlighted in red.

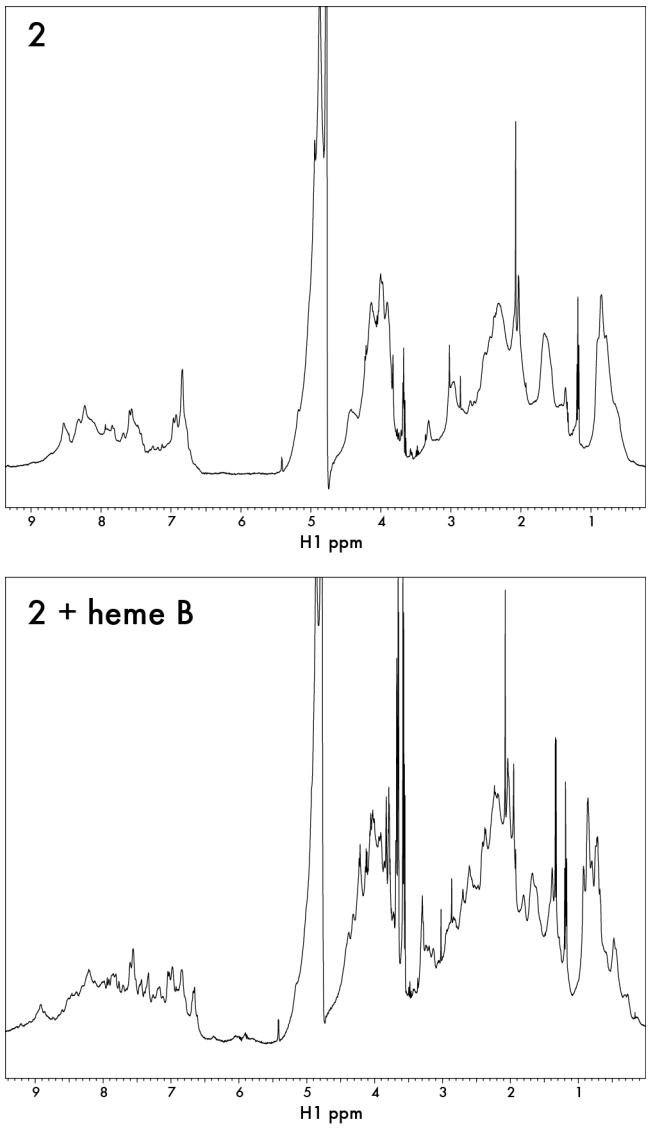

Fig. S9. Heme B-mediated structuring of **2**. ^{1}H NMR spectra of **2** (upper panel) and **2** + heme B (lower panel). Spectra recorded on a Varian INOVA 600 MHz spectrometer at 500 μM **2** in redox buffer containing 10 % v/v D_2O.

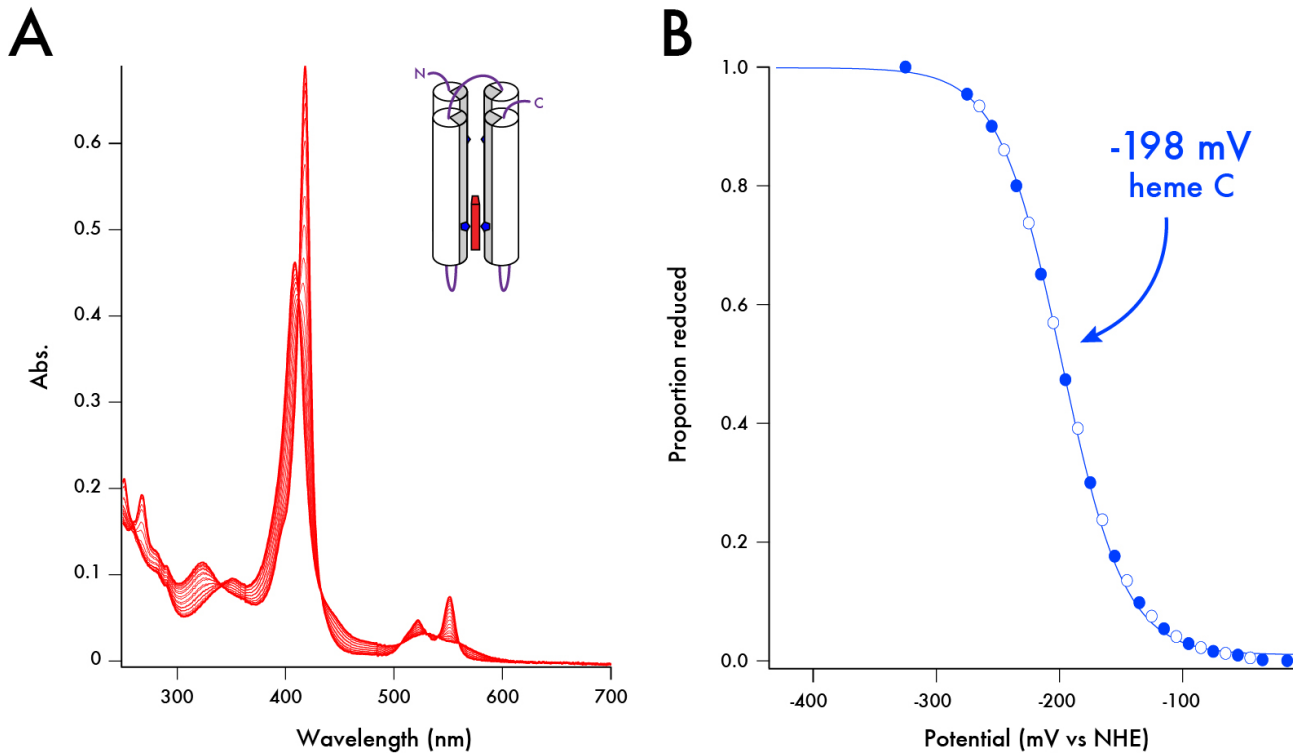

Fig. S10. Redox potential of heme C in **2. A.** UV/visible spectra of potentiometric titrations of **2** (130 µM, 50 mM phosphate, 500 mM KCl, 10% glycerol, pH 7.5) **B.** Determination of redox potential of *c*-type heme in **2**. Reductive (closed circles) and oxidative (open circles) data are fitted with 1-electron Nernst curves and quoted *vs* the Nernst hydrogen electrode (NHE). The proportion of reduced heme (*y-axis)* was determined at 418 nm.

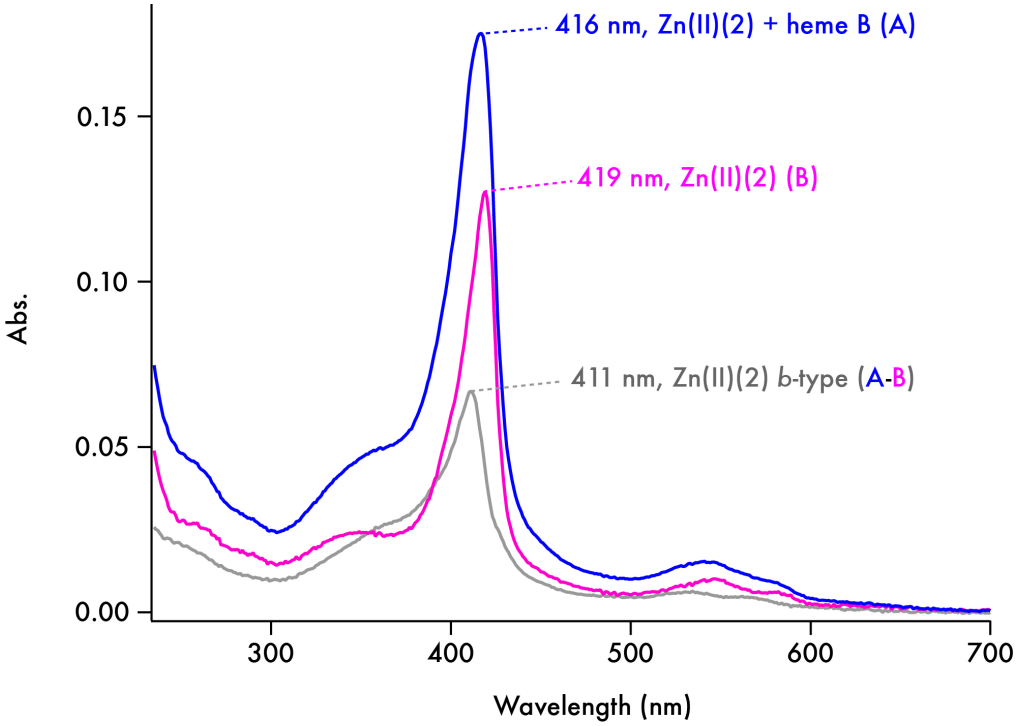

Fig. S11. Formation of a light sensitive porphyrin dyad in **2**. UV/visible spectra of Zn(II)(**2**) and Zn(II)(**2**) + 1 equivalent of heme B (900 nM **2**, 20 mM CHES, 150 mM KCl, pH 9.0). A difference spectrum is displayed, indicating the successful incorporation of a bis-histidine ligated heme B in the available tetrapyrrole binding site.