Published in final edited form as:

J Mol Biol. 2011 October 14; 413(1): 177–194. doi:10.1016/j.jmb.2011.08.019.

Symerythrin structures at atomic resolution and the origins of rubrerythrins and the ferritin-like superfamily

Richard B. Cooley^a, Daniel J. Arp^{b,*}, and P. Andrew Karplus^{a,*}

^aDepartment of Biochemistry and Biophysics, 2011 Ag & Life Sciences Bldg, Oregon State University, Corvallis, OR 97331

^bDepartment of Botany and Plant Pathology, 2082 Cordley Hall, Oregon State University, Corvallis OR 97331

Abstract

Rubrerythrins are diiron-containing peroxidases belonging to the ferritin-like superfamily (FLSF). Here, we describe structures of symerythrin, a novel rubrerythrin variant from the oxygenic phototroph *Cyanophora paradoxa*, at 1.20 to 1.40 Å resolution in three different states: diferric, azide-bound diferric and chemically reduced. The symerythrin metallocenter has a unique eighth ligating residue compared to rubrerythrin: an additional glutamate inserted into helix A of the four-helix bundle that resides on a π -helical segment. Otherwise, the diferric metallocenter structure is highly similar to that of characterized rubrerythrins. Azide binds the diferric center in a μ -1,1 orientation similar to how peroxide binds to diferric rubrerythrin. The structure of the diferrous metallocenter shows heterogeneity that we ascribe to the acidic pH of the crystals. In what we consider the neutral pH conformation, reduction causes a 2.0 Å shift in Fe1 and the toggling of a Glu to a His ligand as seen with rubrerythrins. The function of symerythrin remains unknown, but preliminary tests showing oxidase and peroxidase activity and the similarities of its metallocenter with other rubrerythrins suggest similar functionality between the two despite the additional ligating glutamate in symerythrin. Of particular interest is the high internal symmetry of symerythrin, which supports the notion that its core four-helix bundle was formed by the gene duplication and fusion of a two-helix peptide. Sequence comparisons with another FLSF family also having notable internal symmetry provide compelling evidence that, contrary to previous assumptions, there have been multiple gene-fusion events that have generated the single-chain FLSF fold.

Introduction

The ferritin-like superfamily (FLSF) is a diverse group of metalloproteins named for the ferritins that are responsible for iron storage in prokaryotes and eukaryotes.¹ Additional well-studied proteins within the FLSF include the bacterioferritins, rubrerythrins, Mn-

© 2011 Elsevier Ltd. All rights reserved.

*corresponding authors: P. Andrew Karplus, Department of Biochemistry and Biophysics, 2011 Ag & Life Sciences Bldg, Oregon State University, Corvallis, OR 97331, Phone: (541) 737-3200; Fax (541) 737-0481, karplusp@science.oregonstate.edu. Daniel J. Arp, Department of Botany and Plant Pathology, 2082 Cordley Hall, Oregon State University, Corvallis OR 97331, Phone: (541) 737-1294, Fax: (541) 737-5310, Dan.J.Arp@oregonstate.edu.

Accession Numbers. Coordinates and structure factors dithionite-reduced and azide-bound diferric models have been deposited in the Protein Data Bank with accession numbers 3qhc and 3sid, respectively.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

catalases, DNA-binding proteins from starved cells (DPS), Δ^9 -desaturases, the R2 subunit of class I ribonucleotide reductases, and bacterial multicomponent monooxygenases.¹ Each of these proteins has a conserved core four-helix bundle (with helices referred to as A, B, C and D) that binds a carboxylate-bridged dinuclear metallocenter through six positionally conserved ligating residues. The canonical FLSF metallocenter ligands are four carboxylate and two histidine residues (i.e. a six-ligand motif) symmetrically arranged such that helix pairs A–B and C–D of the four-helix bundle each contribute two carboxylates and one histidine.¹ This symmetry of the ligands has led to the proposal that the core four-helix bundle of the FLSF originated via the duplication and fusion of a gene encoding a simple two-helix protein that formed a symmetric diiron-binding dimer.^{1; 2}

The FLSF dinuclear centers are commonly diiron centers (with sites referred to as Fe1 and Fe2) which facilitate oxygen mediated redox reactions, although some FLSF enzymes have a dimanganese³ or a heteronuclear Mn/Fe cofactor.⁴ In all cases, the reduced form of the dinuclear center is needed to react with substrate, and the redox changes are commonly associated with so-called ‘carboxylate shifts’ that alter the coordination geometry of the metallocenter. This conformational plasticity is a hallmark feature of FLSF metallocenters and has been the subject of numerous structure-function analyses (see review by Sazinsky and Lippard⁵).

The exception to the six-ligand motif pattern is the rubrerythrin family (including rubrerythrins, nigerythrins, and sulrerythrins). These proteins are unusual among FLSF proteins in that they contain a seventh metal ligating residue. This additional ligand is a glutamate which resides within a π -helical bulge of core helix C and reflects an insertion compared to other FLSF proteins⁶; it gives rubrerythrins the unusual ability among FLSF proteins to preferentially react with hydrogen peroxide rather than dioxygen.⁷ As a result, rubrerythrins are believed to act as peroxidases *in vivo* in order to mitigate toxicity from reactive oxygen species,⁸ particularly in anaerobes where they are widely distributed.⁹ Interestingly, rubrerythrins have recognizable internal sequence similarity that goes beyond just the metallocenter ligands and this is thought to reflect the origins of the fold from a gene duplication and fusion event.^{2; 10}

Rubrerythrins have the most plastic metallocenter among FLSF proteins (in terms of their metal-ligating residues and iron atoms) in that although seven residues are involved in metal ligation, only six coordinate the metallocenter at any one time. In the diferric state the His from helix B is not a ligand, but on reduction to the diferrous state, Fe1 shifts ~2.0 Å so that it breaks its bond with the additional glutamate from helix C and makes a new ligating bond with the His from helix B.^{11; 12} This redox-dependent toggling of Fe1 positions is thought to be part of rubrerythrin’s catalytic cycle.¹³

During a recent analysis of π -helices in the FLSF,⁶ we noted that genes found only in the oxygenic phototrophs *Gloeobacter violaceus* and *Cyanophora paradoxa* (NCBI gene IDs 801647 and 2602414, respectively) encode a distinct variant of the rubrerythrin family having even higher internal sequence similarity than the characterized rubrerythrins. The encoded protein sequences include an inserted glutamate in helix A that, according to the internal symmetry of the FLSF four-helix bundle, is placed symmetrically to the ‘additional’ glutamate (inserted in helix C) of the rubrerythrins.⁶ Thus, we proposed this glutamate in helix A reflects an ancestral feature of the rubrerythrin family and serves as an unprecedented eighth metal-ligating residue residing in a π -helical segment of helix A symmetric with the π -helical segment of helix C of rubrerythrins. Given the possibility of discovering a novel rubrerythrin-variant and the implications it could have in uncovering new insights into the origins of an entire protein fold, we set out to test the validity of this hypothesis.

As a first step to characterize these novel rubrerythrin-like proteins, we have cloned and expressed the *C. paradoxa* protein and solved its crystal structure at 1.20 Å resolution with the metallocenter center in the diferric (oxidized) state. We named this protein symerythrin[†] to reflect its high level of internal symmetry and have briefly reported on an unexpected finding in the structure of this protein: an autogenerated carbon-carbon crosslink between the unfunctionalized side chains of valine and phenylalanine.¹⁴ Here, we provide a detailed description of the overall structure of symerythrin and its novel eight-residue diiron metallocenter in oxidized, diferric azide-bound and chemically reduced forms. The plasticity of the metal-ligating residues and the iron atoms of the dinuclear metallocenter are discussed alongside comparisons with the seven-residue metallocenter of rubrerythrin. The structure of symerythrin reveals unprecedented levels of internal symmetry among FLSF proteins, providing new experimental evidence in support of a symerythrin-like ancestor for the rubrerythrin-like family. Unexpectedly, and in contrast with our previous prediction,⁶ we also show evidence that the single-chain four-helix bundle fold of the FLSF has had at least two independent origins.

Results

Expression and purification of recombinant symerythrin

Both *G. violaceus* and *C. paradoxa* symerythrin were expressed without an affinity tag in order to minimize artifactual metal binding during overexpression and handling. All studies reported here are based on the *C. paradoxa* protein, for which the expression was more efficient and reliable. Recombinant *C. paradoxa* symerythrin was obtained in major and minor forms (~10:1) migrating on denaturing gels at apparent molecular weights of 20 and 17 kDa, respectively. Mass spectrometric analyses showed that both forms were full-length symerythrin and both had the N-terminal methionine cleaved during overexpression in *E. coli*. The only difference between the two forms was that the lower band contained the Val127-Phe17 crosslink described earlier.¹⁴ For characterization, these isoforms could be separated from each other through common chromatographic procedures (see *Materials and Methods*).

Solution characterization of crosslinked and non-crosslinked symerythrin

In contrast with known rubrerythrins and rubrerythrin-like proteins, which are dimeric, analytical gel-filtration analysis demonstrates that crosslinked and non-crosslinked symerythrin are monomers at plausibly physiologic conditions (pH 7.2, 150 mM NaCl) (Supplemental Fig. 1). In terms of metal content, crosslinked symerythrin purified with ~2 Fe per monomer (1.7–1.8 by the ferrozine assay,¹⁵ 2.2–2.3 by inductively coupled plasma optical emission spectroscopy (ICP-OES)) and no detectable quantities of manganese or zinc. That the iron remains bound even after elution from a phenyl sepharose column shows that the metallocenter of crosslinked symerythrin is stable in up to 40% ethanol. Non-crosslinked symerythrin, on the other hand, purified with just 0.4–0.5 iron atoms per monomer. Reconstitution of the non-crosslinked form by incubation with excess ferrous iron resulted in 1.6–1.8 iron atoms per monomer. As we reported earlier, the non-crosslinked iron center is labile to the iron chelator 1,10-phenanthroline, but the crosslinked metallocenter is not.¹⁴

[†]The protein now named symerythrin¹⁴ was originally referred to by us⁶ and Andrews² as ‘erythrin’ because it appeared to have a rubrerythrin-like metallocenter and no rubredoxin domain. Later, Andrews¹ used ‘erythrin group’ for a very diverse set of FLSF proteins without a rubredoxin domain and ‘erythrin subfamily’ for a specific subset having notable internal sequence similarity. Here, we use ‘erythrins’ to refer to this latter group.

The UV/visible spectra of resting iron-loaded symerythrin displayed a distinctive absorption band from 300–350 nm (Fig. 1a) typical of the oxo-to-Fe(III) charge transfer of an Fe(III)-Fe(III) (e.g. diferric) center.¹⁶ Anaerobic incubation of this diferric metallocenter with two reducing equivalents of sodium dithionite in the presence of methylviologen resulted in the loss of this band (Fig. 1a and inset). Additionally, the titration of diferric symerythrin with sodium azide yielded a new chromophore at 450 nm (Fig. 1b) as seen with other diferric-azide adducts.^{16; 17} The azide binding showed saturation with a K_d of ~60 mM (Supplemental Fig. 2). In all of these experiments, non-crosslinked and crosslinked symerythrin behaved similarly (Fig. 1 and Supplemental Fig. 3a).

The structure of oxidized crosslinked symerythrin

Overall fold—All structures of symerythrin reported here are of the crosslinked form because the non-crosslinked form did not crystallize in any of the tested conditions. The structure of oxidized crosslinked symerythrin as seen in crystals grown at pH 5.5 was solved by molecular replacement and refined at 1.20 Å resolution to an R/R_{free} of 0.101/0.125 with reasonable geometry (Table 1).¹⁴ The complete chain of crosslinked symerythrin, from residues 2–180, was modeled in both molecules in the asymmetric unit. No substantive differences existed between the two chains in the asymmetric unit (0.04 Å RMSD for 179 C α -atoms), and all descriptions of symerythrin provided here are based on molecule A.

Symerythrin adopts the expected up-down-down-up four-helix bundle fold characteristic of the FLSF with two metal atoms bound at the core of the protein (Figs. 2a and 2b). These metal ions were assigned as irons at 100% occupancy based on quantitative trace metal analyses. Helices B and C are connected by a 31 residue extended ‘linker’ which has a 5-residue stretch of a P_{II}-spiral (i.e. polyproline-II or polypeptide-II¹⁸) and two α -helices. The N-terminal segment preceding helix A adopts an extended conformation, including 5- and 6-residue segments of P_{II}-spirals, and lies in the groove between core helices A and C; it is covalently anchored to helix C via the Val127-Phe17 crosslink. The peptide chain terminates shortly after helix D with a short, 3-residue C-terminal tail.

Like rubrerythrins, symerythrin contains a π -helical segment in helix C (π C) adjacent to the metallocenter. In addition, as predicted,⁶ a π -helical segment (π A) is present in helix A of symerythrin (Fig. 2b). Each of these two π -helices has a Pro at its C-terminal end (Fig. 2a) and is associated with a ~20° kink in its core helix. These two kinked helices pack such that a pocket is created roughly halfway up the four-helix bundle adjacent to the metallocenter. This pocket is isolated from the outside solvent by the N-terminal tail, which is secured to the core four-helix bundle via the Val127-Phe17 crosslink (Fig. 2b). A DALI¹⁹ search revealed nigerythrin (PDB 1yux) as the structurally characterized protein most similar to symerythrin (Z-score = 17.9, 27% sequence identity, C α RMSD of 1.8 Å over 137 residues).

A novel variant of the FLSF metallocenter—Given eight potential metal ligands (six Glu and two His), describing the metallocenter coordination and comparing it with other FLSF proteins (having different residue numbers) can be rather confusing. To facilitate this, we introduce here a nomenclature in which each ligand residue is uniquely identified by the helix from which it comes, with the eight ligands thus being identified as Glu _{α A}, Glu _{π A}, Glu _{α B}, His _{α B}, Glu _{α C}, Glu _{π C}, Glu _{α D}, and His _{α D}. In symerythrin, these correspond to Glu37, Glu40, Glu71, His74, Glu128, Glu131, Glu162, His165, respectively (Fig. 2a). This nomenclature emphasizes the internal symmetry of the fold and allows residues involved in the metallocenter to be uniquely identified without the use of specific residue numbers. In addition to the direct metal ligands, FLSF metallocenters often have a conserved pair of Tyr residues from helices A and C that provide second-shell support for the metallocenter. In

symerythrin, these are Tyr45 and Tyr136. Designating these as Tyr_{αA} and Tyr_{αC}, respectively, Tyr_{αA} hydrogen-bonds to Glu_{αC} and Tyr_{αC} hydrogen-bonds to Glu_{αA}.

As seen at 1.20 Å resolution, the coordination geometry of the carboxylate-bridged diiron center is very well defined (Fig. 3a). The iron atoms are 3.6 Å apart, compared with just 3.3 Å in rubrerythrin. Both metals have octahedral coordination involving a μ -oxo/hydroxo bridge. The coordinating residues match those known in the seven-ligand metallocenters of rubrerythrins plus one additional residue, Glu_{πA}, acting as a monodentate ligand to Fe1. The electron density shows that Glu_{πA} adopts two equally populated conformations (Fig. 3a). The two Fe1 coordination sites occupied by the monodentate interactions of symerythrin's Glu_{αA} and Glu_{πA} are filled by a bidentate interaction of Glu_{αA} in rubrerythrins (Fig. 3b). This change makes the octahedral coordination of Fe1 in symerythrin less distorted than that of rubrerythrin. In contrast, the coordination geometries of Fe2 in symerythrin and rubrerythrin are nearly identical. Also as seen for other rubrerythrins, the His_{αB} side chain is not involved in metal coordination in the diferric state, but in contrast to those structures it enters a hydrogen bond with the non-coordinating carboxylate oxygen of Glu_{αA}. A 2.0 Å dataset collected at room temperature revealed that no significant changes in the metallocenter occur upon freezing of the crystals (data not shown).

The azide-bound diferric complex

A 250 mM sodium azide soak slightly impaired the diffraction of the crystals, and, as little to no density was present for residues 2-4, these were left out of the model. The electron density maps at 1.40 Å resolution revealed a mixture of resting state and azide-bound metallocenters. To guide the model building, an $F_o(\text{azide}) - F_o(\text{resting})$ difference map was generated to provide the most direct image of the changes induced by azide binding (Fig. 4a). This map unambiguously revealed the orientation of the azide as well as the altered coordinating residues; no notable backbone motions are involved. The data fit well with the metallocenter modeled as 50% resting state and 50% azide bound (R/R_{free} of 0.114/0.156, Table 1). Unfortunately, we were unable to obtain a structure of azide bound to the diferrous metallocenter because crystals were unstable in solutions with both azide and sodium dithionite. Also, attempts to generate a diferric peroxide-bound structure, as recently reported for rubrerythrin,¹³ showed no binding of hydrogen peroxide despite using soaks up to 250 mM H₂O₂.

In the azide-bound diferric metallocenter, the azide moiety displaces the bridging oxo/hydroxo-ligand and binds in a μ -1,1 orientation (Fig. 4b) similar to toluene monooxygenase²⁰ and a Y208F variant of the R2 subunit of ribonucleotide reductase.²¹ There are also substantial changes in the coordination geometry of three glutamates. Most notably, Glu_{πC} moves away from the Fe1 and is replaced by a terminally coordinating water molecule (HOH^{az}); this displacement of Glu_{πC} can be understood in that its former position is sterically incompatible with the second nitrogen of the azide. Glu_{αA} undergoes an unusual carboxylate shift in which its Fe1-ligating oxygen changes from Oε1 to Oε2, and Glu_{πA} becomes fixed in a single orientation. These alterations allow both Glu_{αA} and Glu_{πA} to hydrogen bond HOH^{az}. Fe1 and Fe2 shift ~0.3 Å and 0.1 Å, respectively, in the direction indicated by the difference map peaks (Fig. 4a). The geometry of Fe2 ligating residues remain unchanged.

The dithionite-reduced metallocenter adopts two conformers

To determine the structure of the diferrous metallocenter, crystals were soaked in an oxygen-free environment containing sodium dithionite and methyl viologen at pH 5.5. Structural changes happened only at the metallocenter and again involve no backbone shifts. The 1.25 Å resolution data were well accounted for using two distinct conformations of

equal population ($R/R_{\text{free}} = 0.115/0.141$, Fig. 5a, Table 1). We will refer to these as conformations **A** and **B**. In conformation **A** (Fig. 5b), the major change is an Fe1 shift of ~ 2 Å to position Fe1A with the iron-iron distance increasing to 4.1 Å. His_{αB} becomes a ligand and the ligating interactions with Glu_{πA}, Glu_{πC} and the bridging μ -oxo/hydroxo atom are lost. The resulting two open coordination sites are satisfied by a new water molecule very close the old position of Fe1 (HOH-A) and a change in Glu_{αA} coordination from monodentate to bidentate. This shift of Fe1 closely matches the redox dependent shift observed in rubrerythrins (Fig. 5c), however, in those structures 100% of Fe1 undergoes the shift.^{11; 12; 13}

In conformation **B** (Fig. 5c), Fe1 remains in the same position as it was in the oxidized structure, but its ligation changes. Both residues Glu_{πA} and Glu_{πC} remain as ligands but alter their conformation and ligating geometry from the oxidized structure. Because there is no electron density for the positions that Glu_{πA} and Glu_{πC} occupied in the reference oxidized structure we can conclude that this **B** conformation is not simply due to a portion of the crystal that has remained in the diferric state. Furthermore, although dithionite is a weaker reductant at pH 5.5 compared to pH 7 (-550 mV vs -660 mV vs NHE, respectively²²), it with methyl viologen as an electron mediator is a sufficiently strong reducing system at this acidic pH to fully reduce the diiron center of crosslinked symerythrin in solution (Supplemental Fig. 4). Also, trace metal analyses showed symerythrin contains no metals other than iron so that the 50% of Fe1 that did not move cannot be due to a different kind of redox insensitive metal.

Internal symmetry of symerythrin

A sequence alignment of the N- and C-terminal halves of the four-helix bundle of symerythrin demonstrates a 32% internal sequence identity.⁶ The structure of symerythrin shows that, at the level of the folded protein, the internal symmetry is even more striking. First, the backbones of the two halves of the core four-helix bundle overlay to within 0.77 Å (Fig. 6a), a much better internal symmetry than any other structurally characterized FLSF member (Table 2). Second, the π -helical bulges in helices A and C are perfectly symmetric in their placement. Third, the metalcenter is exquisitely symmetric as 102 atoms (51 atom pairs) comprising the irons and their ligands overlay to within 0.71 Å in oxidized symerythrin and to within 0.61 Å in dithionite-reduced symerythrin (Figure 6b).

Exploratory studies of enzymatic activity

The auto-generation of the Val-Phe crosslink near the active site of symerythrin proves its metalcenter has notable catalytic potential,¹⁴ yet the physiologic function of symerythrin remains unknown. The strong structural similarities of the symerythrin metalcenter with that of rubrerythrins makes it plausible that its activity is related. As rubrerythrin is known to have strong peroxidase and much weaker oxidase activity,^{13; 23} we tested symerythrin for such activities. In the absence of a known physiological reductant for symerythrin, for initial tests we exposed stoichiometrically reduced symerythrin to O₂ and monitored its return to the diferric state via recovery of its characteristic absorption band at 350 nm (Fig. 7a). This process of chemically reducing symerythrin followed by oxidation with O₂ could be repeated multiple times. Similarly, recovery of the 350 nm absorption band occurred upon exposure to hydrogen peroxide (Fig. 7b and Supplemental Fig. 3b). Crosslinked and non-crosslinked symerythrin behaved equivalently in these experiments demonstrating that both forms have both catalytic oxidase- and peroxidase-like activities.

Discussion

Since its amino acid sequence was first described sixteen years ago,²⁴ symerythrin was expected to have a rubrerythrin-like diiron center involving seven ligating residues. The structures we report here confirm our prediction⁶ that the symerythrin metallocenter is distinct from rubrerythrins in that it includes as an eighth metal ligating residue, a glutamate that comes from a π -helical segment in helix A of the four-helix bundle.

The metallocenter structure

With crystallographic data to 1.20 Å resolution, the structure of diferric symerythrin provides the highest resolution description of a carboxylate bridged dinuclear center in the FLSF. The high resolution was crucial in allowing us to sort out details of partially occupied conformations associated with each of the structures and especially the azide-bound and dithionite-reduced metallocenters. We suspect that much of this heterogeneity is due to the crystals being at pH 5.5, which compared to a more physiological pH near 7, would alter the protonation state of a substantial portion of any ligands with pK_a values in the range of 5 to 7.5. Although our analyses are not at a sufficient resolution to visualize H-atoms, we can make some inferences of protonation states based on the interactions present, which help in the interpretation of the structures.

First, in the diferric structure we infer that the bridging oxygen is a μ -hydroxo-bridge. This accounts for bond lengths to the irons being 2.1–2.2 Å, compared with values of ~1.8 Å expected for a μ -oxo-bridge,²⁵ and for the close (2.6 Å) H-bonding interaction with Glu $_{\pi C}$ for which we assume the Glu is the acceptor (Fig. 3a). Interestingly, one of the two conformations seen for Glu $_{\pi A}$ H-bonds to the bridging hydroxo, so for 50% of the molecules, Glu $_{\pi A}$ itself is protonated. Assuming these protonation states, the ligands sum to a net charge of –6 or –7 (for the two populations differing in the protonation state of Glu $_{\pi A}$), which matches reasonably well with the +6 charge of the two ferric ions. Finally, an H-bond interaction between His $_{\alpha B}$ and Glu $_{\alpha A}$ is consistent with the His being in the positively charged imidazolium state.

For the azide-bound diferric complex (Fig. 4b), the interactions of HOH^{az} with the metallocenter allow us to suggest that it is a water donating H-bonds to Glu $_{\alpha A}$ and Glu $_{\pi A}$, which are both deprotonated. Assuming these protonation states, with the swapping out of Glu $_{\pi C}$ ligation for HOH^{az} and with Glu $_{\pi A}$ always being deprotonated, the ligands in this complex sum to a net charge of –6, matching the +6 charge of the two ferric ions. That the azide-bound diferric structure is a mix of unliganded and liganded forms is not surprising given that the soak concentration (250 mM) is not much higher than the azide K_d of 60 mM determined at neutral pH.

In both conformations of the dithionite-reduced structure, Glu $_{\pi C}$ has moved so it points at Glu $_{\pi A}$ placing their carboxylate O ϵ 2-atoms just 2.6 Å apart (Fig. 5). A proton must be between these atoms, and we assign Glu $_{\pi C}$ as protonated because its C δ -O ϵ 2 bond length refined to a longer length (1.29 vs 1.25 Å) and has a less polar environment. In conformation **A** (with Fe1 shifted to position Fe1A; Fig. 5b), the two terminal oxygens can be identified as water molecules based both on the longer coordination distances of ~2.3 Å (2.26 – 2.36 Å) and on their network donating two H-bonds each to the O ϵ 1-atoms of Glu $_{\pi A}$ and Glu $_{\pi C}$. With these assignments, the conformation **A** ligands sum to a charge of –4 exactly matching the charge of the diferrous center. For conformation **B** (without the Fe1 shifted) little has changed from the diferric state, except Glu $_{\pi C}$ has become protonated. If the μ -oxo-bridge also became protonated to an aqua-bridge that would help the charge balance, but there is nothing in the structure to support that assignment. As no other rubrerythrin family member has adopted such a diferrous structure, we suggest that this conformation is not of catalytic

interest but is an artifact caused by the low pH of the crystal, which leads to His_{αB} being protonated and unable to serve as a metal ligand in half of the molecules in the crystal. Whereas a crystal structure at pH 7 would be desirable, the current crystal form disintegrated rapidly when placed neutral pH buffers.

Taken together, and discounting the heterogeneities that appear to be associated with the low pH of the crystals, the structures of these three forms of symerythrin reveal that from a structural perspective the metallocenter is remarkably similar to that of rubrerythrins with the diferric and diferrous structures matching those determined for other rubrerythrin family members (Figs. 3b and 5c). And although the azide bound diferric complex provides the first view of such a complex for this family, the μ -1,1 binding mode of azide to the diferric metallocenter of symerythrin is quite similar to how hydrogen peroxide binds to diferric rubrerythrin.¹³ This similarity suggests that the μ -hydroxo ligand of the symerythrin metallocenter is also labile and could be displaced by substrates, products or reaction-intermediates. Unfortunately our efforts to determine an azide complex with diferrous symerythrin were not successful so we cannot compare that with the rubrerythrin result.¹¹

Considerations of symerythrin function

Previously we demonstrated a correlation between the gain and loss of π -helices with the functional diversification of protein families.⁶ The elucidation of symerythrin's structure confirms our prediction that its unique π -helix is involved in the formation of a novel metallocenter, but it is not yet clear if or how these changes influence the chemistry of the symerythrin metallocenter. In terms of gaining insight into symerythrin function, the *in vitro* assays we have carried out show that both oxygen and hydrogen peroxide can be reduced by symerythrin. From a structural viewpoint, most important to consider is the diferrous structure which is both the form from which a catalytic cycle begins and the form that would have predominated in the anaerobic world at the time of the origin of the rubrerythrin family. Compared to the rubrerythrins (Fig. 5c), the symerythrin diferrous structure is a remarkable match, having equivalent coordination geometry and a very similar active site pocket, simply having one more glutamate, Glu _{π A}, present making the active site pocket more polar and more negatively charged. The extra Glu _{π A} would also serve to make the interactions with hydrogen peroxide more symmetric. In this light, the symerythrin active site is perfectly compatible with the peroxidase mechanism proposed for rubrerythrins where the peroxide binds the diferrous enzyme in place of the two terminal water ligands.^{11; 13}

Although inspection of the genomic context of symerythrin in the genomes of *C. paradoxa* and *G. violaceus* did not reveal additional clues about its physiologic function or interacting partner molecules, a reasonable working hypothesis is that symerythrin will also have peroxidase or closely related activity. As is observed for rubrerythrin, it seems non-polar substrates like oxygen and negatively charged substrates like superoxide would be discriminated against by this active site environment. The extra negative charge density could also provide stabilization for the previously proposed¹⁴ high-valent Fe(IV)-oxygen intermediates involved in the formation of the Val127-Phe17 crosslink.

Finally, we note that in these structural studies we have only been able to visualize symerythrin forms containing the Val-Phe crosslink, but we do not yet know whether symerythrin in its natural setting contains the crosslink. Indeed, if the protein is in a microaerobic environment within the cells, the crosslink may not exist since its formation is dependent on oxygen.¹⁴ Both the preferential crystallization of the crosslinked form and the observed stabilization it provides the metallocenter¹⁴ show that the crosslinked form is more conformationally rigid. However, the crosslinked and non-crosslinked proteins behave similarly in all of the spectroscopic and activity studies, which we take as evidence that the structures reported here are also relevant for non-crosslinked symerythrin. Supporting this

conclusion is the observation that the symerythrin metallocenter structures agree well with other rubrerythrin family members, none of which have crosslinks. We are now working to culture these phototrophs in order to identify symerythrin's presence, location, and crosslink status in its natural setting, as well as other properties such as its metal content, its interacting protein partners, and how its expression responds to different environmental conditions.

Insights into the evolution of the rubrerythrins and the FLSF

As first described by Andrews,² rubrerythrin family members (and especially symerythrin[†]) have a notable level of internal two-fold sequence similarity at the subdomain level which when seen for proteins is considered an ancestral feature reflecting that the protein fold originated through the duplication and fusion of a gene encoding a half-protein that assembled into a symmetric dimer (e.g. ^{26; 27}). For the rubrerythrin family, this putative ancestor would have consisted of a two-helix peptide having metal binding residues in the appropriate positions, and given the ubiquitous distribution of rubrerythrins in anaerobic organisms,⁹ the fold would seem to have originated before the advent of oxygen in the atmosphere. One plausible scenario for the origin of the family is that the ancestral form of the rubrerythrin-symerythrin families had a rubrerythrin-like metallocenter and the symerythrins were formed via the insertion of an eighth metal-ligating residue in helix A. Given the very narrow phylogenetic distribution of the 8-residue metallocenter variant compared with the ubiquitous presence of the 7-residue metallocenter,⁹ this scenario would seem attractive.

However, an alternative scenario that we proposed previously,⁶ is that the seven residue metallocenter version was formed by the deletion of the Glu_{πA} metal-ligating residue from a symerythrin-like ancestor. Now, along with the greater internal sequence symmetry of symerythrin compared to rubrerythrins, additional evidence for the ancestral form being symerythrin-like comes from the much higher internal backbone symmetry of symerythrin (Table 2, Fig. 6a), the additional symmetry associated with the two π -helices, and the near perfect internal symmetry of the metallocenter itself (Fig. 6b). By Dollo parsimony,²⁸ the principle that complex features are easier to lose than gain during evolution, the remarkable internal similarities of symerythrin make a compelling case that its eight-ligand motif is an ancestral trait of the family. The narrow phylogenetic distribution of symerythrin would be understandable if the seven-residue version of the metallocenter had improved functionality allowing it to outcompete the ancestral form over time. Interesting in this light is that the two organisms in which symerythrin has been preserved are not anaerobes, but are both aerobic organisms that are among the early organisms that carried out oxygenic photosynthesis. They have both been referred to as “living fossils,” a term coined by Darwin to refer to organisms that have changed little over time because their niche remains constant and have experienced little competition in that niche. Based on this concept, the relatively slow evolution of the rubrerythrin family proteins such that residual internal symmetry from their origin has been maintained suggests that their function has changed little over time.

So if the seven-residue metallocenter was derived from an eight-residue metallocenter, how did the more predominant six-residue metallocenter of the FLSF arise? One possible scenario is the six-residue metallocenter splintered off from a rubrerythrin-like ancestor via the deletion of the seventh metal-ligating residue. Alternatively, the six-residue metallocenter could have formed independently (one or more times) via a distinct gene-duplication and fusion event of a variant of the dimerizing two-helix ancestral peptide that had three rather than four metal-ligating residues. Sequence comparisons lend support to such a model. In particular, there exists one FLSF family with a six-ligand metallocenter for which family members have residual internal sequence identity at a level similar to that of the rubrerythrins (~30% identity), and for which a crystal structure shows high internal

structural similarity (RMSD of 1.1 Å, Table 2). Because of their similarly high internal sequence similarity, Andrews¹ had considered these proteins to be closely grouped with the rubrerythrins, naming them ‘erythrins’ because they do not have a rubredoxin domain.

However, a remarkable finding not noted by Andrews¹ is that these erythrins are actually not very sequence similar to the rubrerythrin family (~16% identity). In fact, aside from the four metalcenter-associated residues in the half-protein (2 Glu, 1 His, and 1 Tyr), there are 14 positions showing strong internal sequence symmetry in the erythrins and 15 positions showing strong internal symmetry in the rubrerythrins but only one match between the two (Fig. 8a). This very low congruence approximates what would be expected for two independent sequences and is not compatible with a model in which the internal sequence similarities are conserved from a common ancestor. Thus, this comparison provides very strong evidence for independent origins for these two families resulting from the fusion of distinct two-helix peptides (Fig. 8c). Further support for this conclusion is another feature symmetric within each family, but differing between them: the loops connecting the A to B (or C to D) helices in erythrins are three residues shorter than those in the rubrerythrin family (Fig. 8a,b).

These observations support the hypothesis that the single-chain FLSF fold originated at least twice via independent gene duplication/fusion events of what may have been homologous two-helix peptides (Fig. 8c). Evidence that this process of FLSF fold generation may continue to this day comes from a modern-day family of proteins that are two-helix peptides that assemble into the topology of the FLSF four-helix bundle containing the key residues of the canonical FLSF six-residue metalcenter. These proteins have not been functionally characterized, but are known from the crystal structure of a protein from *Nitrosomonas europaea* (PDB 3k6c). This two-helix peptide has a loop length matching the erythrins rather than the rubrerythrins (Fig. 8a), but as phylogenetic analyses indicate it is highly divergent from all known FLSF proteins (Supplemental Fig. 6), it cannot be determined how (and even if) it relates to any current single-chain FLSF superfamily members.

As for the more divergent FLSF proteins, these proteins have no notable internal sequence symmetry beyond the metal-ligating residues themselves (Table 2). We would suggest that for these other FLSF superfamily members there is insufficient evidence at this time to conclude whether they derived from the rubrerythrin line, the erythrin line or from one or more additional independent gene-fusion events of homo- or hetero-dimers of two-helix peptides having the appropriate ligands. In any case, the strong evidence for independent gene-fusion events giving rise to the rubrerythrin and erythrin families of the FLSF implies that our earlier proposal that the whole FLSF derived from a symerythrin-like ancestor⁶ was an oversimplification, and that previous assumptions of a single gene-fusion event origin for all FLSF families^{1, 2} must be revisited. Nevertheless, the insights generated through the characterization of symerythrin presented here underscore the value of using π -helices as guides to discover structure-function and evolutionary relationships in proteins.

Materials and Methods

Expression and purification of symerythrin from *Cyanophora paradoxa*

Development of a recombinant expression system—ORF180 from the cyanelle genome of *C. paradoxa* (NCBI Gene ID: 801647) was purchased from GenScript (Piscataway, NJ) in the pUC57 cloning vector with NdeI and BamHI restriction sites added to the 5' and 3' ends of the gene, respectively. This plasmid was transformed into chemically competent *E. coli* DH5 α cells. After cell growth in the presence of 100 μ g/ml ampicillin, plasmid DNA was isolated and digested with NdeI and BamHI. The resulting 0.5 kb insert encoding the symerythrin gene was gel purified using the QIAquick Gel Extraction Kit

(Qiagen) and ligated to the NdeI/BamHI-digested pT7-7 expression vector. The resulting plasmid, pT7CPERY, was transformed by electroporation into *E. coli* BL21(DE3) (Novagen) cells for subsequent expression and purification.

Expression and purification of symerythrin—*E. coli* BL21(DE3) cells containing the pT7CPERY plasmid were grown in 4 L of LB media containing 100 µg/ml Ampicillin at 37 °C to OD₆₀₀ 0.6 – 0.8. Expression of symerythrin was then induced with 1 mM isopropyl β-D-1-thiogalactopyranoside, at which time 150 µM FeSO₄ was also added. After 5 h, cells were harvested by centrifugation at 5000 g for 15 minutes and frozen at –70 °C. Yields of 3–4 g cell paste per liter culture were typical.

For protein purification, 12–14 g cell paste was suspended in 25 mM MOPS, 5% glycerol and 50 mM NaCl pH 7.2 to a total volume of 40 ml. Cells were lysed by two or three passes through a French press at 52,000 Pa, and then centrifuged at 10,000 g for 10 min. The supernatant was carefully decanted, diluted to 60 ml and centrifuged at 150,000 g for 1 h. The resulting supernatant was decanted, adjusted to pH 7.2. This cell free extract was loaded at 2.0 ml/min onto a DEAE Sepharose FF column (2.5 cm × 17 cm) pre-equilibrated with 25 mM MOPS, 5% glycerol and 50 mM NaCl pH 7.2. The column was washed with five column volumes (400 ml) of the same buffer at a flow rate of 1.5 ml/min, after which a linear 400 ml gradient from 0 to 0.5 M NaCl was applied. Each 10 ml fraction was analyzed by SDS-PAGE: non-crosslinked symerythrin migrated with an apparent molecular weight of 20 kDa, while crosslinked symerythrin migrated with an apparent molecular weight of 17 kDa. Both crosslinked and non-crosslinked symerythrin eluted between 0.1 and 0.15 M NaCl (pool I), while enriched non-crosslinked symerythrin eluted between 0.15 and 0.2 M NaCl (pool II).

Purification of non-crosslinked symerythrin—Fractions from pool II of the anion exchange fractionation were concentrated to 2.5 ml by ultrafiltration and applied to a Superdex-75 gel filtration column (2.5 cm × 55 cm) equilibrated with 25 mM MOPS, 5% glycerol and 150 mM NaCl pH 7.2 at a flow rate of 0.25 ml/min. The purity of the resulting fractions was analyzed by SDS-PAGE. Fractions containing pure non-crosslinked symerythrin were pooled, repeatedly exchanged into 10 mM Tris pH 7.4 by ultrafiltration, and concentrated to ~30 mg/ml before being flash frozen in liquid N₂. From 4 L of culture, 15–20 mg of non-crosslinked symerythrin was typically obtained.

Purification of crosslinked symerythrin—Fractions from pool I of the anion exchange chromatography were concentrated to 2.5 ml by ultrafiltration and applied to the same Superdex-75 column described above. Both non-crosslinked and crosslinked symerythrin eluted simultaneously but separate from other contaminants. The pooled fractions were adjusted to 1 M NH₄Cl from a 4 M stock to ensure proper binding to a Phenyl Sepharose CL4B column (2 cm × 90 cm). After application to the column, and washing with two column volumes of buffer (25 mM MOPS and 175 mM NH₄Cl, pH 7.2), the protein was eluted with a 750 ml (five column volume) linear gradient from 0–50% ethanol in 25 mM MOPS, pH 7.2. Crosslinked symerythrin eluted at ~35–40% ethanol while the non-crosslinked form eluted at ~40–45% ethanol. Fractions containing >95% crosslinked symerythrin were pooled together, exchanged into 10 mM Tris pH 7.4 and concentrated to ~15 mg/ml prior to flash freezing in liquid N₂. From 4 L of culture, ~2–3 mg of crosslinked symerythrin was obtained.

Biochemical analyses of purified symerythrin

Protein concentrations were determined by BCA assays (Pierce) using bovine serum albumin as a standard. Analytical gel filtration was used to determine the quaternary

structure of crosslinked and non-crosslinked symerythrin at concentrations ~5 mg/ml. In-solution proteolytic digests of crosslinked and non-crosslinked forms were conducted and analyzed by mass spectrometry as previously described.¹⁴

Trace metal analysis—Iron content was typically determined by the ferrozine colorimetric assay.¹⁵ For determining the content of iron, manganese and zinc by Inductively coupled plasma optical emission spectroscopy (ICP-OES), 10 μ L of 70% trace metal grade HNO₃ was added to an equal volume of 15 mg/ml symerythrin. The mixture was incubated at 60 °C for 3 h and then diluted to 5 ml in metal-free water for subsequent analysis.

Metal reconstitution—Purified non-crosslinked symerythrin was incubated under argon with 1 mM 1,10-phenanthroline and 5 mM EDTA at 2 mg/ml in 25 mM MOPS pH 7.2, 150 mM NaCl, 5% glycerol and 1 mM dithionite. After 3 h at room temperature, the resulting metal-free symerythrin was repeatedly exchanged into 10 mM Tris pH 7.4, concentrated to ~20 mg/ml and flash frozen in liquid nitrogen.

To prepare Fe-reconstituted non-crosslinked symerythrin, metal-free protein was diluted to 2 mg/ml (100 μ M) in a sealed vial containing an anaerobic solution of 25 mM MOPS pH 7.2, 150 mM NaCl and 5% glycerol. Anaerobically prepared 20 mM stock solutions of sodium dithionite and Fe(NH₄)₂(SO₄)₂ were added to this solution to final concentrations of 2 mM and 1 mM, respectively. All manipulations of anaerobic solutions were carried out using gas-tight Hamilton syringes. After 3 h at room temperature under argon, excess dithionite and iron were removed by dialysis. Typically, 1.7 to 1.8 Fe per active site were obtained from this procedure.

UV/Visible spectroscopic characterization—Identical procedures were used for non-crosslinked and crosslinked symerythrin. Iron reconstituted, symerythrin was diluted to 1 mg/ml into an anaerobically prepared solution of 25 mM MOPS pH 7.2, 150 mM NaCl and 50 μ M methyl viologen. This solution was transferred to a 1 cm path length quartz cuvette purged with argon. Spectra were recorded at room temperature using a Beckman DU-640 spectrophotometer. Sodium dithionite was then added from an anaerobically prepared 1 mM stock solution and incubated for 3 minutes, at which time the spectrum was recorded. Subsequent dithionite additions were repeated until absorption peaks characteristic of reduced methyl viologen were observed, indicating all symerythrin metallocenters were reduced. At this time, the cuvette was exposed to either O₂ by removing the septa on the cuvette or H₂O₂ by stepwise additions of an anaerobically prepared 1 mM hydrogen peroxide stock solution. Azide-bound spectra were recorded after aerobically incubating a 1 mg/ml symerythrin solution with various quantities of sodium azide from a 5 M stock solution in 25 mM MOPS, 150 mM NaCl, pH 7.2.

Consensus sequences for rubrerythrin- and erythrin-like sequences

Non-redundant sequences of rubrerythrin family members were obtained from the NCBI database by performing a PSI-BLAST search using *D. vulgaris* rubrerythrin (PDB 1lkm). Using a 50% sequence identity threshold, 250 rubrerythrin sequences were obtained and aligned via the CLUSTALW algorithm. The consensus sequence of these aligned sequences was derived using a SeqLogo plot²⁹ generated by the WebLogo server.³⁰ Residues in each position were considered highly, moderately, or poorly conserved if the identity of the top amino acid had a bit value of at least 2/3, between 2/3 and 1/3, and less than 1/3 the maximum bit value, respectively. A similar PSI-BLAST search was performed using the erythrin sequence from *Pyrococcus furiosus* (PDB 2fzf) for which 11 sequences matched the same 50% sequence identity threshold used for rubrerythrins. For the oligomerizing $\alpha\alpha$ -

hairpins, the sequence of PDB 3k6c was used for the PSI-BLAST search. This resulted in 80 sequences with at least 50% sequence identity. Alignments and consensus sequences for these protein families were obtained as for rubrerythrins. For the consensus sequence of symerythrins, residues were considered highly conserved if the residues were identical in the two symerythrin sequences. For the internal similarity assessment, a residue was considered moderately conserved if it matched an internally symmetric highly conserved residue.

Crystallization and Structure Determination

Crystallization and crystal handling—Crystals were grown at 6 °C in hanging drops on siliconized cover slips from 2.5 μ l of purified crosslinked symerythrin (8 mg/ml in 10 mM Tris pH 7.4) and 1 μ l of reservoir solution consisting of 22–28% PEG 3350, 0.1 M Bis-Tris pH 5.5 and 0.35–0.4 M NH_4Cl . Diffraction quality crystals ($400 \times 75 \times 75 \mu\text{m}$) were observed within 4–10 days. No crystals grew when non-crosslinked protein was used, and irregular crystal growth was observed when heterogeneous mixtures of crosslinked and non-crosslinked symerythrin were used. Purification of crosslinked symerythrin was therefore essential for obtaining diffraction quality crystals under these conditions. For data collection of all forms, single crystals were suspended on a rayon loop, briefly dipped into paratone oil and immediately frozen in liquid N_2 .

To form the azide-bound diferric complex, crystals were incubated for 20 min in an artificial mother liquor (AML) of 28% PEG 3350, 0.37 M NH_4Cl and 0.1 M BisTris pH 5.5 supplemented with 250 mM sodium azide. Attempts to obtain greater occupancies of azide binding were not successful as the use of higher concentrations of sodium azide or longer incubation periods resulted in the disintegration of the crystals.

For preparing the dithionite-reduced form, crystals before freezing were incubated for 45 min under argon in anaerobically prepared solution of 5 mM dithionite and 0.5 mM methylviologen in AML. The blue color of reduced methyl viologen was present during the course of the incubation, confirming the solution remained anaerobic. To ensure the results were not due to partial or incomplete reduction, we varied the dithionite concentration from 1 mM to 20 mM and the length of incubation from 20 min to 3 h; no differences were found between these treatments. Furthermore, to be sure dithionite was a sufficiently strong reductant to fully reduce the diiron center at the pH of the AML, in solution UV/Visible spectroscopic characterizations of crosslinked symerythrin were performed as described above except that symerythrin was diluted in 100 mM BisTris pH 5.5 and 50 μM methyl viologen before the titration with dithionite. To address the possibility that the paratone oil in which the crystals were frozen contained small quantities of oxygen that could affect the structure of the reduced metallocenter, crystals were soaked as above except with 15% PEG400 added as a cryoprotectant so that after the 45 min incubation, crystals could be directly frozen in liquid N_2 . The results were unchanged. Lastly, symerythrin crystals were similarly incubated in reducing agents Ti(III)citrate and ascorbic acid, however no differences from the resting diferric structure were observed.

Data Collection and structure determination—Data were collected using beamline 5.0.3 and 5.0.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory). Data were processed and scaled using iMosflm v1.0.040³¹ and SCALA³², and 5% of the data were flagged for use in R_{free} . Statistics are in Table 1.

The structure of differic symerythrin was solved by molecular replacement using the Phenix software package (v.1.7-650).³³ Rubrerythrin (PDB 1lkm) without the rubredoxin domain, iron atoms and water molecules was used as the initial search model and a solution was obtained with space group $\text{P}3_2$ and 2 molecules in the asymmetric unit (initial $R/R_{\text{free}} = 0.294/0.353$). In each monomer, two $2\text{F}_o - \text{F}_c$ peaks of $>20 \rho_{\text{rms}}$ were observed in a similar

position as the dinuclear center of rubrerythrin. Modeling two iron atoms in these peaks dropped the R/R_{free} to 0.288/0.323, after which the sequence of symerythrin from residues 2-180 was built manually using Coot³⁴ and refined with REFMAC.³⁵ For this and all subsequent structures, anisotropic B-factors and riding hydrogens were used during refinements. Standard criteria were used for modeling water molecules ($>1 \rho_{\text{rms}}$ intensity in the $2F_o - F_c$ map, $>2.4 \text{ \AA}$ distance from nearest contact). Molprobity³⁶ was used to monitor the model geometry. For the final rounds of refinement, B-factor and geometry restraint weights were optimized. The final R/R_{free} of the diferric structure was 0.101/0.125 (Table 1). For the $2F_o - F_c$ electron density maps shown in Figures 3 and 5, composite omit maps using a simulated annealing protocol were calculated using Phenix. These composite $2F_o - F_c$ omit maps are nearly identical to the final $2F_o - F_c$ maps, indicating the phases of the latter were not significantly model-biased.

For refinement of the dithionite-treated structure, the diferric structure (with water molecules) was used as the starting model except that both iron atoms and all eight side chains of the inner sphere iron-ligating residues from both chains were removed to minimize model bias (initial $R/R_{\text{free}} = 0.194/0.205$). Iron atoms and side chains of the metallocenter were then manually built into both chains. Water molecules were added or removed according to the criteria described for the oxidized structure. As with the oxidized structure, interpretable electron density was observed for residues 2-180 in both chains. Final refinements were done as for the diferric state, leading to a final R/R_{free} of 0.115/0.141 (Table 1).

The azide bound structure (initial $R/R_{\text{free}} = 0.238/0.286$) was solved using the same process as the dithionite structure. The positions of both iron atoms and the side chains of Glu_{αB}, His_{αB}, Glu_{αC}, Glu_{αD} and His_{αD} were not substantially different from the unliganded structure and were therefore modeled first ($R/R_{\text{free}} = 0.137/0.175$). $2F_o - F_o$ density for the remaining ligating residues revealed a mixture of states consisting of the unliganded conformation and an azide-induced conformation. To facilitate building the azide and the azide-induced conformations of Glu_{αA}, Glu_{πA} and Glu_{πC}, an $F_o(\text{azide}) - F_o(\text{unliganded})$ difference map was generated using phases from the unliganded model (Fig. 4a). Modeling the occupancies of the unliganded and azide-bound conformations at 0.5 each gave consistent B-factors between the iron atoms and all ligands, including the azide molecule. With the slightly lower resolution of these data compared with the oxidized and dithionite-treated data, residues 2-4 had much weaker density and were not modeled. π -helices were confirmed by analysis with DSSP³⁷ and π -HUNT.⁶

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the General Medical Institute, NIH grant GM R01-13 083136 and the Oregon Agricultural Experiment Station.

References

1. Andrews SC. The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor. *Biochim Biophys Acta*. 2010; 1800:691–705. [PubMed: 20553812]
2. Andrews, S. Iron Storage in Bacteria. In: Poole, RK., editor. *Advances in Microbial Physiology*. Vol. 40. Academic Press; San Diego: 1998. p. 281-349.

3. Cotruvo JA, Stubbe J. *Escherichia coli* class Ib ribonucleotide reductase contains a dimanganese(III)-tyrosyl radical cofactor *in vivo*. *Biochemistry*. 2011; 50:1672–81. [PubMed: 21250660]
4. Jiang W, Yun D, Saleh L, Barr EW, Xing G, Hoffart LM, Maslak MA, Krebs C, Bollinger JM Jr. A manganese(IV)/iron(III) cofactor in *Chlamydia trachomatis* ribonucleotide reductase. *Science*. 2007; 316:1188–91. [PubMed: 17525338]
5. Sazinsky MH, Lippard SJ. Correlating structure with function in bacterial multicomponent monooxygenases and related diiron proteins. *Acc Chem Res*. 2006; 39:558–66. [PubMed: 16906752]
6. Cooley RB, Arp DJ, Karplus PA. Evolutionary Origin of a Secondary Structure: π -Helices as Cryptic but Widespread Insertional Variations of α -Helices That Enhance Protein Functionality. *J Mol Biol*. 2010; 404:232–46. [PubMed: 20888342]
7. Coulter ED, Shenvi NV, Beharry ZM, Smith JJ, Prickril BC. Rubrerythrin-catalyzed substrate oxidation by dioxygen and hydrogen peroxide. *Inorg Chim Acta*. 2000; 297:231–241.
8. Zhao W, Ye Z, Zhao J. RbrA, a cyanobacterial rubrerythrin, functions as a FNR-dependent peroxidase in heterocysts in protection of nitrogenase from damage by hydrogen peroxide in *Anabaena* sp. PCC 7120. *Mol Microbiol*. 2007; 66:1219–30. [PubMed: 18001348]
9. Gomes CM, Le Gall J, Xavier AV, Teixeira M. Could a diiron-containing four-helix-bundle protein have been a primitive oxygen reductase? *Chembiochem*. 2001; 2:583–7. [PubMed: 11828492]
10. Kurtz DM Jr, Prickril BC. Intrapeptide sequence homology in rubrerythrin from *Desulfovibrio vulgaris*: identification of potential ligands to the diiron site. *Biochem Biophys Res Comm*. 1991; 181:337–41. [PubMed: 1958203]
11. Jin S, Kurtz DM Jr, Liu ZJ, Rose J, Wang BC. X-ray crystal structures of reduced rubrerythrin and its azide adduct: a structure-based mechanism for a non-heme diiron peroxidase. *J Am Chem Soc*. 2002; 124:9845–55. [PubMed: 12175244]
12. Iyer RB, Silaghi-Dumitrescu R, Kurtz DM Jr, Lanzilotta WN. High-resolution crystal structures of *Desulfovibrio vulgaris* (Hildenborough) nigerythrin: facile, redox-dependent iron movement, domain interface variability, and peroxidase activity in the rubrerythrins. *J Biol Inorg Chem*. 2005; 10:407–16. [PubMed: 15895271]
13. Dillard BD, Demick JM, Adams MW, Lanzilotta WN. A cryo-crystallographic time course for peroxide reduction by rubrerythrin from *Pyrococcus furiosus*. *J Biol Inorg Chem*. 2011 E-pub ahead of print.
14. Cooley RB, Rhoads TW, Arp DJ, Karplus PA. A diiron protein autogenerates a valine-phenylalanine cross-link. *Science*. 2011; 332:929. [PubMed: 21596985]
15. Percival MD. Human 5-lipoxygenase contains an essential iron. *J Biol Chem*. 1991; 266:10058–61. [PubMed: 2037564]
16. Fox BG, Shanklin J, Somerville C, Munck E. Stearoyl-acyl carrier protein Δ^9 -desaturase from *Ricinus communis* is a diiron-oxo protein. *Proc Natl Acad Sci U S A*. 1993; 90:2486–90. [PubMed: 8460163]
17. Gupta N, Bonomi F, Kurtz DM Jr, Ravi N, Wang DL, Huynh BH. Recombinant *Desulfovibrio vulgaris* rubrerythrin. Isolation and characterization of the diiron domain. *Biochemistry*. 1995; 34:3310–8. [PubMed: 7880826]
18. Hollingsworth SA, Karplus PA. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomol Concepts*. 2010; 1:271–283. [PubMed: 21436958]
19. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. Searching protein structure databases with DaliLite v.3. *Bioinformatics*. 2008; 24:2780–1. [PubMed: 18818215]
20. Sazinsky MH, Bard J, Di Donato A, Lippard SJ. Crystal structure of the toluene/o-xylene monooxygenase hydroxylase from *Pseudomonas stutzeri* OX1. Insight into the substrate specificity, substrate channeling, and active site tuning of multicomponent monooxygenases. *J Biol Chem*. 2004; 279:30600–10. [PubMed: 15096510]
21. Andersson ME, Hogbom M, Rinaldo-Matthis A, Andersson KK, Sjöberg BM, Nordlund P. The crystal structure of an azide complex of the diferrous R2 subunit of ribonucleotide reductase displays a novel carboxylate shift with important mechanistic implications for diiron-catalyzed oxygen activation. *Journal of the American Chemical Society*. 1999; 121:2346–2352.

22. Mayhew SG. The redox potential of dithionite and SO₂ from equilibrium reactions with flavodoxins, methyl viologen and hydrogen plus hydrogenase. *Eur J Biochem.* 1978; 85:535–47. [PubMed: 648533]
23. Riebe O, Fischer RJ, Wampler DA, Kurtz DM Jr, Bahl H. Pathway for H₂O₂ and O₂ detoxification in *Clostridium acetobutylicum*. *Microbiology.* 2009; 155:16–24. [PubMed: 19118342]
24. Stirewalt VL, Michalowski CB, Löffelhardt W, Bohnert HJ, Bryant DA. Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. *Plant Mol Biol Repr.* 1995; 13:327–332.
25. Kurtz DM. Oxo-Bridged and Hydroxo-Bridged Diiron Complexes - a Chemical Perspective on a Biological Unit. *Chemical Reviews.* 1990; 90:585–606.
26. Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M. Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science.* 2000; 289:1546–50. [PubMed: 10968789]
27. Dauter Z, Wilson KS, Sieker LC, Meyer J, Moulis JM. Atomic resolution (0.94 Å) structure of *Clostridium acidurici* ferredoxin. Detailed geometry of [4Fe-4S] clusters in a protein. *Biochemistry.* 1997; 36:16065–73. [PubMed: 9405040]
28. Farris J. Phylogenetic analysis under Dollo's Law. *Syst Zool.* 1977; 26:77–88.
29. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 1990; 18:6097–100. [PubMed: 2172928]
30. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004; 14:1188–90. [PubMed: 15173120]
31. Leslie A. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4+ ESF-EAMCB newsletter on protein crystallography.* 1992:26.
32. Evans, PR. Recent advances in phasing. In: Wilson, KS.; Davies, G.; Ashton, AW.; Bailey, S., editors. *Proc CCP4 Study Weekend.* Daresbury Laboratory; Warrington, UK: 1997. p. 97-102.
33. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* 2010; 66:213–21. [PubMed: 20124702]
34. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr.* 2010; 66:486–501. [PubMed: 20383002]
35. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr.* 1997; 53:240–55. [PubMed: 15299926]
36. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 2007; 35:W375–83. [PubMed: 17452350]
37. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 1983; 22:2577–637. [PubMed: 6667333]
38. Diederichs K, Karplus PA. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol.* 1997; 4:269–75. [PubMed: 9095194]
39. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. 2010.

Figure 1.

Spectra of oxidized, reduced and azide bound symerythrin. (A) Reduction (dotted arrow) converts spectra of diferric non-crosslinked (solid gray line) and crosslinked (solid black line) symerythrin to spectra of diferrous symerythrin (dashed gray and black lines, respectively). Inset: The change in absorption at 340 nm as a function of reducing equivalents during a dithionite titration of diferric, crosslinked symerythrin. Equivalent results were obtained for the non-crosslinked isoform (Supplemental Fig. 3a). (B) Spectra of diferric non-crosslinked and crosslinked symerythrin before (gray and black lines, respectively) and after (dashed gray and black lines, respectively) addition of 0.5 M sodium azide.

Figure 2.

Structure of crosslinked diferric symerythrin. (A) The sequence of symerythrin with secondary structure segments annotated: P_{II} spirals¹⁸ (tilda shapes), α -helices (cylinders – with A and B green, C and D blue, and linker helices grey), and π -helices (orange rectangles). Metal ligands (including the second-shell Tyr residues) are outlined in red with the seventh and eighth metal-ligating residues distinguished by purple and pink backgrounds, respectively. The Val127-Phe17 crosslink is indicated using yellow boxes and lines. Black dots are located above every tenth residue. (B) Stereoview ribbon diagram of symerythrin emphasizing the α -units of helices A–B (green) and C–D (blue), the N-terminal tail (dark red), the linker between helices B and C (gray), and the π -helical segments π A and π C (orange). Iron atoms (brown), atoms in the Val127-Phe17 crosslink (yellow) and metalcenter residues (colored as in Fig. 3) are shown as balls-and-sticks. The N- and C-termini are labeled. Molecular graphics were prepared using PYMOL.³⁹

Figure 3.

Stereoviews of the diferric metallocenter of symerythrin and its comparison with rubrerythrin. (A) Ball-and-stick model of the symerythrin metallocenter showing the six canonical metal-ligating residues of the FLSF (gray carbons), the seventh ligating residue characteristic of rubrerythrins (Glu131; purple carbons), and the eighth ligating residue unique to symerythrin (Glu40; pink carbons). Nitrogen and oxygen atoms are blue and red, respectively. A $2F_o - F_c$ composite annealed omit map (see *Materials and Methods*) is shown for the side chains of Glu131 and Glu40 (contoured at $2.0 \rho_{rms}$) and for Fe1 and Fe2 (brown spheres) and the μ -hydroxo bridge (contoured at $5.0 \rho_{rms}$). (B) Overlay of the diferric metallocenter of symerythrin (as in panel A) with the diferric metallocenter of rubrerythrin from *D. vulgaris* (PDB 1lkm) shown semi-transparently with carbon colored cyan and other atoms colored as for symerythrin. In panel B and further figures, metallocenter residues are labeled according to the helix on which they reside, as described in the main text.

Stereoviews of the azide-bound diferric metallocenter of symerythrin. (A) The unliganded diferric metallocenter model overlayed on the $F_o(\text{azide}) - F_o(\text{unliganded})$ difference map contoured at $+4.5 \rho_{\text{rms}}$ (green mesh) and $-4.5 \rho_{\text{rms}}$ (red mesh). Atom coloring and residue labeling are as described in Fig. 3. (B) Model for azide-bound diferric metallocenter (seen at 0.5 occupancy) including the newly bound HOH^{az} . Lines and coloring are as in Fig. 3a.

Figure 5.

Stereoview of the diferrous symerythrin metallocenter and its comparison with rubrerythrin. (A) The reduced symerythrin metallocenter modeled as two equally populated conformations, **A** and **B**, each at 50% occupancy. The $2F_o - F_c$ composite annealed omit map contoured at $2.0 \rho_{rms}$ (blue mesh) reveals single conformations for residues $E_{\pi A}$ and $E_{\pi C}$ and two conformations for $E_{\alpha A}$ and density contoured to $5.0 \rho_{rms}$ shows two positions for Fe1 (Fe1**A** and Fe1**B**). An $F_o - F_c$ omit map (omitting HOH-A) contoured to $+4.0 \rho_{rms}$ (green mesh) shows evidence for modeling an Fe1**A**-ligating water molecule (labeled HOH-A). Lines and atom coloring are as in Fig. 3a. (B) The reduced conformation **A** including the omit map density for HOH-A described in panel A. (C) Overlay of conformation **A** of diferrous symerythrin with the diferrous metallocenter of rubrerythrin (transparent atoms with same coloring as Fig. 3b; PDB 1lko).

Figure 6.

Internal symmetry in diferrous symerythrin. (A) Overlay of helix pair A–B (green) onto C–D (blue) highlighting iron-ligating residues (sticks), the two π -helices (orange) and the two iron atoms (brown and light brown spheres). (B) Orthogonal views (direction of rotation indicated) of an overlay of conformation A of the diferrous symerythrin metallocenter (coloring is the same as Fig. 3) with itself (yellow carbon atoms), based on the superimposing of helix pairs A–B and C–D onto helix pairs C–D and A–B, respectively. Direct metal ligands and the second-shell Tyr residues are shown and labeled in the panel in which they are best seen.

Figure 7.

Oxidase and peroxidase cycling of *C. paradoxa* symerythrin. (A) Absorption spectra of diferrous non-crosslinked (gray lines) and crosslinked (black lines) symerythrin before (solid lines) and after (dashed lines) exposure to O_2 . Dotted arrow emphasizes the direction of the spectral changes. (B) As for A but showing spectra before and after exposure to hydrogen peroxide. Inset: stoichiometry of the change in absorption coefficient at 340 nm as a function of hydrogen peroxide added to 50 μM diferrous, crosslinked symerythrin. Equivalent results were obtained for the non-crosslinked isoform (Supplemental Fig. 3b).

Figure 8.

Internal sequence similarities of three FLSF families and a model for their origins. (A) Internal alignment of consensus sequences from helix pairs A–B and C–D of symerythrins, of the rubrerythrin family members having seven metal ligating residues, of erythrins, and of the oligomerizing α -peptide family. Upper- and lower-case letters give the identities of highly and moderately conserved residues, respectively, while "x" designates positions for which there is lesser sequence conservation (see *Material and Methods*). Conserved metallocenter residues (including the secondary sphere Tyr residues) are highlighted in red and are connected across families with solid red lines. Residues that are internally symmetric in symerythrins, 7-ligand rubrerythrins, and erythrins are highlighted in gray, green and purple, respectively. A dark highlight coloring is used for positions that match highly conserved residues and a light highlight color is used for positions matching a moderately conserved residue with either a moderately conserved or highly conserved residue. The complete SeqLogos plot²⁹ of these consensus sequences are shown in Supplemental Fig. 5. Dashed black lines connect residue pairs that match between the groups. In parenthesis below the family name is the number of sequences used to generate the consensus sequences for that group. (B) Overlay of helix pairs A–B and C–D of symerythrin (dark green and dark blue, respectively) on the equivalent helices of erythrin (light green and light blue, respectively; PDB 2fzf) shows the structural difference associated with the three residue gap in erythrins compared to rubrerythrins. Side chains of metallocenter residues are shown as sticks. (C) Schematic for the origins of modern day symerythrins, 7-ligand rubrerythrins, erythrins and the oligomerizing α -peptides. Solid lines represent a simple direct pathway for the formation of each modern day family assuming that the dimerizing two-helix peptides are all homologous with each other. The dotted lines (with "?") are alternatives to the thinner solid lines (with "?") recognizing that with the small number of residues involved in the metallocenter, it is also possible that the distinct dimerizing peptides were not homologous.

Table 1

Data Collection and Refinement Statistics for Symerythrin

	Diferric ^a	Dithionite-Treated	Azide-bound diferric
<i>Data collection ^b</i>			
Space group	P3 ₂	P3 ₂	P3 ₂
Unit cell axes (Å)	a=b=81.47, c=46.26	a=b=81.50 c=46.34	a=b=82.01 c=46.44
Resolution Limits (Å)	30.57-1.20 (1.25-1.20)	30.60-1.25 (1.31-1.25)	30.74-1.40 (1.48-1.40)
Unique Observations	107,444 (15,684)	95,243 (13,875)	68,857 (10,086)
Multiplicity	17.8 (10.7)	11.9 (10.4)	12.2 (12.0)
Average I/σ	23.4 (6.3)	15.0 (5.4)	18.9 (5.6)
R _{meas} ^c (%)	8.1 (35.2)	10.6 (42.6)	7.7 (46.1)
<i>Refinement</i>			
R _{cryst} /R _{free} (%)	10.1/12.5	11.5/14.1	11.5/15.6
No. protein molecules	2	2	2
No. protein residues	358	358	352
No. water molecules	555	555	562
Total number atoms	3534	3549	3543
rmsd bond angles (°)	3.6	2.8	2.1
rmsd bond lengths (Å)	0.019	0.022	0.026
 protein (Å ²)	12.9	14.7	16.2
 water (Å ²)	26.2	27.3	30.24
 Fe (Å ²)	6.73	10.1	10.0
<i>Ramachandran Plot (%) ^d</i>			
Favored	97.2	97.2	98.0
Outliers	0.0	0.0	0.0
PDB code	3qhb	3qhc	3sid

^aThe statistics for this structure are as reported in Cooley et al. (2011)^bNumbers in parentheses correspond to values in the highest resolution bin^cR_{meas} is the multiplicity-weighted merging R-factor ³⁸^dRamachandran plot generated using Molprobity ³⁶

Table 2

Internal symmetry statistics for structure-based comparisons of the A–B with the C–D core-helix pairs from representatives of FLSF proteins. Alignments and the reported statistics were generated by the DALI server.¹⁹

Protein	PDB	No. aligned residues	% Sequence Identity	RMSD (Å)	Z-score
Symerythrin	3qhb	57	32	0.77	10.6
Ruberythrin	1lkm	58	29	1.5	9.8
Erythrin	2fzf	58	31	1.1	8.5
DPS	2d5k	59	7	2.0	8.9
Bacterioferritin	1bfr	62	16	1.3	9.3
Ferritin	1vlg	62	11	1.5	8.4
Δ^9 -desaturase	2uw1	55	4	2.5	4.3
RNR Ib	1uzr	59	15	2.8	6.4
MMOH α -subunit	1mtv:D	60	8	2.8	5.8
MMOH β -subunit	1mtv:B	57	9	2.7	4.6