PROTEINS: Structure, Function, and Bioinformatics 57:878–882 (2004)

# STRUCTURE NOTE

# Structural Genomics of *Pyrococcus furiosus*: X-Ray Crystallography Reveals 3D Domain Swapping in Rubrerythrin

**Wolfram Tempel,[1] Zhi-Jie (James) Liu,[1] Florian D. Schubot,[1] Ashit Shah,[1] Michael V. Weinberg,[1] Francis E. Jenney, Jr.,[1] W. Bryan Arendall, III,[2] Michael W. W. Adams,[1] Jane S. Richardson,[2] David C. Richardson,[2] John P. Rose,[1*] and Bi-Cheng Wang[1]**

[1]*Southeast Collaboratory for Structural Genomics, Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia*

[2]*Department of Biochemistry, Duke University, Durham, North Carolina*

***Introduction.*** Rubrerythrin was first isolated from the anaerobic bacterium *Desulfovibrio vulgaris* [Protein Data Bank (PDB) entry: 1RYT].[1,2] It is a homodimeric protein and each subunit contains two different types of metal sites. One is mononuclear and located in the C-terminal domain. It is rubredoxin-like and consists of 1 metal ion bound to the side-chains of 4 cysteinyl residues. The second site, in the N-terminal domain, is binuclear. It is hemerythrin-like and consists of 2 metal ions bound by the side-chains of 5 glutamyl and 1 histidinyl residue. The nature of the metal ions in the 2 sites of *D. vulgaris* rubrerythrin is not clear, since both $3Fe$[3,4] and 1Zn/2Fe forms[5] of the protein have been isolated. The mononuclear site appears to be occupied only by Fe in both forms.[5] A variety of catalytic activities have also been proposed for *D. vulgaris* rubrerythrin, including pyrophosphatase,[3] ferroxidase,[6] and nicotinamide adenine dinucleotide hydride (NADH) peroxidase.[7–9] Its physiological function has also not been established, although a role in oxidative stress has been implicated for the *D. vulgaris* protein, as well as for those from several other anaerobic and microaerophilic bacteria.[10]

The nature and function of rubrerythrin in anaerobic archaea was recently clarified by the purification and biochemical characterization of the protein from the hyperthermophile, *Pyrococcus furiosus*.[11] It was shown that the native protein purified from *P. furiosus* contains only iron, although the recombinant protein, obtained from *Escherichia coli,* contained one zinc atom/subunit even when the organism was grown on a minimal medium supplemented with iron. In addition, native (but not the zinc-containing recombinant forms of) *P. furiosus* rubrerythrin had peroxidase activity and reduced hydrogen peroxide to water using *P. furiosus* rubredoxin as the electron donor. An NADH-dependent peroxidase system was reconstituted using rubrerythrin, rubredoxin, and NADH rubredoxin oxidoreductase (from *P. furiosus*). This system is proposed to be part of a pathway for the detoxification of reactive oxygen species involving the novel enzyme superoxide reductase.[12]

A structural study of rubrerythrin from *P. furiosus* was initiated as part of the structural genomics intiative that is under way with this hyperthermophilic archaeon.[13] The sequence of *P. furiosus* rubrerythrin shows 32% sequence identity with that of the *D. vulgaris* protein.[14] A direct alignment is shown in Figure 1 and reveals a 13-residue deletion in the *P. furiosus* sequence but very similar secondary structural motifs,[16] suggesting virtually identical structures. However, we show here that the structure of *P. furiosus* rubrerythrin, determined to a resolution of 2.35 Å, reveals a previously unobserved swapping of its structural domains.

***Experimental.*** *Expression and purification:* The gene encoding *P. furiosus* rubrerythrin (PF1283) was cloned and expressed in *E. coli* and the recombinant protein was purified according to the high-throughput protocols established for *P. furiosus* protein production at the Southeast Collaboratory for Structural Genomics (SECSG).[13] The resultant protein contained an N-terminal $His_6$ tag. Protein identity and purity were assessed using mass spectroscopy and polyacrylamide gel electrophoresis (PAGE).

*Crystallization:* Crystallization experiments were performed using the modified microbatch under oil method.[17]

Grant sponsor: National Institutes of Health; Grant numbers: GM62407 and 60329. Grant sponsor: IBM Life Sciences. Grant sponsor: Georgia Research Alliance and the University of Georgia Research Foundation. Grant sponsor: Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences; Contract number: W-31-109-ENG-38.

*Correspondence to: John P. Rose, Southeast Collaboratory for Structural Genomics, Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602. E-mail: rose@bcl4.bmb.uga.edu

Received 19 May 2004; Accepted 30 June 2004

Published online 1 October 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20280

```
      HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH HHHHHHHHHHHHHHHHHHHHH
1NNQ: -MVVKRTMTKKFLEEAFAGESMAHMRYLIFAEKAEQEGFPNIAKLFRAIAYAEFVHAKNH
         :* : *:* :  ****** *: **  *. :*:::** :*:.:*   *  *  ***.
1RYT: MKSLKGSRTEKNILTAFAGESQARNRYNYFGGQAKKDGFVQISDIFAETADQEREHAKRL
        HHHHHHHHHHHHHHHHHHHHHHHHHHHHHH  HHHHHHHHHHHHHHHHHHHH

      HHHH                  HHHHHHHHHHHHHHHHHHHHHHHHHHHHHH HHHHHHH
1NNQ: FIAL-------------GKLGKTPENLQMGIEGETFEVEEMYPVYNKAAEFQGEKEAVRT
      *  *               * :..*  **  .  ** .*  **** : : *. :* :* .*.
1RYT: FKFLEGGDLEIVAAFPAGIIADTHANLIASAAGEHHEYTEMYPSFARIAREEGYEEIARV
      HHH                   HHHHHHHHHHHHHHHHHH HHHHHHHHHHH   HHHHHH

      HHHHHHHHHHHHHHHHHHHHHHHHH          EEE      EEE            H
1NNQ: THYALEAEKIHAELYRKAKEKAEKG--EDIEIKKVYICPICGYTAVDE-APEYCPVCGAP
           **::* : : .  .: ::*     *  . : *  ***.  .  *** **.*. *
1RYT: FASIAVAEEFHEKRFLDFARNIKEGRVFLREQATKWRCRNCGYVHEGTGAPELCPACAHP
      HHHHHHHHHHHHHHHHHHHHHHHHH          EEEEE     EEEE

      HHHHEEE
1NNQ: KEKFVVFE---
      * :* ::
1RYT: KAHFELLGINW
      HHHEEE
```

Fig. 1. Sequence alignment with CLUSTALW[15] and secondary structure for PDB entries 1NNQ (*P. furiosus*, red/top) and 1RYT (*D. vulgaris*, blue/bottom).

**TABLE I. Data Collection Parameters and Statistics**

| | |
|---|---|
| Wavelength (Å) | 0.97 |
| Resolution (outer shell, Å) | 50-2.35 (2.43–2.35) |
| $\langle I \rangle / \langle \sigma(I) \rangle$ | 36.7 (6.3) |
| Completeness (%) | 99.9 (100.0) |
| Observations | 311097 |
| $R_{sym} = \Sigma(ABS(I-\langle I \rangle))/\Sigma(I)$ | 7.1 (36.9) |
| Unique HKLs data ($R_{free}$) | 18575 (1004) |
| Number of atoms in model | 2744 |
| Rfactor ($R_{free}$, %) | 21.1 (25.3) |
| RMSD bond lengths (Å)/angles (°) | 0.013/1.23 |

For these experiments, 0.5 μL of crystallization reagent were mixed with 0.5 μL of protein (20 mg/mL) solution, 10 m*M* NaCl, 2 m*M* dithiothreitol (DTT), and 20 m*M* Tris, pH 8. The wells were sealed with a 70:30 paraffin:silicone oil mixture to retard dehydration. Initial crystallization conditions were identified using the Wizard-1 screening kit (Emerald Biostructures, EBS-WIZ-1) and optimized. Diffraction-quality crystals were obtained from a 1.2 μL drop containing 0.5 μL of the above protein solution, 0.5 μL precipitant solution, 20% w/v polyethylene glycol (PEG-1000), 200 m*M* calcium acetate, 100 m*M* imidazole, pH 8,

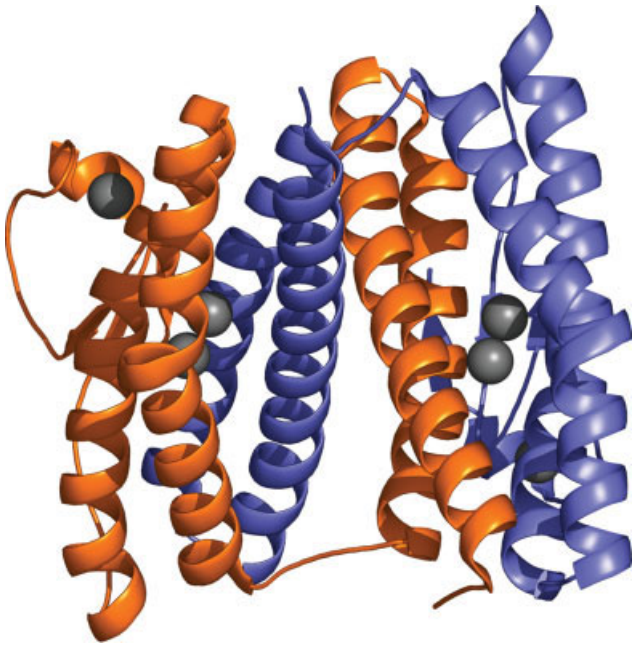

Fig. 2. Cartoon representation of the *P. furiosus* homodimer generated with PyMOL.[26] Metal cations are shown as gray spheres.

and 0.2 μL of additive solution, and 25 m*M* $K_2PtCl_4$ (Hampton Research, HR2-422).

*Data Collection:* A single crystal was harvested with a cryoloop (Hampton Research, HR4-747) and briefly immersed in a 1 μL drop containing a 1:4 mixture of glycerol and the above-described precipitant solution prior to incubation (10 min) under 100 PSI xenon (Rigaku/MSC CryoXeSiter), before being flash cooled in liquid Freon, retrieved, and stored in liquid nitrogen.

Data that lead to the structure solution were collected at cryogenic temperatures on beamline 22ID (SER-CAT), Advanced Photon Source, Argonne National Laboratory using a MAR 165 charge-coupled device (CCD) detector and 0.97 Å X-rays. The detector was positioned 170 mm from the crystal at a 2θ angle of 0°, providing data to the 2.35 Å resolution, the limit of the detector. Two data sets consisting of 200 images each (0.5° ω steps) were collected with ϕ offsets of 0° and 180°, respectively. The exposure time was 2 s. Data were indexed, integrated, and scaled using the HKL2000 software suite;[18] see Table I.

The positions of 2 anomalous scatterers and initial phases based on these sites were obtained with SOLVE V2.02[19] using the single-wavelength anomalous scattering option. Phase refinement and initial tracing of the peptide chains were accomplished using RESOLVE.[20] The model was adjusted manually using XFIT,[21] refined using REFMAC5,[22] and validated using MOLPROBITY[23] and PROCHECK.[24] The refined model is available from the PDB,[25] entry 1NNQ.

***Results and Discussion.*** A crystal of recombinant rubrerythrin was incubated under high pressure (100 psi) xenon, flash-cooled in liquid Freon, and shipped to SER-CAT for data collection. A set of high-resolution (2.35 Å) data was collected on the xenon-incubated crystal using the SER-CAT undulator beamline and 0.97 Å X-rays for refinement as described. The SER-CAT data set was also submitted to the SECSG sca2structure high-throughput structure determination pipeline. The pipeline spawned 180 SOLVE/RESOLVE jobs, on the SECSG Linux cluster. Parameters surveyed include space group ($P4_12_12$, $P4_22_12$, and $P4_32_12$), high-resolution data cutoffs for both the initial phasing (4.0–2.4 Å in 0.4-Å increments) and phase refinement, and extension (also screened from 4.0 to 2.4 Å in 0.4-Å increments, but always at the same or higher resolution cutoff than used in initial phasing), the number of anomalous scatterers within the asymmetric unit (2 and 4), and the number of molecules in the asymmetric unit (2 or 3). All anomalous scatterers were assumed to be xenon for the calculations. The total phasing and fitting process (180 parallel jobs) took approximately 3.5 h on a 128 processor IBM Linux cluster with the initial fitted map being produced in under 4.5 h from mounting the crystal on the goniometer.

The SCA2STRUCTURE results gave two sites from the SOLVE analysis and a fitted sequence (80% complete) from automated RESOLVE fitting. The inspection of the RESOLVE maps and the model produced quickly revealed that the anomalous scatterers identified by SOLVE were in fact metal ions located in the protein's iron-binding sites. Whether the unsuccessful attempts at derivatization with $K_2PtCl_4$ or xenon were in fact relevant to the successful structure determination has not been revisited. The experimental details for these procedures are listed here for completeness only.

**TABLE II. Comparison of the *P. furiosus* and *D. vulgaris* 4-Helix Bundles[a]**

| | *Pyrococcus furiosus* | | *Desulfovibrio vulgaris* | |
|---|---|---|---|---|
| Helix | Residues | Chain | Chain | Residues |
| HA | 5–36 | A | A | 8–37 |
| HB | 38–63 | A | A | 40–63 |
| HC | 68–98 | B | A | 83–111 |
| HD | 100–130 | B | A | 115–143 |

[a]PDB entry 1RYT contains coordinates for a single peptide chain.

*Overall structure and domain swapping:* The *P. furiosus* crystals contain 1 rubrerythrin homodimer in the asymmetric unit. The rubrerythrin monomer (cartoon of homodimer shown in Fig. 2) is mostly helical, consisting of 2 helical domains, each containing a 2-helix bundle (HA-HB and HC-HD; see Table II), separated by a short 4-residue linker. A 7-residue linker connects helix HD with a small C-terminal domain, comprising residues 138 through 171, that contains a short, 3-strand β-sheet.

During examination of initial electron density maps, discrepancies between the initial *P. furiosus* trace and the topology of *D. vulgaris* structure (1RYT) became apparent. While the topologies of both structures are conserved to residue 62 (the end of helix HB), the structures are distinct as to how the peptide chain connects to the next helix (HC). This is due to the 13-residue deletion observed in the *P. furiosus* sequence (Fig. 1) that corresponds to the loop linking HB and HC in the *D. vulgaris* structure. Because of this deletion, the relationship of the 2 helical domains in the monomer is markedly different in the *P. furiosus* structure compared with *D. vulgaris* structure and affects the overall dimer structure (Fig. 3). The *D. vulgaris* dimer can be described as a simple 2-fold (crystallographic) related dimer (HD↓-HC↑-HB↑-HA↓ · HA′↑-HB′↓-HC′↓-HD′↑). In the *P. furiosus* case, domain swapping occurs between the 2-rubrerythrin monomers, creating a very similar (HD′↓-HC′↑-HB&uaarr;-HA↓ · HA′↑-HB′↓-HC↓-HD↑) dimer. Because of this, the relationship of the two 4-helix bundles comprising the rubrerythrin dimer remains essentially unchanged [root-mean-square deviation (RMSD) Cα 1.5 Å] in the *P. furiosus* and *D. vulgaris* structures (Fig. 4).

During the preparation of this manuscript, three-dimensional (3D) domain swapping in the rubrerythrin-like protein sulerythrin from the aerobic archeon *Sulfolobus tokodaii* was reported.[29] The report included a prediction that "rubredoxin-like proteins from *Pyrococcus* species" would not exhibit domain swapping. The authors of the sulerythrin structure emphasize the "as isolated" in vivo character of their model. In this context, it is noteworthy that the *P. furiosus* structure discussed here is the result of recombinant expression. Whether 3D domain swapping in this case is an artifact[27] of *E. coli* expression will be conclusively addressed with the structure determination of native *P. furiosus* rubrerythrin. However, a

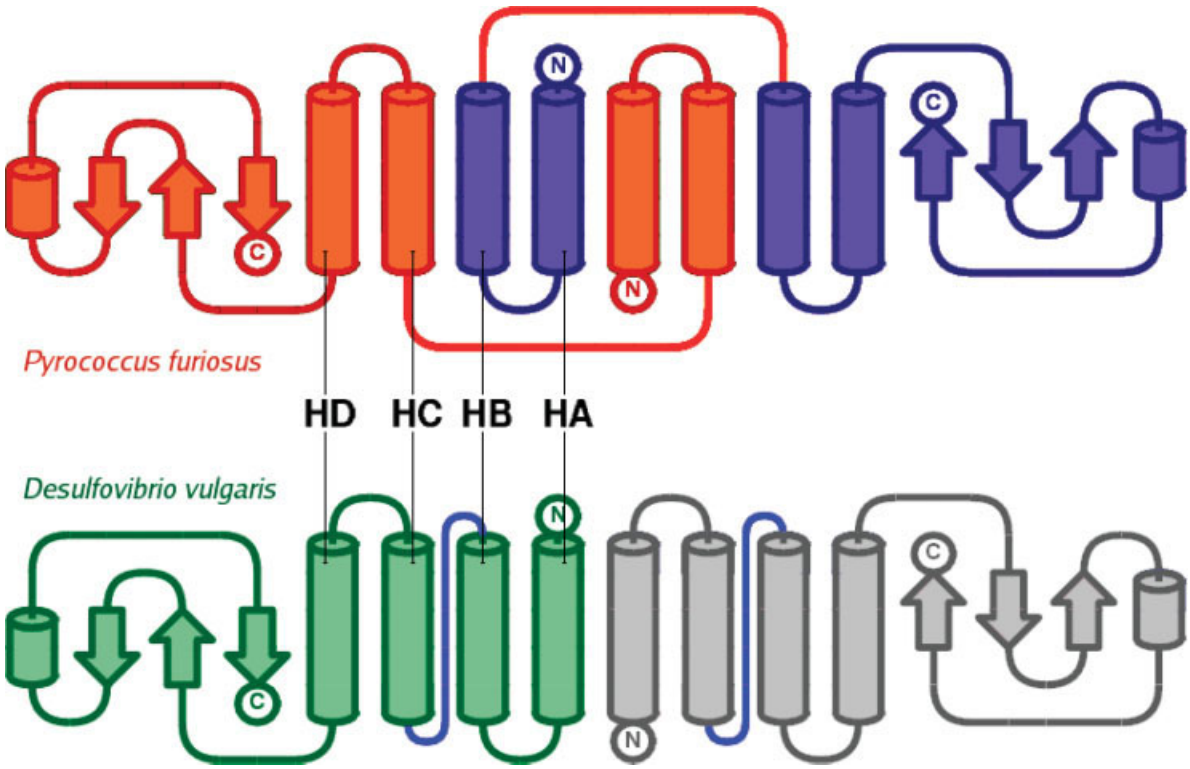


Fig. 3. 3D domain swapping[27] shown in topology diagrams of PDB entries 1NNQ (*P. furiosus*, top) and 1RYT (*D. vulgaris*, bottom). Colors orange, purple, green, and gray were applied to distinguish the monomers. The linker segments relevant to domain swapping are colored red (1NNQ) and blue (1RYT). The diagram was generated with the program TOPDRAW.[28]

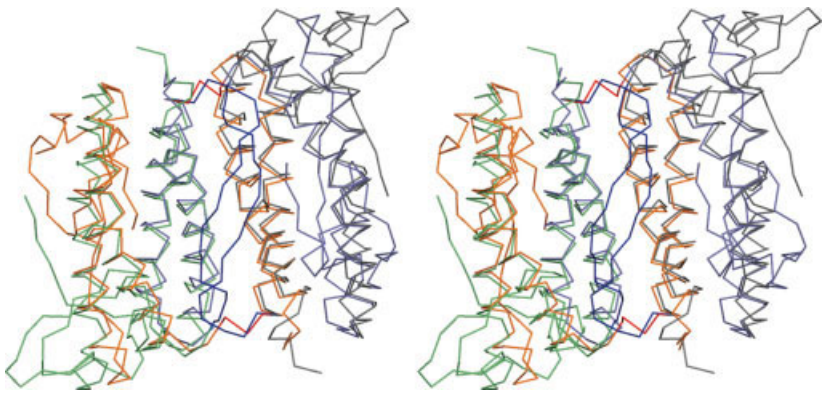

Fig. 4. Stereographic representation of the superimposition of helix bundles in PDB entries 1NNQ (*P. furiosus*) and 1RYT (*D. vulgaris*). The monomers are colored in orange/purple for 1NNQ and green/gray for 1RYT, respectively. This image was created using PyMOL.[26]

nonswapped assembly conserving the typical helix bundle is unlikely due to the very short loop connecting helices HB and HC.

*Metal-binding sites:* Each peptide chain accommodates 3 metal ions in 2 metal-binding centers. The coiled C-termini of each subunit accommodate a single metal ion in a rubredoxin-like $CX_2CX_nCPXC$ center. As a consequence of domain swapping[27] in the structure of the *P. furiosus* protein, the glutamyl and histidinyl side-chains that form the binuclear metal center are contributed by residues from both rubrerythrin monomers.

For refinement, all metal ions found in the metal-binding sites were assigned as divalent zinc, since inductively coupled plasma spectroscopy detected significant amounts of zinc in the crystallization sample (data not shown[11]). The presence of zinc is also supported by Bijvoet difference Fourier analysis using data collected with 0.97 Å radiation. The peaks corresponding to the 3 metal sites were stronger for data collected using 0.97 Å X-rays than for data collected with 1.54 Å X-rays (data not shown). This would be expected for zinc-containing crystals, since zinc ($\Delta f''_{Zn} = 2.4$) should show anomalous signal from the $\lambda = 0.97$ Å data and not from the $\lambda = 1.54$ Å data ($\Delta f''_{Zn} = 0.7$). In the case of iron, one would expect a stronger signal from

the $\lambda = 1.54$ Å data ($\Delta f''_{Fe} = 3.2$) than from the $\lambda = 0.97$ Å data ($\Delta f''_{Fe} = 1.5$), which runs counter to what was observed. The resolution of the structure prevents identification based on binding geometry.

***Acknowledgments.*** Data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members.html.

## REFERENCES

1. Legall J, Prickril BC, Moura I, Xavier AV, Moura JJG, Huynh BH. Isolation and characterization of rubrerythrin, a non-heme iron protein from *Desulfovibrio vulgaris* that contains rubredoxin centers and a hemerythrin-like binuclear iron cluster. Biochemistry 1988;27:1636–1642.
2. deMare F, Kurtz DM, Nordlund P. The structure of *Desulfovibrio vulgaris* rubrerythrin reveals a unique combination of rubredoxin-like FeS4 and ferritin-like diiron domains. Nat Struct Biol 1996;3: 539–546.
3. Liu MY, Legall J. Purification and characterization of 2 proteins with inorganic pyrophosphatase activity from *Desulfovibrio vulgaris* rubrerythrin and a new, highly-active, enzyme. Biochem Biophys Res Commun 1990;171:313–318.
4. Ravi N, Prickril BC, Kurtz DM, Huynh BH. Spectroscopic characterization of Fe-57-reconstituted rubrerythrin, a nonheme iron protein with structural analogies to ribonucleotide reductase. Biochemistry 1993;32:8487–8491.
5. Sieker LC, Holmes M, Le Trong I, Turley S, Liu MY, LeGall J, Stenkamp RE. The 1.9 angstrom crystal structure of the "as isolated" rubrerythrin from *Desulfovibrio vulgaris*: some surprising results. J Biol Inorg Chem 2000;5:505–513.
6. Bonomi F, Kurtz DM, Cui XY. Ferroxidase activity of recombinant *Desulfovibrio vulgaris* rubrerythrin. J Biol Inorg Chem 1996;1:67–72.
7. Coulter ED, Kurtz DM. A role for rubredoxin in oxidative stress protection in *Desulfovibrio vulgaris*: Catalytic electron transfer to rubrerythrin and two-iron superoxide reductase. Arch Biochem Biophys 2001;394:76–86.
8. Coulter ED, Shenvi NV, Kurtz DM. In vitro and in vivo activities of rubrerythrin. J Inorg Biochem 1999;74:197.
9. Coulter ED, Shenvi NV, Beharry ZM, Smith JJ, Prickril BC, Kurtz DM. Rubrerythrin-catalyzed substrate oxidation by dioxygen and hydrogen peroxide. Inorg Chim Acta 2000;297:231–241.
10. Alban PS, Krieg NR. A hydrogen peroxide resistant mutant of *Spirillum volutans* has NADH peroxidase activity but no increased oxygen tolerance. Can J Microbiol 1998;44:87–91.
11. Weinberg MV, Jenney FE, Jr., Cui X, Adams MWW. Rubrerythrin from the hyperthermophilic archaeon *Pyrococcus furiosus* is a rubredoxin-dependent, iron-containing peroxidase. J Bacteriol 2004. In press.
12. Jenney FE, Verhagen MFJM, Cui XY, Adams MWW. Anaerobic microbes: oxygen detoxification without superoxide dismutase. Science 1999;286:306–309.
13. Adams MW, Dailey HA, DeLucas LJ, Luo M, Prestegard JH, Rose JP, Wang BC. The Southeast Collaboratory for Structural Genomics: a high-throughput gene to structure factory. Acc Chem Res 2003;36:191–198.
14. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–410.
15. Thompson JD, Higgins DG, Gibson TJ. Clustal-W—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.
16. Rost B. PHD: predicting one-dimensional protein structure by profile-based neural networks. Computer Methods for Macromolecular Sequence Analysis. Methods Enzymol 1996;266:525–539.
17. Baldock P, Mills V, Stewart PS. A comparison of microbatch and vapour diffusion for initial screening of crystallization conditions. J Cryst Growth 1996;168:170–174.
18. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
19. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr 1999;55:849–861.
20. Terwilliger TC. Automated structure solution, density modification and model building. Acta Crystallogr D Biol Crystallogr 2002;58:1937–1940.
21. McRee DE. XtalView Xfit—a versatile program for manipulating atomic coordinates and electron density. J Struct Biol 1999;125: 156–165.
22. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 1997;53:240–255.
23. Lovell SC, Davis IW, Adrendall WB, Bakker PIWd, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by C alpha geometry: phi,psi and C beta deviation. Proteins 2003;50:437–450.
24. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. Procheck—a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
25. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucl Acids Res 2000;28:235–242.
26. DeLano WL. The PyMOL molecular graphics system. San Carlos, CA: DeLano Scientific; 2002.
27. Liu Y, Eisenberg D. 3D domain swapping: as domains continue to swap. Protein Sci 2002;11:1285–1299.
28. Bond CS. TopDraw: a sketchpad for protein structure topology cartoons. Bioinformatics 2003;19:311–312.
29. Fushinobu S, Shoun H, Wakagi T. Crystal structure of sulerythrin, a rubrerythrin-like protein from a strictly aerobic archaeon, *Sulfolobus tokodaii* strain 7, shows unexpected domain swapping. Biochemistry 2003;42:11707–11715.