The cyanobacterial protoporphyrinogen oxidase HemJ is a new *b*-type heme protein functionally coupled with coproporphyrinogen III oxidase

Petra Skotnicová, Roman Sobotka, Mark Shepherd, Jan Hájek, Pavel Hrouzek and Martin Tichý

SUPPORTING INFORMATION

Supplementary results

HemJ modeling

The coevolutionary protein modeling is based on the assumption that during evolution, if one amino acid changes, then another amino acid changes as well to keep the structure and activity of the protein intact (3). By identifying such pairs of coevolving amino acids, it is possible to predict which residues are close in the three-dimensional structure of the protein. This information can be used to generate a structural model of a protein. The accuracy of such prediction is dependent on the number of homologs in the multiple sequence alignment.

Cyanobacterial HemJ proteins contain 5 transmembrane helices unlike most HemJ proteins containing only 4 helices. The modeling was performed on HemJ peptide from *R. sphaeroides* (WP_023003745) with 4 helices as the cyanobacterial 5^{th} helix did not provide enough coverage for coevolutionary modeling (see below). From automatic structure prediction servers, Robetta employs both ab initio folding and template-based modeling, RaptorX employs template-based modeling and RaptorX-Contact is based on ab initio modeling. Interestingly, all the methods provided a very similar fold – a four helix bundle unit (Fig. S8). Robetta models were based on the PF03653 coevolutionary data model (1). Interestingly, the three best templates chosen by RaptorX for modeling were heme-binding subunits of membrane redox complexes. These templates were subunits of cytochrome ba_3 oxidase (PDB ID 3eh3:A), cytochrome c oxidase (PDB ID 1m56:A) and the best template - cytochrome *bd*-type oxidase (PDB ID 5doq:A). Cytochrome *bd*-I ubiquinol oxidase is a terminal oxidase that generates a proton-motive force across the plasmatic membrane and mediates transfer of electrons from quinol to oxygen. The enzyme has two integral membrane subunits CydA and CydB (2) with three hemes (heme b_{595}, heme b_{558}, and heme *d*) mediating electron transfer. The resulting RaptorX model of HemJ aligned with both CydA and CydB as they possess the same fold (3) forming heterodimer. Modeled HemJ also revealed strong similarities to another four helix bundle of cytb_{561} (PDB ID: 4O7G). Strikingly, in our HemJ model, invariable His16 (His12 in the Fig. S7) is aligned with conserved His21 in CydA and with His51 in cyt b_{561} (Fig. S8). Both these histidines provide axial ligands for heme *b* (4,5).

Table S1. Mass spectrometry identification of protein bands detected in the HemJ.f eluate after separation by SDS-PAGE (see Fig. 4A)

Protein bands were identified by nano LC-MS/MS and by database search

Gene name	Protein description	Mass (Da)	Theoretical Peptides	Peptides found	Coverage (%)
slr0228	cell division protein FtsH2	68453	48	18	37.1611
sll1463	cell division protein FtsH4	68157	60	14	32.0064
slr1390	cell division protein FtsH1	69261	49	11	26.4798
slr1604	cell division protein FtsH3	67209	45	22	46.1039
slr1834	P700 apoprotein subunit Ia	82882	34	4	6.2583
slr1835	P700 apoprotein subunit Ib	81240	24	4	7.1135
slr0228	cell division protein FtsH2	68453	48	2	2.8708
slr1329	ATP synthase beta subunit	51700	31	11	34.7826
sll1326	ATP synthase alpha chain	53932	33	8	23.2604
slr1790	Protoporphyrinogen oxidase HemJ	22049	14	3	20.7254
slr1790	Protoporphyrinogen oxidase HemJ	22049	14	7	29.0155
sll1106	protein with unknown function	18019	10	8	46.7836
slr1790	Protoporphyrinogen oxidase HemJ	22049	14	1	6.6667

Table S2. Main fragments obtained in HRMS/MS analysis of Harderoporphyrin and Copro

[M+H]+ corresponds to the molecular ion charged by proton.

Compound (Elemetal composition)	**Ion (*m/z*)**	**Formula**	**Error (ppm)**	**Loss designation**
Harderoporphyrin	609.2730	$C_{35}H_{37}N_4O_6$	-2.8	[M+H]+
($C_{35}H_{37}N_4O_6$)	550.2622	$C_{33}H_{34}N_4O_4$	-8.6	·CH_2COOH radical loss
	491.2466	$C_{31}H_{31}N_4O_2$	-5.1	·CH_2COOH radical loss
	432.2313	$C_{29}H_{28}N_4$	-1.0	·CH_2COOH radical loss
Coproporphyrin	655.2784	$C_{36}H_{39}N_4O_8$	-3.3	[M+H]+
($C_{36}H_{39}N_4O_8$)	596.2649	$C_{34}H_{36}N_4O_6$	-3.2	·CH_2COOH radical loss
	537.2517	$C_{32}H_{33}N_4O_4$	-3.8	·CH_2COOH radical loss
	477.2304	$C_{30}H_{29}N_4O_2$	-3.9	·CH_2COOH radical loss
	419.2232	$C_{28}H_{27}N_4$	-0.4	·CH_2COOH radical loss

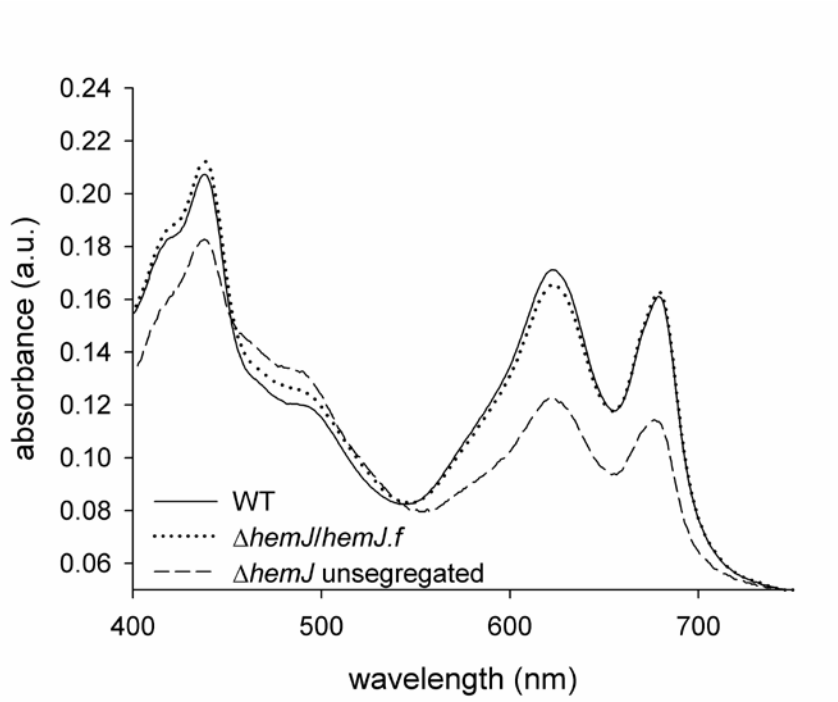

Figure S1. Whole cell absorption spectra suggesting complementation of Δ*hemJ* by FLAG-tagged *hemJ*

A whole cell absorption spectra of the *Synechocystis* 6803 WT, Δ*hemJ*/*hemJ.f* and unsegregated Δ*hemJ* (supplemented with 20 µg/ml of chloramphenicol) grown at 40 µmol of photons $m^{-2}s^{-1}$. Chlorophyll is represented by 680 nm peak and phycobiliproteins by the 625 nm peak. Spectra were normalised to light scattering at 750 nm.

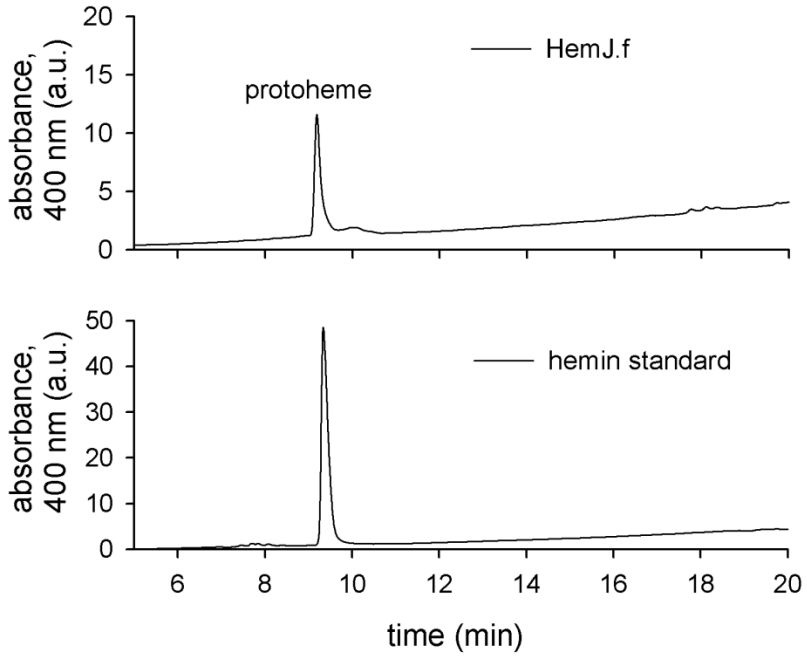

Figure S2. Determination of protoheme in the purified HemJ.f protein

Heme was extracted by acetone/2 % HCl from purified HemJ.f from ΔPSI background and separated by HPLC essentially as described in (6) (upper chromatogram). The lower chromatogram shows separation of a hemin standard (Sigma, Germany).

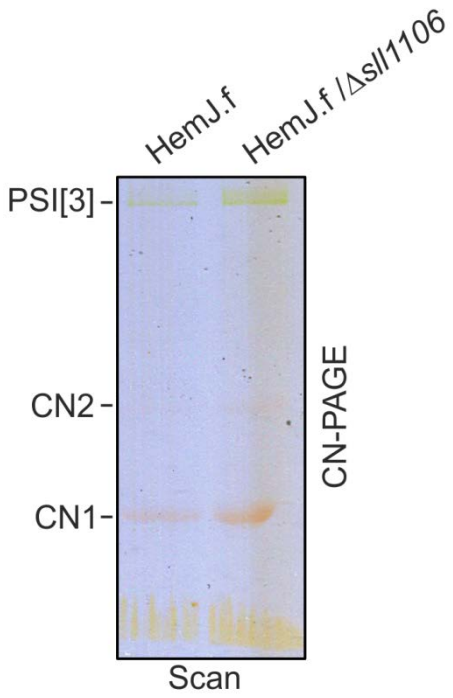

Figure S3. Separation of the purified HemJ.f obtained from Δ*sll1106* mutant of *Synechocystis* 6803 by CN-PAGE

The HemJ pulldown obtained from Δ*sll1106* mutant of *Synechocystis* 6803 was separated by 4-14 % clear native gel electrophoresis (CN-PAGE). The gel was scanned in transmittance mode (Scan) using an LAS 4000 Imager (Fuji).

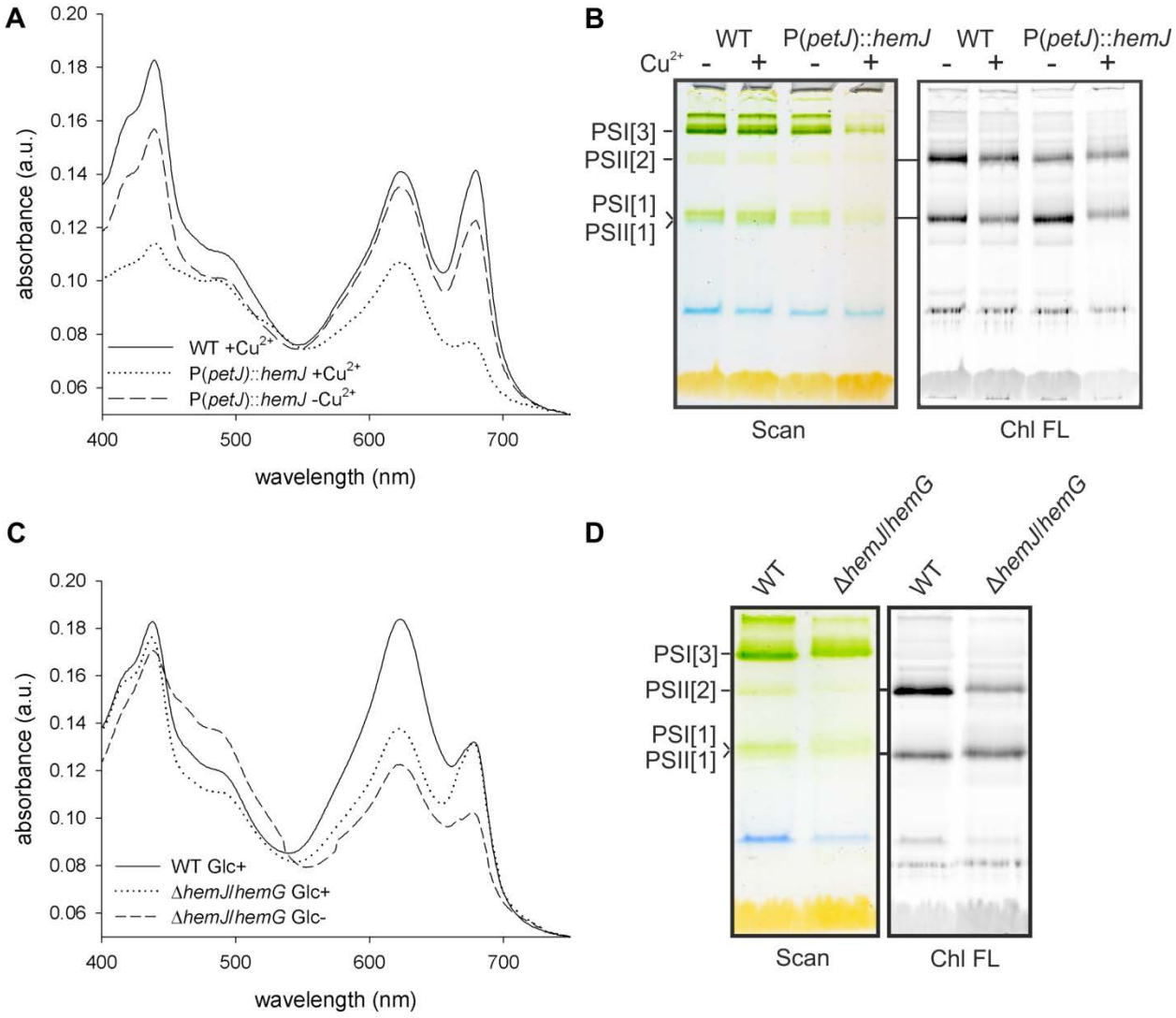

Figure S4. Whole cell absorption spectra of P(*petJ*)::*hemJ* and Δ*hemJ*/*hemG* strains and CN-PAGE of solubilized membrane complexes

Synechocystis 6803 strains grown at 40 µmol of photons $m^{-2}s^{-1}$ were used for whole cell absorption spectra (A, C) and CN-PAGE (B, D). Chlorophyll is represented by 680 nm peak and phycobilinoproteins by the 625 nm peak. Spectra were normalised to light scattering at 750 nm. P(*petJ*)::*hemJ* strain was grown in the medium without copper; $+Cu^{2+}$ refers to the cells to which medium were added 1 µM $CuSO_4$ for 2 days. Glc+ strains were grown with 5mM glucose and Glc- were measured 3 days after exchange of the medium for the one without glucose. Membranes from Glc- strains were used for CN-PAGE (D). The gels (B, D) were scanned in transmittance mode (Scan) using an LAS 4000 Imager (Fuji), chlorophyll fluorescence (Chl FL) emitted by PSII was excited by blue light and detected also by LAS 4000.

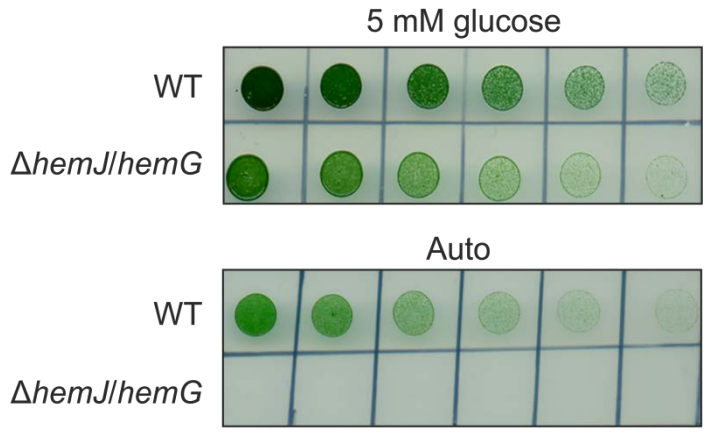

Figure S5. Mixotrophic (5mM glucose) and photoautotrophic growth (Auto) of WT and Δ*hemJ*/*hemG* strains of *Synechocystis* 6803 on agar plates

Strains were cultivated for 4 days under normal light conditions (40 µmol photons m^{-2} s^{-1}) at 28 °C. 5 µl of the cultures OD_{730}= 0.2 was applied to the plate on the left side followed with two-fold serial dilutions.

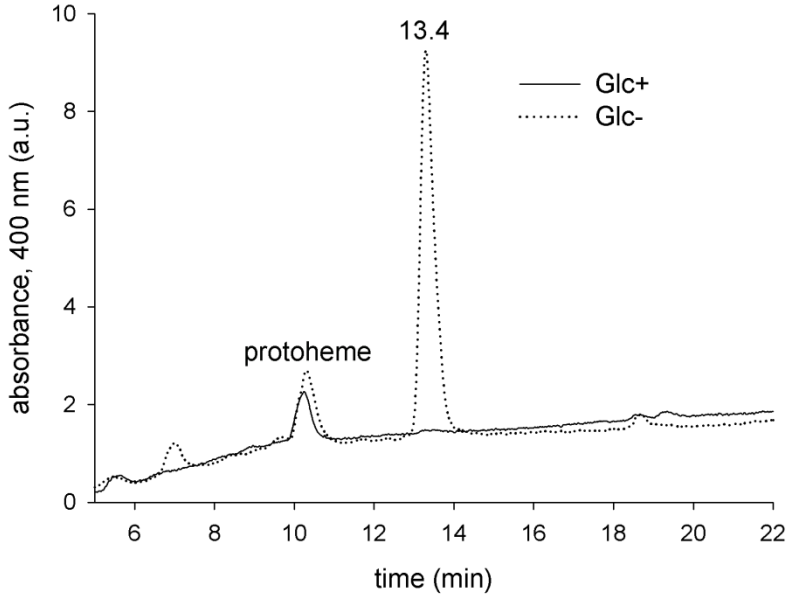

Figure S6. Accumulation of Harderoporphyrin in the Δ*hemJ*/*hemG* strain under photoautotrophic conditions

Pigments were extracted from the mutant cells grown in the presence of 5mM glucose (Glc+) or from cells incubated for three days without glucose (Glc-). Extracted pigments were separated by HPLC and detected by a diode-array detector at 400 nm (see Experimental Procedures).

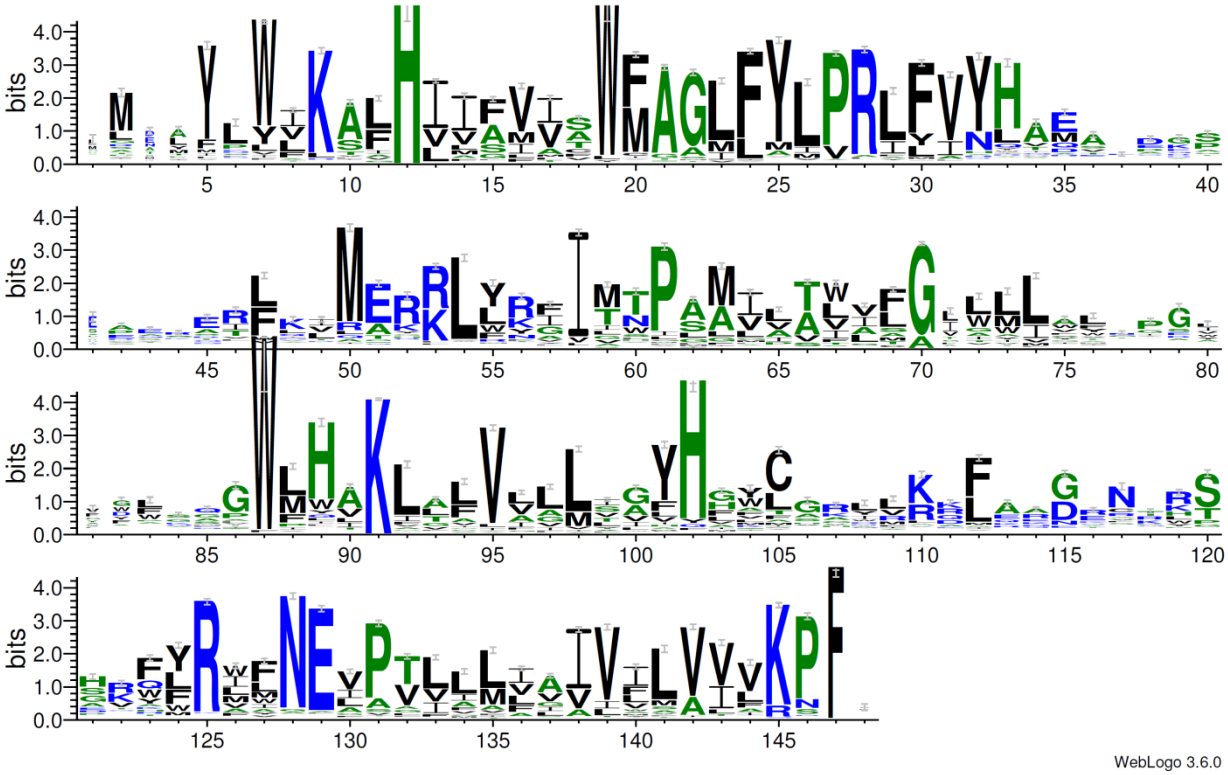

Figure S7. Sequence conservation of the proteins from PF03653 Pfam family

Consensus sequence of the Pfam family of proteins containing Slr1790, generated with Weblogo 3.6.0. (7). The height of the letters correlates with their conservation. Black, blue and green letters indicate hydrophobic, hydrophilic and other residues, respectively. The data for this logo consist of 1498 sequences from the full Pfam alignment of this family (Accession number PF03653). Amino acid residues His12, Trp87 and Lys91 are discussed in the text.

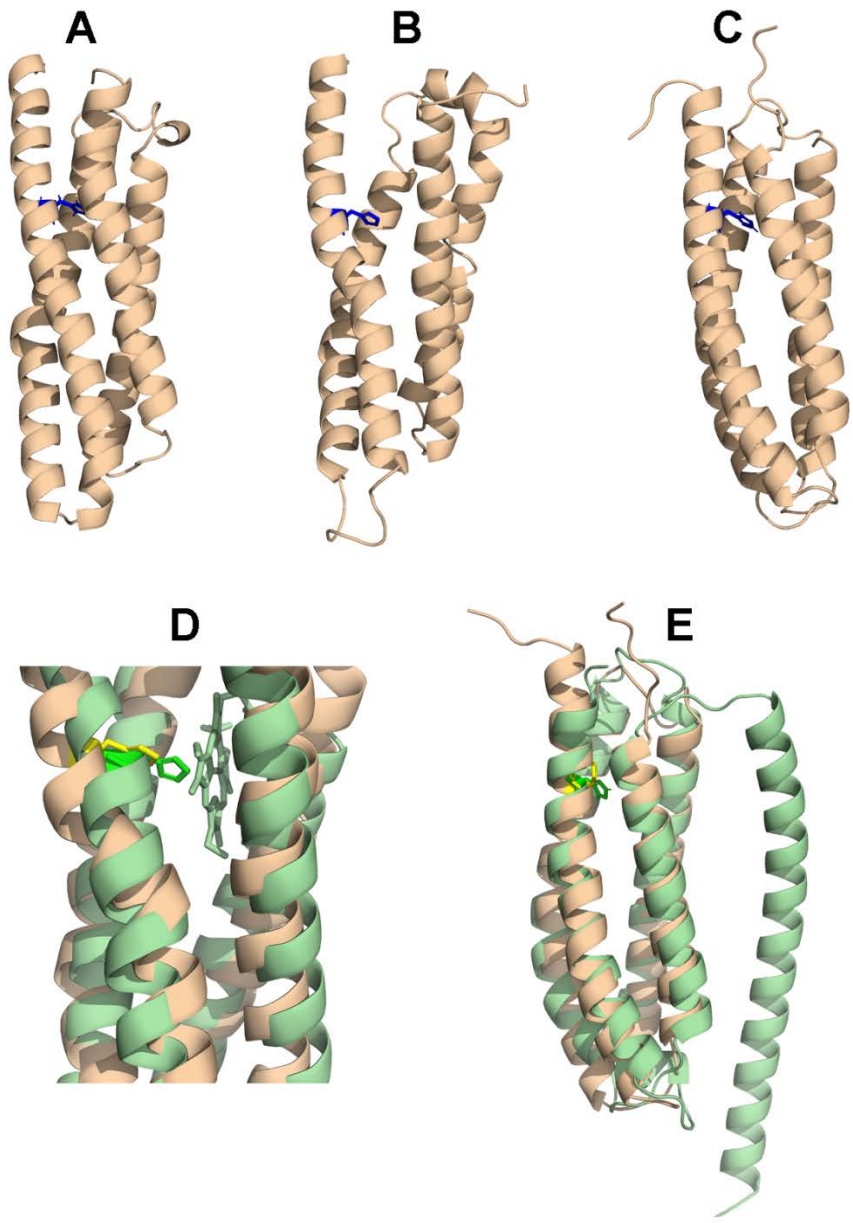

Figure S8. Comparison of model structures of *R. sphaeroides* HemJ

HemJ models predicted by automatic structure prediction servers – (A) Robetta (8), (B) RaptorX (9) and (C) RaptorX-Contact (10).The His16 proposed to bind heme is shown in blue. (D) Alignment of His16 in *R. sphaeroides* HemJ model with conserved His51 in cytb_{561} serving as a heme ligand. HemJ is shown in beige, cytb_{561} in olive green. His16 is yellow, His51 is green. (E) Alignment of HemJ from *R. sphaeroides* (beige) and *Synechocystis* 6803 (olive green) HemJ RaptorX-Contact model with conserved His.

References

1. Ovchinnikov, S., Park, H., Varghese, N., Huang, P.-S., Pavlopoulos, G. A., Kim, D. E., Kamisetty, H., Kyrpides, N. C., and Baker, D. (2017) Protein structure determination using metagenome sequence data. *Science* **355**, 294-298
2. Miller, M. J., and Gennis, R. B. (1983) The purification and characterization of the cytochrome *d* terminal oxidase complex of the *Escherichia coli* aerobic respiratory chain. *J Biol Chem* **258**, 9159-9165
3. Ovchinnikov, S., Kinch, L., Park, H., Liao, Y., Pei, J., Kim, D. E., Kamisetty, H., Grishin, N. V., and Baker, D. (2015) Large-scale determination of previously unsolved protein structures using evolutionary information. *eLife* **4**, e09248
4. Safarian, S., Rajendran, C., Müller, H., Preu, J., Langer, J. D., Ovchinnikov, S., Hirose, T., Kusumoto, T., Sakamoto, J., and Michel, H. (2016) Structure of a *bd* oxidase indicates similar mechanisms for membrane-integrated oxygen reductases. *Science* **352**, 583-586
5. Lu, P., Ma, D., Yan, C., Gong, X., Du, M., and Shi, Y. (2014) Structure and mechanism of a eukaryotic transmembrane ascorbate-dependent oxidoreductase. *Proceedings of the National Academy of Sciences* **111**, 1813-1818
6. Kořený, L., Sobotka, R., Kovářová, J., Gnipova, A., Flegontov, P., Horvath, A., Oborník, M., Ayala, F. J., and Lukes, J. (2012) Aerobic kinetoplastid flagellate *Phytomonas* does not require heme for viability. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 3808-3813
7. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190
8. Kim, D. E., Chivian, D., and Baker, D. (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research* **32**, W526-W531
9. Kallberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., and Xu, J. (2012) Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* **7**, 1511-1522
10. Wang, S., Sun, S., Li, Z., Zhang, R., and Xu, J. (2017) Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. *PLOS Computational Biology* **13**, e1005324