


A New Paradigm of Multiheme Cytochrome Evolution by Grafting and Pruning Protein Modules

Ricardo Soares,^{1,2} Nazua L. Costa,¹ Catarina M. Paquete,¹ Claudia Andreini,³ and Ricardo O. Louro ^{*,1}

¹Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal

²Instituto Nacional de Investigação Agrária e Veterinária, Oeiras, Portugal

³Magnetic Resonance Center and Department of Chemistry, University of Florence, Sesto Fiorentino, Italy

*Corresponding author: E-mail: louro@itqb.unl.pt.

Associate editor: Crystal Hepp

Abstract

Multiheme cytochromes play key roles in diverse biogeochemical cycles, but understanding the origin and evolution of these proteins is a challenge due to their ancient origin and complex structure. Up until now, the evolution of multiheme cytochromes composed by multiple redox modules in a single polypeptide chain was proposed to occur by gene fusion events. In this context, the pentaheme nitrite reductase NrfA and the tetraheme cytochrome *c*₅₅₄ were previously proposed to be at the origin of the extant octa- and nonaheme cytochrome *c* involved in metabolic pathways that contribute to the nitrogen, sulfur, and iron biogeochemical cycles by a gene fusion event. Here, we combine structural and character-based phylogenetic analysis with an unbiased root placement method to refine the evolutionary relationships between these multiheme cytochromes. The evidence show that NrfA and cytochrome *c*₅₅₄ belong to different clades, which suggests that these two multiheme cytochromes are products of truncation of ancestral octaheme cytochromes related to extant octaheme nitrite reductase and MccA, respectively. From our phylogenetic analysis, the last common ancestor is predicted to be an octaheme cytochrome with nitrite reduction ability. Evolution from this octaheme framework led to the great diversity of extant multiheme cytochromes analyzed here by pruning and grafting of protein modules and hemes. By shedding light into the evolution of multiheme cytochromes that intervene in different biogeochemical cycles, this work contributes to our understanding about the interplay between biology and geochemistry across large time scales in the history of Earth.

Key words: multiheme cytochromes, biogeochemical cycles, gene fusion, gene fission, modular evolution.

Introduction

Multiheme cytochrome *c* (MHC) catalyzes diverse chemical reactions in prokaryotes providing them with a remarkable metabolic versatility (Simon et al. 2011; Bewley et al. 2013; Deng et al. 2018). These metalloproteins exploit the redox, spin, and acid–base properties of the heme cofactors to perform chemical reactions that are pivotal to many biogeochemical cycles, including those of nitrogen, sulfur, and iron (Pereira and Xavier 2006; Mayfield et al. 2011; Paquete et al. 2019). Some MHC families share sequence and structural similarities that unequivocally reflect a common ancestral origin (Sharma et al. 2010). Understanding their evolution has been a challenge due to the low preserved amino acid sequence and limited available 3D structures that are better preserved than sequence across large timescales (Chothia and Lesk 1986; Illergård et al. 2009; Ingles-Prieto et al. 2013). Nevertheless, the current paradigm for the emergence of the different known MHC follows the proposed fusion of redox modules, going from simple MHC to more complex and containing more heme cofactors per polypeptide

chain. This comes from the observation of the presence of repetitive redox modules in MHC that has enabled to relate different MHC families (Roldán et al. 1998; Carrondo et al. 2006; Santos-Silva et al. 2007; Clarke et al. 2011; Pokkuluri et al. 2011; Pereira et al. 2017; Edwards et al. 2020).

In the case of the pentaheme nitrite reductase (NrfA) and the octaheme hydroxylamine oxidoreductase (HAO), structural similarities of the heme-core and of the three interface-forming helices have long been pointed out as a sign of the common origin between these two MHCs that have distinct functions in the nitrogen cycle (Igarashi et al. 1997; Einsle et al. 1999). Octaheme nitrite reductases (ONRs) later discovered showed even more similarities with NrfA with respect to its structure and function (Polyakov et al. 2009; Tikhonova et al. 2012). Cytochrome *c*₅₅₄ (cyt *c*₅₅₄) that is unrelated to NrfA, also showed similarities with HAO regarding the heme-core structure (Iverson et al. 1998), suggesting that diverse evolutionary mechanisms took place for the emergence of these proteins. In all of these proteins, catalysis takes

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

place at a heme with one open coordination site (penta-coordinated) in the iron for access of the substrate.

Using phylogenetics and placing NrfA at the root, an ancestral ONR was proposed to be an intermediary for the evolution of NrfA to the presumably more recent HAO family of proteins (Klotz et al. 2008). In that proposal, an unknown triheme cytochrome fused with NrfA to originate ONR. That simple evolutionary scheme required revision as more homologous MHCs were characterized. For example, the structure of the copper-containing sulfite reductase MccA (Hermann et al. 2015), which also reduces nitrite, and the octaheme nitrite reductase lhOCC (Parey et al. 2016) revealed similarities with NrfA, ONR, and HAO. However, unlike ONR and HAO, the penta-coordinated heme of MccA is located in the N-terminal region (Hermann et al. 2015). This feature is also found in the octaheme tetrathionate reductase (OTR; Mowat et al. 2004) that was proposed to have diverged from NrfA via a different route based on the lack of the interface-forming helices characteristic of HAO and ONR (Klotz et al. 2008).

Table 1. List of the Homologous MHC.

Protein	PDB Code	Number of Hemes	Organism/Class
Cyt _{c₅₅₄}	1FT5	4	<i>Nitrosomonas europaea</i> / Betaproteobacteria
NrfA	1QDB	5	<i>Sulfurospirillum deleyianum</i> / Epsilonproteobacteria
	1FS7	5	<i>Wolinella succinogenes</i> / Epsilonproteobacteria
	1GU6	5	<i>Escherichia coli</i> / Gammaproteobacteria
	1OAH	5	<i>Desulfovibrio desulfuricans</i> / Deltaproteobacteria
	2J7A	5	<i>Desulfovibrio vulgaris</i> / Deltaproteobacteria
	3UBR	5	<i>Shewanella oneidensis</i> / Gammaproteobacteria
	6V0A	5	<i>Geobacter lovleyi</i> / Deltaproteobacteria
ONR	2OT4	8	<i>Thioalkalivibrio nitratireducens</i> / Gammaproteobacteria
	3SXQ	8	<i>Thioalkalivibrio paradoxus</i> / Gammaproteobacteria
HAO	1FGJ	8	<i>Nitrosomonas europaea</i> / Betaproteobacteria
	4N4J	8	<i>Candidatus Kuenenia stuttgartiensis</i> / Candidatus Brocadia
HDH	6HIF	8	<i>Candidatus Kuenenia stuttgartiensis</i> / Candidatus Brocadia
lhOCC	4QO5	8	<i>Ignicoccus hospitalis</i> /Thermoprotei (Archaea)
OTR	1SP3	8	<i>Shewanella oneidensis</i> / Gammaproteobacteria
MccA	4RKM	8	<i>Wolinella succinogenes</i> / Epsilonproteobacteria
OcwA	6I5B	9	<i>Thermincola potens</i> /Clostridia
OmhA	6QVM	11	<i>Carboxydotherrnus ferrireducens</i> / Clostridia

For each MHC, the PDB code of the 3D structure, the number of hemes, and the source organism and the taxonomic class are indicated.

More recently, the structure of the cell-surface nonaheme OcwA from the electroactive bacterium *Thermincola potens* JR was determined (Costa et al. 2019). This protein, which has iron-oxide reductase activity in vitro in agreement with its proposed physiological role, did not reveal structural similarities with other structurally characterized cell-surface iron reductases (Edwards et al. 2012; Liu et al. 2014; Edwards et al. 2015; Wang et al. 2019; Edwards et al. 2020). In contrast, OcwA showed similarities with NrfA at the C-terminus and cyt_{c₅₅₄} at the N-terminus, conserving the penta-coordinated hemes of both proteins. These observations led to the proposal that an ancestral OcwA-like protein could represent the evolutionary link between NrfA and cytochrome c₅₅₄ with the different octaheme cytochromes (Costa et al. 2019). In this scenario, OcwA would have originated from a gene fusion event between ancestral NrfA and cyt_{c₅₅₄}. Subsequent loss of one of the two penta-coordinated hemes and diversification would have led to the different extant octaheme cytochromes (Costa et al. 2019). However, these previous studies (Klotz et al. 2008; Costa et al. 2019) lacked a phylogenetic analysis with a combination of structural and character (sequence) information, and an unbiased root placement method. Here, we used these criteria combined with minimal functional-site characterization of the hemes to refine the previous evolutionary proposals. Our phylogenetic analysis revealed that cyt_{c₅₅₄} and NrfA are products of a truncation event from different octaheme cytochromes and that the last common ancestor (LCA) is inferred to be an octaheme cytochrome able to reduce nitrite.

Results

Identification of the Homologous Group of MHC

The MHCs analyzed in this work (table 1) are involved in diverse metabolic pathways. NrfA, ONR, lhOCC, MccA, and OTR are involved in dissimilatory nitrite and/or sulfite and/or tetrathionate reduction. Cyt_{c₅₅₄}, HAO, and hydrazine dehydrogenase (HDH) are involved in ammonia oxidation reactions. HAO oxidizes hydroxylamine within the aerobic ammonia oxidation metabolism; HDH oxidizes hydrazine in the anaerobic ammonia oxidation (anammox) metabolism; cyt_{c₅₅₄} functions as an electron transfer protein from HAO to membrane-bound cytochrome c₅₅₂ (for a review, see Paquete et al. 2019). OcwA is involved in dissimilatory iron reduction by extracellular electron transfer (Carlson et al. 2012; Costa et al. 2019). Nevertheless, all of them display nitrite or nitric oxide reductase activity in vitro (Costa et al. 2019; Paquete et al. 2019). Sequence and structural similarities allowed the assignment of the recently identified undecaheme OmhA (Gavrilov et al. 2021) as belonging to this homologous group (table 1). This protein was isolated from the S-layer of the thermophilic and Gram-positive bacterium *Carboxydotherrnus ferrireducens* (Gavrilov et al. 2012). OmhA shares 29.5% semi-global sequence identity and its structure has a C- α root mean square deviation

(RMSD) of 2.45 Å (PyMOL; Schrödinger, Inc.) with OcwA. The amino acid sequence of OmhA contains an extra N-terminal extension harboring two extra hemes in comparison with OcwA ([supplementary fig. S1, Supplementary Material](#) online). Sequence and structural searches did not show any homologous MHC to this small region.

The MHCs listed in [table 1](#) share one or several structurally homologous features ([fig. 1](#)). The exception is *cyt c₅₅₄* and NrfA that lack similarity between each other as these two cytochromes align at opposing ends (N- and C-terminal, respectively) of the octa-, nona-, and undeca-heme cytochromes. Based on this observation, *cyt c₅₅₄* and NrfA are used to define N- and C-terminal modules of the MHC analyzed in this work ([fig. 1](#)), respectively. All other MHCs share a similar heme-core arrangement with three or more hemes within these modules displaying

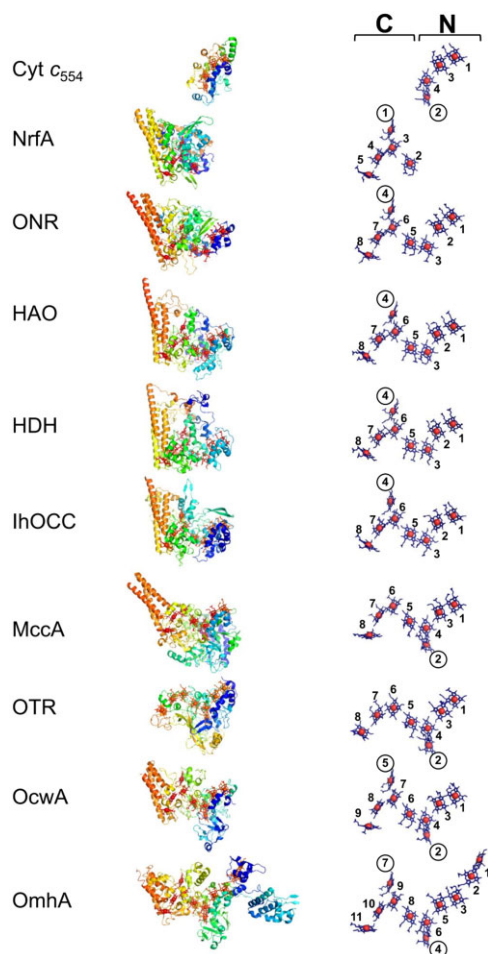


Fig. 1. Aligned structures for the representative MHC of the three main clusters. PDB structures used were *Cyt c₅₅₄*, 1FT5_A; NrfA, 1QDB_A; ONR, 2OT4_A; HAO, 1FGJ_A; HDH, 6HIF_A; lhOCC, 4QO5_A; OTR, 1SP3_A; MccA, 4RKM_A; OcwA, 6I5B_A; OmhA, 6QVM_A. Cartoon and heme-core representations are depicted. Cartoon representations are rainbow colored from N- (right) to C- (left) terminal. Heme-core representation contains heme position number and penta-coordinated hemes (penta-coordinated) are identified by a circle. The C- and N-terminal modules are indicated by the respective letters.

good structural superimposition. One key difference is the presence of one or two penta-coordinated hemes. NrfA, ONR, HAO, HDH, and lhOCC contain the penta-coordinated heme within the “C-terminal module,” whereas *cyt c₅₅₄*, MccA, and OTR contain the penta-coordinated heme within the “N-terminal module” ([fig. 1](#)). OcwA and OmhA are currently the only proteins that possess both penta-coordinated hemes. Another notable observation is the presence or absence of the three C-terminal oligomerization-forming helices that are absent in *cyt c₅₅₄* (lacks the C-terminal module) and in OTR ([fig. 1](#)). The distal axial ligands’ position is conserved in all MHCs except in OTR where the distal ligands of hemes 1 and 7 are found in different relative positions in the amino acid sequence ([supplementary fig. S2, Supplementary Material](#) online).

Comparison of the Backbone Trace and Heme-Core Architecture

In order to assess the structural distance within these MHCs, the backbone and heme-core arrangements were compared. As shown by the network representation in [figure 2](#) (corresponding distance matrices are available in [supplementary tables S1–S4, Supplementary Material](#) online), OTR is significantly different from all other MHCs. This protein appears as an outlier with respect both to the backbone and heme-core structural comparisons ([fig. 2](#)). Based on the backbone trace and heme-core architecture, the remaining MHCs clustered in three distinct clades: HAO, HDH, and lhOCC (clade 1); NrfA and ONR (clade 2); *cyt c₅₅₄*, MccA, OcwA, and OmhA (clade 3). In clades 1 and 2, intracluster structural distances are globally shorter than in clade 3, which did not always meet the selected threshold for node (structure) connection (gray lines in [fig. 2](#)). Nevertheless, clade 3 MHCs were consistently found together when compared with the rest of the MHC ([fig. 2](#)), with the exception of the backbone trace comparisons at the C-terminal where clades 2 and 3 are close together by the short distance between OcwA and one NrfA structure (PDB entry 3UBR; [fig. 2B](#)). Within clade 3, the cell-surface MHC OcwA and OmhA presented the shortest distances, with the exception for the structural comparisons of the heme-core arrangement at the C-terminal ([fig. 2D; supplementary table S4, Supplementary Material](#) online). In this case, OcwA and MccA pair showed the shortest distances.

Comparison of the Heme Centers as Minimal Functional Sites

Minimal functional sites in metalloproteins (MFSs) are portions of the 3D structure that focus on the region around the metal site(s). MFSs of heme centers were extracted from representative structures of the MHC listed in [table 1](#), compared all versus all with the Metals² tool ([Andreini et al. 2013](#)), and clustered into groups of structurally similar MFSs. This analysis resulted in 10 groups, comprising 72 MFSs out of the 77 included in the analysis,

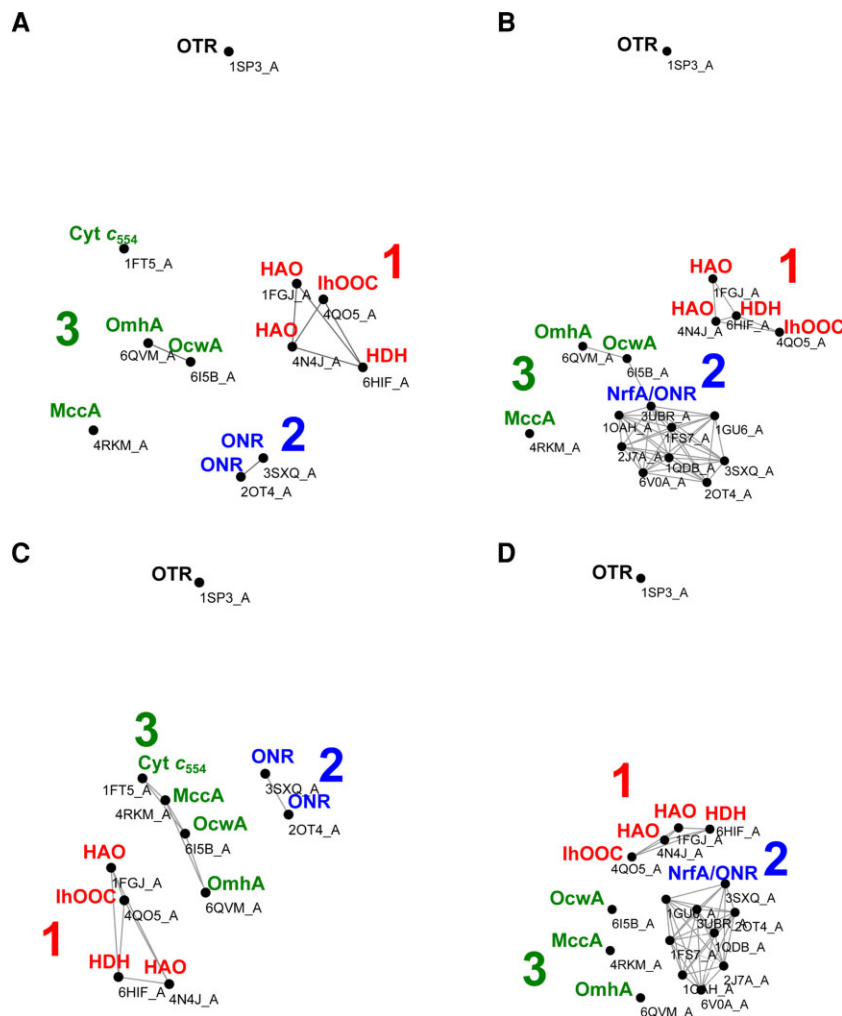


Fig. 2. Distance network analysis using structural information. Structural comparisons were performed using distance comparisons of the backbone structure at the N-terminal (A) and C-terminal (B) and of the heme-core at the N-terminal (C) and C-terminal (D). Each node (black dot) represents one structure. Gray lines connect structures with distances ≤ 3.5 and 0.75 Å for the backbone and heme-core structures, respectively.

Table 2. MFS Groups by Heme Position in the Amino Acid Sequence within Each MHC.

Protein	PDB	Heme Position										
		1	2	3	4	5	6	7	8	9	10	11
HAO	1fgj_A			H		B	A	C	G	A	B	A
HDH	6hif_A			H		B	A	C	G	A	B	A
lhOCC	4qo5_A			H		B	A	C	G	A	B	A
ONR	2ot4_A			E		Other	A	C	D	A	B	A
NrfA	1qdb_A							C	D	A	B	A
OcwA	6i5b_A			E	Other	F	A	C	D	A	B	A
OmhA	6qvm_A	D	Other	E	C	F	A	C	D	A	B	A
MccA	4rkm_A			A	Other	F	A		D	A	B	A
Cyt c ₅₅₄	1ft5_A			E	C	F	A					
OTR	1sp3_A			E	Other	G	I		J	I	J	A

For each MHC, the PDB code of the 3D structure, the position of hemes in the sequence (numbered according to OmhA numbering in [fig. 1](#)) and the group of MFSs to which each heme center belongs are indicated.

which were arbitrarily named group A (26 MFSs), group B (11 MFSs), group C (9 MFSs), group D (6 MFSs), group E (5 MFSs), group F (4 MFSs), group G (4 MFSs), group H (3 MFSs), group I (2 MFSs), and group J (2 MFSs; [supplementary fig. S3, Supplementary Material](#) online). [Table 2](#) shows the groups to which each heme center in the MHC belongs. As MFSs describe the local environment around a metal cofactor, each group of MFSs identifies a

set of heme centers in MHC with common structural features, which for the sake of simplicity we refer to as a heme “type.” Observation of [table 2](#) confirms that (1) OTR represents an outlier with respect to all the other MHCs, as only the heme centers at positions 3 and 11 are of a type shared by other MHCs; and (2) HAO, HDH, and lhOCC represent a clear subgroup (clade 1 above), as all of them have the same pattern of heme types, which includes three centers

(at positions 3, 5, and 8) different from all the other MHCs. The relationships between the other MHCs (ONR, NrfA, OcwA, OmhA, MccA, and cyt c_{554}) are not clearly defined with this analysis, as the differences are mainly related with absence of specific hemes or the presence of heme types that are not clustered.

Overall, the results of the analysis of MFSs are consistent with those obtained for the *Comparison of the backbone trace and heme-core architecture*, which found HAO and lhOCC clustered in clade 1 and the remaining MHCs divided in clades 2 and 3 (fig. 2).

Character-Based Phylogeny

A total of 5700 sequences (accession codes and sequences are provided in [supplementary files, Supplementary Material](#) online) were collected from NCBI RefSeq database ([supplementary table S5, Supplementary Material](#) online). These correspond to highly homologous sequences to at least one of the reference MHC listed in [table 1](#). These sequences ranged from 23.1 to 100% of global identity. Globally the diversity of prokaryotes harboring homologous MHC was greatly expanded from those whose MHCs have been structurally characterized ([table 1](#)). A total of 22 phyla and two domains (Bacteria and Archaea) were found in this analysis ([supplementary table S5, Supplementary Material](#) online). NrfA was the most represented family within the database in terms of number of sequences and diversity, whereas the cell-surface MHCs (OcwA and OmhA) were very poorly represented. As OTR was shown to be an outlier in the structural comparisons (see backbone trace, heme-core architecture, and minimal functional-site comparisons), it was not included in this analysis. HAO and HDH homologs greatly overlapped as HAO and HDH from *Kuenenia stuttgartiensis* are phylogenetically closer than HAO from *Nitrosomonas europaea*. In this sense, this group was joined hereon (HAO/HDH).

After redundancy reduction (see Sequence Collection and Clustering) phylogenetic inference was performed using multiple sequence alignments (MSAs) either containing NrfA (NrfA+) or cyt c_{554} (cyt c_{554} +) along with the rest of the MHC, as these cytochromes align at opposing ends of the sequence of the remaining MHCs. MSAs are provided in [supplementary files, Supplementary Material](#) online. Reconstructed phylogenies consistently indicated the presence of three main monophyletic clades composed by HAO/HDH and lhOCC (clade 1), NrfA and ONR (clade 2), and OcwA, OmhA, MccA, and Cyt c_{554} (clade 3; [fig. 3A–D](#)). All monophyletic proposed clades are supported with SH-aLTR/Bootstrap or posterior probabilities above 70%. For clades 1 and 2, these statistical support values are above 90%. Within the last clade, the cell-surface MHCs were clustered together (OcwA and OmhA). Likewise, MccA and cyt c_{554} were clustered together within this clade. Minimal ancestor deviation rooting analysis ([Tria et al. 2017](#)) consistently placed the root between clade 1 and the junction of clades 2 and 3.

Mismatch between the phylogenetic position of some sequences and the taxonomic classification of the organisms from which these sequences derive were observed across all clades analyzed ([supplementary figs. S4–S6, Supplementary Material](#) online). This suggests widespread occurrence of horizontal gene transfer (HGT) for these MHCs. Notably, OcwA and OmhA sequences only presented vertical transfer in the current data set. However, this may be a consequence of the poor representation of these two MHCs in the public database analyzed. It contains only two and four highly homologous sequences, respectively, and from bacteria belonging the same genus, *Thermincola* and *Carboxydotherrmus*, respectively ([supplementary table S5, Supplementary Material](#) online).

Conservation of Critical Residues for Function

In order to gain further insights about the emergence of the unique features that differentiate these MHCs, the conservation of critical residues for function was searched and compared within each MHC clade ([fig. 4](#)). The ability to oxidize hydroxylamine or hydrazine appears to have a relatively recent origin within clade 1, judging by the branching points of the homologous sequences that conserve the tyrosine residue at the crosslink position ([supplementary fig. S7, Supplementary Material](#) online). In clade 2, all ONRs and NrfAs contain the typical known features for nitrite reduction, namely the lysine as the proximal ligand of the penta-coordinated heme and the catalytic tyrosine, histidine, and arginine, which correspond to K131, Y217, H282, and R113 in *Sulfurospirillum delleyianum* (PDB entry 1QDB_A), respectively. The only exception is the NrfA variants without the lysine proximal ligand of the penta-coordinated heme ([supplementary fig. S8, Supplementary Material](#) online) that appear to have emerged recently within clade 2. Clade 3 is more diversified in terms of function and structure than clades 1 and 2. Nevertheless, the identified residues that are important for catalysis are all conserved in OcwA and MccA, which have catalytic residues assigned in their structures. All OcwA and OmhA sequences conserve the nonbonding histidines that are localized in the distal side of the penta-coordinated hemes. These correspond to H282 and H381 in *T. potens* JR (PDB entry 6I5B_A). OcwA and OmhA conserve also two extra histidines nearby the first and second penta-coordinated hemes (positions 4 and 7 in [table 2](#)), H281 and H380 in *T. potens* JR (PDB entry 6I5B_A). *T. potens* JR H380 is conserved in all OcwA and OmhA sequences, whereas *T. potens* JR H281 is conserved in all OcwA sequences and in half of the OmhA sequences. All MccA sequences conserve the known residues that are important for catalysis, namely Y123, K208, Y285, Y301, R366, K393, C399, and C495 in *Wolinella succinogenes* (PDB entry 4RKM_A). The eighth heme (position 11 in [table 2](#)) has the uncommon CX₁₅CH motif ([Hermann et al. 2015](#)). In addition, we found also the presence of the CX₁₇CH heme-binding motif for the eighth heme of the MccA sequences in our data set.

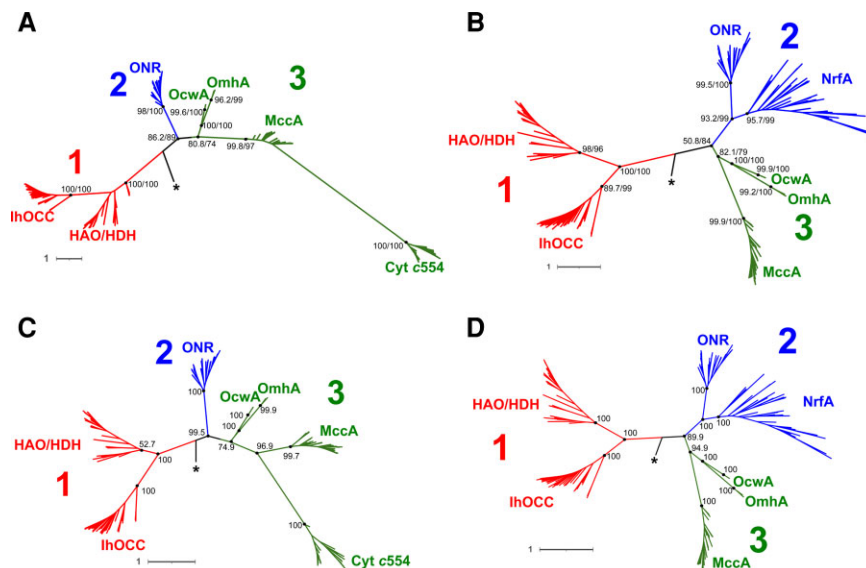


FIG. 3. Character-based phylogenetic analysis for the homologous MHC. Maximum likelihood phylogenetic trees based on *cyt c554*+ (A) and *NrfA*+ (B) multiple sequence alignments. Bayesian phylogenetic trees based on *cyt c554*+ (C) and *NrfA*+ (D) multiple sequence alignments. Protein sequences fell in three main clades according with their tree positioning and are represented in different colors (red, blue, and green) and labelled with 1, 2 and 3. Asterisks indicate the root placement according with the minimal ancestor deviation method (Tria et al. 2017). Bootstrap/SH-aLTR (ML trees) or posterior probabilities (Bayesian trees) confidence percentages values are presented near each node for the major splits. Tree scale represents number of substitutions per site.

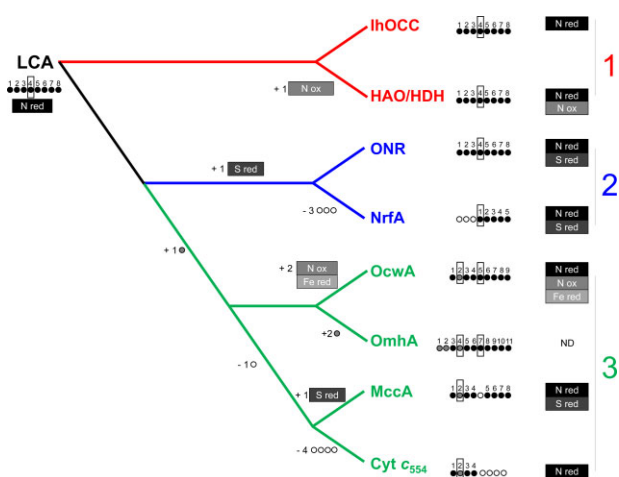


FIG. 4. Cladogram with the most parsimonious solution for the evolution of the MHC after phylogenetic analysis. Circles and rectangles represent the hemes and catalytic activities, respectively, for each MHC. Dark circles and rectangles represent hemes and activities inherited from the LCA, whereas gray and white represent heme or activity gain and loss, respectively. Gains and losses of hemes and activities from the LCA to extant MHC are depicted near each respective position in the cladogram. Penta-coordinated hemes are marked within black rectangles. ND represents no data.

Discussion

The dissection of the origin and evolution of MHC has attracted significant scientific interest but has also been characterized by numerous challenges due to the small sequence similarity that was preserved among these proteins along time. Nevertheless, a great number of MHCs have been identified as homologs based on structural similarities. The timeline for their emergence is typically thought to progress by fusion of redox modules towards more

complex MHC. These can reach up to 16 hemes for structurally characterized MHC, but genes coding putative MHC containing as much as 113 hemes per polypeptide chain have been reported (Roldán et al. 1998; Carrondo et al. 2006; Santos-Silva et al. 2007; Clarke et al. 2011; Pokkuluri et al. 2011; Pereira et al. 2017; Edwards et al. 2020; Leu et al. 2020). Towards understanding the evolution of MHC, we focused our attention on a group of MHC that is involved in the biogeochemical cycles of iron, nitrogen, and sulfur.

In addition to gene fusion, gene fission is also a well-established mechanism for diversity of domain combination in all domains of life (Kummerfeld and Teichmann 2005; Pasek et al. 2006; Weiner and Bornberg-Bauer 2006). However, only gene fusion was considered when proposing the origin of the octa and nonaheme cytochromes by *NrfA* and *cyt c554* (Klotz et al. 2008; Costa et al. 2019). Indeed, the previous proposals for MHC evolution were biased towards a fusion event (Klotz et al. 2008; Costa et al. 2019). By performing a thorough phylogenetic inference with both structural and sequence data, we now show that *NrfA* and *cyt c554* were likely formed by truncation of octaheme cytochromes from different clades. Phylogenetic analysis revealed three main clades composed by HAO/HDH and IhOCC (clade 1), *NrfA* and ONR (clade 2), and *OcwA*, *OmhA*, *MccA*, and *cyt c554* (clade 3). *NrfA* and *cyt c554* belong to different clades and consequently, the MHCs that are phylogenetically closer to *NrfA* are not the same for *cyt c554*. The closest homolog of *NrfA* is ONR, whereas for *cyt c554*, the closest homolog is *MccA*. This implies that a fusion event between *NrfA* and *cyt c554* was not responsible for *OcwA* origin.

Although previous studies presented unrooted trees (Bergmann et al. 2005; Klotz et al. 2008), *NrfA* was chosen

as the implicit root for those phylogenetic analyses, based on the premise that nitrate/nitrite ammonification is the most ancestral reaction when compared with ammonia oxidation reactions. However, for a global cycle dating perspective, isotope signatures cannot at present, discriminate between all of the different reactions within the nitrogen cycle (Stüeken et al. 2016). Moreover, since those studies, other MHCs capable of nitrite reduction besides NrfA were identified (e.g., lhOCC; Parey et al. 2016) and dissimilatory iron reduction is also considered to be a very early form of bioenergetic metabolism (Vargas et al. 1998; Johnson et al. 2008). Attempts at a rigorous assessment of relative dating using taxonomics in order to solve the root placement is complicated by the evidence that most of these MHCs have undergone considerable HGT, which was detected in the present analysis. Likewise, HGT of MHC has been extensively reported in the literature, including for NrfA (Welsh et al. 2014), HAO (Bergmann et al. 2005), and several MHCs involved in extracellular electron transfer, such as MtrCAB and OmcA from *Shewanella* spp. (Zhong et al. 2018; Baker et al. 2022).

Outgroup rooting is by far the most used method to infer the root position in a given phylogenetic tree but it is unsuitable when the ingroup is already very divergent and no appropriate outgroup is available. Considering this, we used minimal ancestor deviation method for rooting our trees (Tria et al. 2017). This method is an updated version of the midpoint rooting method by enabling perturbations to the molecular clock and therefore heterotachy. This analysis placed the root between clade 1 and the junction of clades 2 and 3. Using this information, we can conclude that clade 1 is the most ancient, as it represents the basal clade from the root position, followed by clades 2 and 3.

Using the most parsimonious solution for inferring ancestral states, the LCA is inferred to be an octaheme cytochrome that was able to reduce nitrite. In the timeline that we propose (fig. 4), clade 1 was the first to branch into HAO/HDH and lhOCC, both having the same heme-core architecture of the LCA. The penta-coordinated heme of HAO/HDH is a P460 heme and most of the sequences collected for HAO/HDH conserve the tyrosine residue at crosslinking position of the representative structures (supplementary fig. S7, Supplementary Material online). Contrarily, the P460 heme is absent in lhOCC and it is predicted to be absent also in the LCA. In this scenario, the P460 heme evolved de novo in the HAO/HDH subclade. Indeed, an HAO family protein that lacks the tyrosine crosslink (HAOr) was recently isolated from *K. stuttgartiensis* and showed no hydroxylamine oxidation activity, while presenting nitrite reduction to nitric oxide (Ferousi et al. 2021). The second branching clade was clade 2, that latter diversified into ONR that conserved the overall heme-core arrangement of the LCA, and NrfA that further lost the three N-terminal hemes. The lysine residue as the proximal ligand of the penta-coordinated heme seems to be an innovation of this clade, conserved in all ONR sequences.

In contrast, not all sequences that were collected for NrfA conserve this lysine residue even though no structure is available for these variants. Clade 3 appeared more recently by the addition of a second penta-coordinated heme (position 4 in table 2). OcwA and OmhA still conserve also the first penta-coordinated heme (position 7 in table 2) that is common to the two more ancient clades (1 and 2), and OmhA gained two extra hemes at the N-terminus. In contrast, MccA and cyt c_{554} have lost the first penta-coordinated heme (position 7 in table 2), and cyt c_{554} further lost the last four C-terminal hemes. Inference that the LCA for the MHC involved in the nitrogen and sulfur cycles was an octaheme cytochrome and that loss of hemes took place, was previously proposed and discussed (Kern et al. 2011; Simon and Klotz 2013). However, figure 1 shows that the sequence alignments used by Kern et al. (2011) are not congruent with the respective 3D structure alignments. Kern et al. (2011) aligned the penta-coordinated hemes of MccA and cyt c_{554} (position 4 in table 2) with the ones of NrfA, ONR, and HAO (position 7 in table 2), which results in an 180° γ -axis misalignment from the aligned protein structures of these two groups presented in figure 1. Two consequences arise from this: the tree topology artificially exacerbates the distance between these two groups and by proposing an alignment of NrfA with cyt c_{554} Kern et al. (2011) failed to realize the existence of redox modules, which is clearly apparent in figure 1. Since Kern et al. (2011), many MHC structures have been solved, which facilitates the correct identification of the mentioned redox modules.

The cladogram of figure 4 provides additional support to the analysis performed here given that it recapitulates the patterns of heme “types,” as described by MFSs, observed in extant MHC (supplementary fig. S9, Supplementary Material online). In this scenario, clade 1 conserved the local heme environments from the LCA, whereas clades 2 and 3 underwent successive modifications. The common branching of clades 2 and 3 from clade 1 involved the modification of the two first hemes (positions 3 and 5 in table 2). The subsequent origin of clade 3 involved the addition of a new penta-coordinated heme (position 4 in table 2) that had similar characteristics of the first penta-coordinated heme (position 7 in table 2) and further modification of the heme at position 5 (table 2). Diversification of clade 3 involved further modifications of the second penta-coordinated heme (OcwA and MccA) to new specific heme types (position 4 in table 2). Addition of two extra N-terminal hemes (positions 1 and 2 in table 2) for OmhA and modification of the heme at position 3 to the same type of positions 6, 9, and 11 (table 2) in the case of MccA also led to the diversification of the heme patterns of clade 3.

Conclusion

MHC were abundantly employed by nature in the development of multiple biogeochemical cycles across large

time scales, which were of high importance for the colonization of all extant ecological niches of life on earth. Previous proposals for the evolution of MHC have been biased towards gene fusion events (Roldán et al. 1998; Carrondo et al. 2006; Santos-Silva et al. 2007; Clarke et al. 2011; Pokkuluri et al. 2011; Pereira et al. 2017; Edwards et al. 2020). Our study changes this perspective by showing that fission of heterogeneous redox modules also drives the evolution of MHC and should be *a priori* equally considered. It shows that NrfA and cyt c_{554} likely resulted from truncation events of an ancestral ONR and MccA, respectively, and that the common ancestor of the MHC analyzed here was likely an octaheme cytochrome similar to extant lhOCC with nitrite reductase activity. Evolution from this ancestral octaheme cytochrome included pruning and grafting of heme-binding polypeptide modules, which led to the emergence of extant MHC that catalyze very distinct reactions within the nitrogen, sulfur, and iron biogeochemical cycles. Altogether, this work opens a new perspective in our understanding about the evolution of MHC and their changing role in the interplay between biology and geochemistry across large timescales.

Materials and Methods

Backbone and Heme-Core Comparisons

Protein structures were retrieved from the PDB database (table 1). Chain A of each protein was selected for subsequent structural comparisons. The data were divided in two subsets for analysis, one including all MHCs except NrfA and another including all MHCs except cyt c_{554} , as these align at opposing ends of the octa-, nona-, and undecaheme MHC. Heme-core comparisons were performed using PyMOL Molecular Graphics System (version 2.0 Schrödinger, LLC). The structural positions of the 33 non-mobile atoms of the porphyrin ring were selected for each heme. For the analysis including cyt c_{554} , the three common N-terminal hemes to all MHCs except NrfA were selected (positions 3, 5, and 6 in table 2). For the analysis including NrfA, the four common C-terminal hemes to all MHCs except cyt c_{554} were selected (positions 8, 9, 10, and 11 in table 2). Pair-fit function was used to align the corresponding positions of each pair. Pairwise RMSD values were used to construct a distance matrix that was then fed into Cytoscape 3.7.1 to construct networks based on structural distance (Shannon et al. 2003). Representation was generated using the prefuse force directed layout algorithm with 1 normalized weight values option and 1,000 iterations. Backbone structure of the MHC was compared using mTM-align (Dong, Pan, et al. 2018; Dong, Peng, et al. 2018). From the multiple structure alignment generated, pairwise RMSD were retrieved and used to build a matrix that was then used as input data for network representation in Cytoscape, in a similar procedure to that performed for the heme-core structural comparisons.

Minimal Functional Sites Analysis

MFSs were extracted for each heme in the PDB structures listed in figure 1 and compared using the Metals² tool (Andreini et al. 2013). The MFSs were then clustered by average linkage clustering using a threshold value of 2.25, which has been previously shown to indicate significant structural similarity between sites (Rosato et al. 2016). This procedure resulted in the clustering of 69 out of a total of 77 MFSs. A further three MFSs were included in clusters as all their similarity scores below 2.50 (Valasatava et al. 2015) were associated with MFSs all belonging to the same cluster. The residue numbers for each heme in the reference PDB structures are listed in supplementary table S6, Supplementary Material online.

Sequence Collection and Clustering

Sequences were retrieved from NCBI Reference Sequence database (RefSeq; O'Leary et al. 2016) using Blastp algorithm (Altschul et al. 1990). Sequences from all the proteins whose structure has been determined were used as queries (access date: September 15, 2021). An *e*-value of 1^{-50} was selected as a threshold. Protein sequences were filtered accordingly with the expected number of heme-binding motifs (e.g. five heme-binding motifs for NrfAs). The common heme-binding motif (CX₂CH) and also other less common heme-binding motifs (CX₂CK, CX₃CH, CX₄CH, CX₁₁CH, CX₁₅CH, and CX₁₇CH) were used in the filtering steps. Collected sequences for each protein group were aligned using MAFFT 7 (Kato and Standley 2013). Using MEGA 7 platform (Kumar et al. 2016), overall mean distances were individually computed using the *p*-distance model (Nei and Kumar 2000), uniform rates, and pairwise deletion methods. Standard errors were calculated using Bootstrap method with 1,000 replications (supplementary table 5, Supplementary Material online). CD-HIT (Li et al. 2001; Huang et al. 2010) was used to cluster highly homologous sequences in order to reduce the size of the data set but maintain the overall diversity. A total of 25 representative sequences without reference sequences (that were previously used as queries) were selected for each group. Reference sequences that contain 3D structures were added separately.

Sequence Alignments

For phylogenetic analyses, two MSAs were performed. One MSA contained protein sequences for all targeted MHC except NrfA (c_{554} + MSA) and the second MSA contained all MHCs except cyt c_{554} (NrfA+ MSA). MSA are provided in supplementary files, Supplementary Material online. Globally, protein sequences that included examples of reported 3D structures were aligned using PROMALS 3D (Pei et al. 2008) with default parameters. This program integrates sequence information derived from predicted secondary structure, profile-profile Hidden Markov Models, and structural information derived from sequence-structure and structure-structure alignments as constraints for consistency-based progressive alignments. In addition,

user-defined constraints of alignable heme-binding motifs and other important and identified structural elements, for example, distal axial ligands and active-site residues were used as anchors for the alignment. This MSA was used as a structural reference alignment in MAFFT 7 (Kato and Standley 2013) to align the sequences retrieved from NCBI (see Sequence Collection and Clustering). The L-INS-I algorithm was used with the “leave gappy regions” option. Resulting MSA was inspected in MEGA 7 and manual refinement was performed when necessary. Low-quality positions containing >75% gaps were filtered with trimAl version 1.3 (Capella-Gutiérrez et al. 2009) within the Phylemon 2.0 platform (Sánchez et al. 2011). For the pairwise sequence alignment of OcwA and OmhA, the Needleman–Wunsch algorithm (Needleman and Wunsch 1970) was used and represented in ESPript 3.0 (Robert and Gouet 2014).

Phylogenetic Analysis

As NrfA and cyt c_{554} align at opposing ends of the octa-, nona-, and undecaheme MHC, phylogenetic inference was performed using two MSA in separate (cyt c_{554} + and NrfA+ MSA; see Sequence Alignments) generating two phylogenetic trees. Each tree was reconstructed using maximum likelihood and Bayesian methods. Cyt c_{554} + MSA contained 749 positions and NrfA+ MSA contained 793 positions. For maximum likelihood inference, phylogeny was reconstructed using IQ-tree (version 2.1.2 COVID-edition built October 22, 2020; Minh et al. 2020) on XSEDE (Townes et al. 2014) and CIPRES (Miller et al. 2010) platforms. For model selection, ModelFinder (Kalyanamoorthy et al. 2017) method was selected. The best model according to Bayesian information criterion (BIC) was WAG+R8 and WAG+R7, for the cyt c_{554} + and NrfA+ MSAs, respectively. For generation of branch support values, ultra-fast bootstrap (Hoang et al. 2018) and SH-aLRT (Guindon et al. 2010) statistical methods were used. Confidence values were based on 1,000 replications for each method. For Bayesian inference, phylogeny was reconstructed using MrBayes (version 3.2.7a x86_64; Ronquist et al. 2012) on XSEDE (Townes et al. 2014) and CIPRES (Miller et al. 2010) platforms. As MrBayes does not include the FreeRate model for heterogeneity across sites available at the ModelFinder (Kalyanamoorthy et al. 2017; Minh et al. 2020), the best model was selected using the smart model selection (Lefort et al. 2017) at the ATGC: Montpellier Bioinformatics Platform (<http://www.atgc-montpellier.fr/sms/>). The best model according to BIC was WAG+G+I for both cyt c_{554} + and NrfA+ MSA. Furthermore, best-fit parameters for cyt c_{554} + MSA were: proportion of invariable sites of 0.035; number of substitution rate categories of 4 and gamma shape parameter of 1.738. For NrfA+ MSA best-fit parameters were proportion of invariable sites of 0.030; number of substitution rate categories of 4; gamma shape parameter of 1.993. Analysis ran using two independent runs with 12 Metropolis-Coupled Markov chain Monte Carlo each for 2 million generations,

sampling from the posterior distribution every 5000 generations. Chain convergence was assessed using Tracer version 1.7.2 (Rambaut et al. 2018). Minimum estimated sample size (ESS) and potential scale reduction factor (PSRF; Gelman and Rubin 1992) values were within acceptable values. ESS values were higher than 200 and PSRF values were between 1.000 and 1.001. A majority-rule consensus tree with all compatible groups and posterior probabilities of the bipartitions were used to reconstruct the MHC phylogeny, after discarding the first 50% of the sampled trees as burn-in. Tree rooting was performed by the minimal ancestor deviation method (Tria et al. 2017). HGT was assessed by comparing the phylogenetic position from the different protein sequences and the classification of the belonging organisms based on NCBI taxonomy database (Federhen 2012). Mismatch in phyla and class positions were considered as an evidence for HGT. Taxonomic classes were assigned to each tip of the maximum likelihood and Bayesian trees previously constructed. Tree visualization and representation were performed using the Interactive Tree Of Life tool (Letunic and Bork 2019).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Financial support was provided by European EC Horizon2020 TIMB3 (Project 810856). Financial support was also provided by Project MOSTMICRO-ITQB with references UIDB/04612/2020 and UIDP/04612/2020. Fundação para a Ciência e a Tecnologia (FCT) Portugal is also acknowledged for project PTDC/BIA-BQM/4143/2021.

Data Availability

All relevant data are available within the article and supplementary files, [Supplementary Material](#) online. Upon request to the corresponding author, additional data can be provided.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* **215**:403–410. doi:10.1016/S0022-2836(05)80360-2
- Andreini C, Cavallaro G, Rosato A, Valasatava Y. 2013. MetalS2: a tool for the structural alignment of minimal functional sites in metal-binding proteins and nucleic acids. *J Chem Inf Model.* **53**:3064–3075. doi:10.1021/ci400459w
- Baker IR, Conley BE, Gralnick JA, Girguis PR. 2022. Evidence for horizontal and vertical transmission of mtr-mediated extracellular electron transfer among the bacteria. *mBio.* e0290421. doi:10.1128/mbio.02904-21
- Bergmann DJ, Hooper AB, Klotz MG. 2005. Structure and sequence conservation of hao cluster genes of autotrophic ammonia-oxidizing

- bacteria: evidence for their evolutionary history. *Appl Environ Microbiol.* **71**:5371–5382. doi:10.1128/AEM.71.9.5371-5382.2005
- Bewley KD, Ellis KE, Firer-Sherwood MA, Elliott SJ. 2013. Multi-heme proteins: nature's electronic multi-purpose tool. *Biochim Biophys Acta–Bioenerg.* **1827**:938–948. doi:10.1016/j.bbabo.2013.03.010
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* **25**:1972–1973. doi:10.1093/bioinformatics/btp348
- Carlson HK, Iavarone AT, Gorur A, Yeo BS, Tran R, Melnyk RA, Mathies RA, Auer M, Coates JD. 2012. Surface multiheme c-type cytochromes from *Thermincola potens* and implications for respiratory metal reduction by Gram-positive bacteria. *Proc Natl Acad Sci U S A.* **109**:1702–1707. doi:10.1073/pnas.1112905109
- Carrondo MA, Soares CM, Matias PM. 2006. Nine-heme cytochrome c. In: Messerschmidt A, Huber R, Wieghardt K, Cygler M, Bode W, editors. *Handbook of metalloproteins.* John Wiley & Sons, Ltd. p. 1–12. doi:10.1002/0470028637.met112
- Chothia C, Lesk AM. 1986. The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5**:823–826. doi:10.1002/j.1460-2075.1986.tb04288.x
- Clarke TA, Edwards MJ, Gates AJ, Hall A, White GF, Bradley J, Reardon CL, Shi L, Beliaev AS, Marshall MJ, et al. 2011. Structure of a bacterial cell surface decaheme electron conduit. *Proc Natl Acad Sci U S A.* **108**:9384–9389. doi:10.1073/pnas.1017200108
- Costa NL, Hermann B, Fourmond V, Faustino MM, Teixeira M, Einsle O, Paquete CM, Louro RO. 2019. How thermophilic gram-positive organisms perform extracellular electron transfer: characterization of the cell surface terminal reductase OcwA. *mBio.* **10**. doi:10.1128/mBio.01210-19
- Deng X, Dohmae N, Nealson KH, Hashimoto K, Okamoto A. 2018. Multi-heme cytochromes provide a pathway for survival in energy-limited environments. *Sci Adv* **4**(2):ea05682. doi:10.1126/sciadv.aao5682
- Dong R, Pan S, Peng Z, Zhang Y, Yang J. 2018. mTM-align: a server for fast protein structure database search and multiple protein structure alignment. *Nucleic Acids Res.* **46**:W380–W386. doi:10.1093/nar/gkx1013
- Dong R, Peng Z, Zhang Y, Yang J. 2018. mTM-align: an algorithm for fast and accurate multiple protein structure alignment. *Bioinformatics.* **34**:1719–1725. doi:10.1093/bioinformatics/btx828
- Edwards MJ, Fredrickson JK, Zachara JM, Richardson DJ, Clarke TA. 2012. Analysis of structural MtrC models based on homology with the crystal structure of MtrF. *Biochem Soc Trans.* **40**:1181–1185. doi:10.1042/BST20120132
- Edwards MJ, White GF, Butt JN, Richardson DJ, Clarke TA. 2020. The crystal structure of a biological insulated transmembrane molecular wire. *Cell.* **181**:665–673.e10. doi:10.1016/j.cell.2020.03.032
- Edwards MJ, White GF, Norman M, Tome-Fernandez A, Ainsworth E, Shi L, Fredrickson JK, Zachara JM, Butt JN, Richardson DJ, et al. 2015. Redox linked flavin sites in extracellular decaheme proteins involved in microbe-mineral electron transfer. *Sci Rep.* **5**:11677. doi:10.1038/srep11677
- Einsle O, Messerschmidt A, Stach P, Bourenkov GP, Bartunik HD, Huber R, Kroneck PMH. 1999. Structure of cytochrome c nitrite reductase. *Nature.* **400**:476–480. doi:10.1038/22802
- Federhen S. 2012. The NCBI Taxonomy database. *Nucleic Acids Res.* **40**:D136–D143. doi:10.1093/nar/gkr1178
- Ferousi C, Schmitz RA, Maalcke WJ, Lindhoud S, Versantvoort W, Jetten MSM, Reimann J, Kartal B. 2021. Characterization of a nitrite-reducing octaheme hydroxylamine oxidoreductase that lacks the tyrosine cross-link. *J Biol Chem.* **296**:100476. doi:10.1016/j.jbc.2021.100476
- Gavrilov SN, Lloyd JR, Kostrikin AI. 2012. Fe(III) oxide reduction by a Gram-positive thermophile: physiological mechanisms for dissimilatory reduction of poorly crystalline Fe(III) oxide by a thermophilic Gram-positive bacterium *Carboxydotherrmus ferrireducens.* *Geomicrobiol J.* **29**:804–819. doi:10.1080/01490451.2011.635755
- Gavrilov SN, Zavarzina DG, Elizarov IM, Tikhonova TV, Dergousova NI, Popov VO, Lloyd JR, Knight D, El-Naggar MY, Pirbadian S, et al. 2021. Novel extracellular electron transfer channels in a gram-positive thermophilic bacterium. *Front Microbiol.* **11**. doi:10.3389/fmicb.2020.597818
- Gelman A, Rubin DB. 1992. Inference from iterative simulation using multiple sequences. *Stat Sci.* **7**:457–472.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* **59**:307–321. doi:10.1093/sysbio/syq010
- Hermann B, Kern M, La Pietra L, Simon J, Einsle O. 2015. The octaheme MccA is a haem c-copper sulfite reductase. *Nature.* **520**:706–709. doi:10.1038/nature14109
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* **35**:518–522. doi:10.1093/molbev/msx281
- Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics.* **26**:680–682. doi:10.1093/bioinformatics/btq003
- Igarashi N, Moriyama H, Fujiwara T, Fukumori Y, Tanaka N. 1997. The 2.8 Å structure of hydroxylamine oxidoreductase from a nitrifying chemoautotrophic bacterium, *Nitrosomonas europaea.* *Nat Struct Biol.* **4**:276. doi:10.1038/nsb0497-276
- Illergård K, Ardell DH, Elofsson A. 2009. Structure is three to ten times more conserved than sequence—a study of structural response in protein cores. *Proteins.* **77**:499–508. doi:10.1002/prot.22458
- Ingles-Prieto A, Ibarra-Molero B, Delgado-Delgado A, Perez-Jimenez R, Fernandez JM, Gaucher EA, Sanchez-Ruiz JM, Gavira JA. 2013. Conservation of protein structure over four billion years. *Structure.* **21**:1690–1697. doi:10.1016/j.str.2013.06.020
- Iverson TM, Arciero DM, Hsu BT, Logan MS, Hooper AB, Rees DC. 1998. Heme packing motifs revealed by the crystal structure of the tetra-heme cytochrome *c*₅₅₄ from *Nitrosomonas europaea.* *Nat Struct Biol.* **5**:1005–1012. doi:10.1038/2975
- Johnson CM, Beard BL, Roden EE. 2008. The iron isotope fingerprints of redox and biogeochemical cycling in modern and ancient earth. *Annu Rev Earth Planet Sci.* **36**:457–493. doi:10.1146/annurev.earth.36.031207.124139
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* **14**:587–589. doi:10.1038/nmeth.4285
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* **30**:772. doi:10.1093/molbev/mst010
- Kern M, Klotz MG, Simon J. 2011. The *Wolinella succinogenes* mcc gene cluster encodes an unconventional respiratory sulphite reduction system. *Mol Microbiol.* **82**:1515–1530. doi:10.1111/j.1365-2958.2011.07906.x
- Klotz MG, Schmid MC, Strous M, op den Camp HJM, Jetten MSM, Hooper AB. 2008. Evolution of an octaheme cytochrome c protein family that is key to aerobic and anaerobic ammonia oxidation by bacteria. *Environ Microbiol.* **10**:3150–3163. doi:10.1111/j.1462-2920.2008.01733.x
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* **33**:1870–1874. doi:10.1093/molbev/msw054
- Kummerfeld S, Teichmann S. 2005. Relative rates of gene fusion and fission in multi-domain proteins. *Trends Genet.* **21**:25–30. doi:10.1016/j.tig.2004.11.007
- Lefort V, Longueville J-E, Gascuel O. 2017. SMS: smart model selection in PhyML. *Mol Biol Evol.* **34**:2422–2424. doi:10.1093/molbev/msx149
- Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**:W256–W259. doi:10.1093/nar/gkz239

- Leu AO, Cai C, Mcllroy SJ, Southam G, Orphan VJ, Yuan Z, Hu S, Tyson GW. 2020. Anaerobic methane oxidation coupled to manganese reduction by members of the Methanoperedenaceae. *ISME J.* **14**:1030–1041. doi:10.1038/s41396-020-0590-x
- Li W, Jaroszewski L, Godzik A. 2001. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **17**:282–283. doi:10.1093/bioinformatics/17.3.282
- Liu Y, Wang Z, Liu J, Levar C, Edwards MJ, Babauta JT, Kennedy DW, Shi Z, Beyenal H, Bond DR, et al. 2014. A trans-outer membrane porin-cytochrome protein complex for extracellular electron transfer by *Geobacter sulfurreducens* PCA. *Environ Microbiol Rep.* **6**:776–785. doi:10.1111/1758-2229.12204
- Mayfield JA, Dehner CA, DuBois JL. 2011. Recent advances in bacterial heme protein biochemistry. *Curr Opin Chem Biol.* **15**:260–266. doi:10.1016/j.cbpa.2011.02.002
- Miller MA, Pfeiffer W, Schwartz T. 2010, editors. *Creating the CIPRES Science Gateway for inference of large phylogenetic trees.* In: *2010 Gateway Computing Environments Workshop (GCE).* New Orleans, LA: IEEE. p. 1–8.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* **37**:1530–1534. doi:10.1093/molbev/msaa015
- Mowat CG, Rothery E, Miles CS, Mclver L, Doherty MK, Drewette K, Taylor P, Walkinshaw MD, Chapman SK, Reid GA. 2004. Octaheme tetrahydroxymethyl reductase is a respiratory enzyme with novel heme ligation. *Nat Struct Mol Biol.* **11**:1023–1024. doi:10.1038/nsmb827
- Needleman SB, Wunsch CD. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol.* **48**:443–453. doi:10.1016/0022-2836(70)90057-4
- Nei M, Kumar S. 2000. *Molecular evolution and phylogenetics.* Oxford, New York: Oxford University Press.
- O’Leary NA, Wright MW, Brister JR, Ciufu S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**:D733–D745. doi:10.1093/nar/gkv1189
- Paquete CM, Rusconi G, Silva AV, Soares R, Louro RO. 2019. Chapter Three—a brief survey of the “cytochrome.” In: Poole RK, editor. *Advances in microbial physiology.* Vol. 75. Academic Press. (Advances in Microbial Physiology). p. 69–135 doi:10.1016/b.s.ampbs.2019.07.005
- Parey K, Fielding AJ, Sörgel M, Rachel R, Huber H, Ziegler C, Rajendran C. 2016. In meso crystal structure of a novel membrane-associated octaheme cytochrome *c* from the Crenarchaeon *Ignicoccus hospitalis*. *FEBS J.* **283**:3807–3820. doi:10.1111/febs.13870
- Pasek S, Risler J-L, Brézellec P. 2006. Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. *Bioinformatics.* **22**:1418–1423. doi:10.1093/bioinformatics/btl135
- Pei J, Kim B-H, Grishin NV. 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**:2295–2300. doi:10.1093/nar/gkn072
- Pereira IC, Xavier AV. 2006. Multi-heme cytochromes & enzymes. In: King RB, Crabtree RH, Lukehart CM, Atwood DA, Scott RA, editors. *Encyclopedia of inorganic chemistry.* Chichester, UK: John Wiley & Sons, Ltd. p. 3360–3376.
- Pereira L, Saraiva IH, Oliveira ASF, Soares CM, Louro RO, Frazão C. 2017. Molecular structure of FoxE, the putative iron oxidase of *Rhodospirillum rubrum* SW2. *Biochim Biophys Acta Bioenerg.* **1858**:847–853. doi:10.1016/j.bbabi.2017.07.002
- Pokkuluri PR, Londer YY, Duke NEC, Pessanha M, Yang X, Orshonsky V, Orshonsky L, Erickson J, Zagayanskiy Y, Salgueiro CA, et al. 2011. Structure of a novel dodecaheme cytochrome *c* from *Geobacter sulfurreducens* reveals an extended 12 nm protein with interacting hemes. *J Struct Biol.* **174**:223–233. doi:10.1016/j.jsb.2010.11.022
- Polyakov KM, Boyko KM, Tikhonova TV, Slutsky A, Antipov AN, Zvyagilskaya RA, Popov AN, Bourenkov GP, Lamzin VS, Popov VO. 2009. High-resolution structural analysis of a novel octaheme cytochrome *c* nitrite reductase from the haloalkaliphilic bacterium *Thioalkalivibrio nitratireducens*. *J Mol Biol.* **389**:846–862. doi:10.1016/j.jmb.2009.04.037
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst Biol.* **67**:901–904. doi:10.1093/sysbio/syy032
- Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**:W320–W324. doi:10.1093/nar/gku316
- Roldán MD, Sears HJ, Cheesman MR, Ferguson SJ, Thomson AJ, Berks BC, Richardson DJ. 1998. Spectroscopic characterization of a novel multiheme *c*-type cytochrome widely implicated in bacterial electron transport. *J Biol Chem.* **273**:28785–28790. doi:10.1074/jbc.273.44.28785
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* **61**:539–542. doi:10.1093/sysbio/sys029
- Rosato A, Valasatava Y, Andreini C. 2016. Minimal functional sites in metalloproteins and their usage in structural bioinformatics. *Int J Mol Sci.* **17**:E671. doi:10.3390/ijms17050671
- Sánchez R, Serra F, Tarraga J, Medina I, Carbonell J, Pulido L, de María A, Capella-Gutiérrez S, Huerta-Cepas J, Gabaldón T, et al. 2011. Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Res.* **39**:W470–W474. doi:10.1093/nar/gkr408
- Santos-Silva T, Dias JM, Dolla A, Durand M-C, Gonçalves LL, Lampreia J, Moura I, Romão MJ. 2007. Crystal structure of the 16 heme cytochrome from *Desulfovibrio gigas*: a glycosylated protein in a sulphate-reducing bacterium. *J Mol Biol.* **370**:659–673. doi:10.1016/j.jmb.2007.04.055
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**:2498–2504. doi:10.1101/gr.1239303
- Sharma S, Cavallaro G, Rosato A. 2010. A systematic investigation of multiheme *c*-type cytochromes in prokaryotes. *J Biol Inorg Chem.* **15**:559–571. doi:10.1007/s00775-010-0623-4
- Simon J, Kern M, Hermann B, Einsle O, Butt JN. 2011. Physiological function and catalytic versatility of bacterial multiheme cytochromes *c* involved in nitrogen and sulfur cycling. *Biochem Soc Trans.* **39**:1864–1870. doi:10.1042/BST20110713
- Simon J, Klotz MG. 2013. Diversity and evolution of bioenergetic systems involved in microbial nitrogen compound transformations. *Biochim Biophys Acta.* **1827**:114–135. doi:10.1016/j.bbabi.2012.07.005
- Stüeken EE, Kipp MA, Koehler MC, Buick R. 2016. The evolution of Earth’s biogeochemical nitrogen cycle. *Earth-Sci Rev.* **160**:220–239. doi:10.1016/j.earscirev.2016.07.007
- Tikhonova T, Tikhonov A, Trofimov A, Polyakov K, Boyko K, Cherkashin E, Rakitina T, Sorokin D, Popov V. 2012. Comparative structural and functional analysis of two octaheme nitrite reductases from closely related *Thioalkalivibrio* species. *FEBS J.* **279**:4052–4061. doi:10.1111/j.1742-4658.2012.08811.x
- Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, Hazlewood V, Lathrop S, Lifka D, Peterson GD, et al. 2014. XSEDE: accelerating scientific discovery. *Comput Sci Eng.* **16**:62–74. doi:10.1109/MCSE.2014.80
- Tria FDK, Landan G, Dagan T. 2017. Phylogenetic rooting using minimal ancestor deviation. *Nat Ecol Evol.* **1**:193. doi:10.1038/s41559-017-0193
- Valasatava Y, Andreini C, Rosato A. 2015. Hidden relationships between metalloproteins unveiled by structural comparison of their metal sites. *Sci Rep.* **5**:9486. doi:10.1038/srep09486
- Vargas M, Kashefi K, Blunt-Harris E, Lovley D. 1998. Microbiological evidence for Fe(III) reduction on early Earth. *Nature.* **395**:65–67. doi:10.1038/25720

- Wang F, Gu Y, O'Brien JP, Yi SM, Yalcin SE, Srikanth V, Shen C, Vu D, Ing NL, Hochbaum AI, *et al.* 2019. Structure of microbial nanowires reveals stacked hemes that transport electrons over micrometers. *Cell*. **177**:361–369.e10. doi:10.1016/j.cell.2019.03.029
- Weiner J 3rd, Bornberg-Bauer E. 2006. Evolution of circular permutations in multidomain proteins. *Mol Biol Evol*. **23**:734–743. doi:10.1093/molbev/msj091
- Welsh A, Chee-Sanford JC, Connor LM, Löffler FE, Sanford RA. 2014. Refined NrfA phylogeny improves PCR-based nrfA gene detection. *Appl Environ Microbiol*. **80**:2110–2119. doi:10.1128/AEM.03443-13
- Zhong C, Han M, Yu S, Yang P, Li H, Ning K. 2018. Pan-genome analyses of 24 *Shewanella* strains re-emphasize the diversification of their functions yet evolutionary dynamics of metal-reducing pathway. *Biotechnol Biofuels*. **11**:193. doi:10.1186/s13068-018-1201-1