

Conserved evolutionary units in the heme-copper oxidase superfamily revealed by novel homologous protein families

Jimin Pei,^{1†} Wenlin Li,^{2,3†} Lisa N. Kinch,¹ and Nick V. Grishin^{1,2,3*}

¹Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75390

²Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, Texas 75390

³Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75390

Received 19 May 2014; Accepted 11 June 2014

DOI: 10.1002/pro.2503

Published online 16 June 2014 proteinscience.org

Abstract: The heme-copper oxidase (HCO) superfamily includes HCOs in aerobic respiratory chains and nitric oxide reductases (NORs) in the denitrification pathway. The HCO/NOR catalytic subunit has a core structure consisting of 12 transmembrane helices (TMHs) arranged in three-fold rotational pseudosymmetry, with six conserved histidines for heme and metal binding. Using sensitive sequence similarity searches, we detected a number of novel HCO/NOR homologs and named them HCO Homology (HCOH) proteins. Several HCOH families possess only four TMHs that exhibit the most pronounced similarity to the last four TMHs (TMHs 9–12) of HCOs/NORs. Encoded by independent genes, four-TMH HCOH proteins represent a single evolutionary unit (EU) that relates to each of the three homologous EUs of HCOs/NORs comprising TMHs 1–4, TMHs 5–8, and TMHs 9–12. Single-EU HCOH proteins could form homotrimers or heterotrimers to maintain the general structure and ligand-binding sites defined by the HCO/NOR catalytic subunit fold. The remaining HCOH families, including NnrS, have 12-TMHs and three EUs. Most three-EU HCOH proteins possess two conserved histidines and could bind a single heme. Limited experimental studies and genomic context analysis suggest that many HCOH proteins could function in the denitrification pathway and in detoxification of reactive molecules such as nitric oxide. HCO/NOR catalytic subunits exhibit remarkable structural similarity to the homotrimers of MAPEG (membrane-associated proteins in eicosanoid and glutathione metabolism) proteins. Gene duplication, fusion, and fission likely play important roles in the evolution of HCOs/NORs and HCOH proteins.

Keywords: Heme-copper oxidase superfamily; HCO homology proteins; denitrification; detoxification; MAPEG proteins; gene duplication; gene fusion and fission

Additional Supporting Information may be found in the online version of this article.

[†]J.P. and W.L. contributed equally to this work

Grant sponsor: National Institutes of Health; Grant number: GM094575; Grant sponsor: Welch Foundation; Grant number: I-1505.

*Correspondence to: Nick V. Grishin, Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, 6001 Forest Park Road, Dallas, TX 75390.
E-mail: grishin@chop.swmed.edu

Introduction

Aerobic respiration has evolved to use oxygen, which produces about 16 times more adenosine triphosphates (ATPs) than anaerobic respiration.¹ This advantage likely facilitated some critical evolutionary steps, such as the origin of eukaryotes and the increase of body size.^{2,3} Heme-copper oxidases (HCOs) are membrane-bound enzyme complexes functioning in the terminal step of aerobic respiratory chains.^{4–6} They catalyze the reduction of dioxygen to water using electrons transferred from cytochrome c or a quinol derivative. The released

energy is coupled to the translocation of protons across the membrane to generate an electrochemical gradient that can be used for ATP synthesis. All HCOs possess a catalytic subunit, an integral membrane protein with 12 core transmembrane helices (TMHs). Six conserved histidines in the TMHs of the catalytic subunit coordinate three co-factors: a high-spin heme and a copper ion in the binuclear catalytic site, and an additional low-spin heme functioning in the electron transfer pathway.^{7,8} Two of the six histidines function as axial ligands to coordinate the low-spin heme, while the rest participate in the catalytic site, with three histidines positioning the copper on one side of the high-spin heme and one histidine serving as the axial ligand on the other side. The 12-TMH core structure of the HCO catalytic subunit displays three-fold rotational pseudosymmetry and distributes the two heme groups into two of the three proposed pseudosymmetric units,^{9,10} each of which consists of four TMHs.

HCOs from various organisms have been discovered. They differ in heme types, electron donors (such as cytochrome *c* and ubiquinol), proton transfer pathways, and subunit composition. Three major types of HCOs (A, B, and C) have been defined based on sequence and structural analyses.^{5,11} A-type HCOs include cytochrome *c* oxidases in mitochondria, and cytochrome *c* oxidases and quinol oxidases in many bacteria and some archaea.¹² B-type HCOs are mainly found in the archaeal phylum of Crenarchaeota and appear to use one proton pathway compared to two proton pathways in A-type HCOs. A- and B-type HCOs share a conserved tyrosine residue residing in the sixth TMH. This tyrosine is covalently linked to a copper-binding histidine and was proposed to donate a fourth electron to the binuclear center in the catalytic process.¹³ C-type HCOs, mainly from Proteobacteria, use a catalytic tyrosine residue located in a structurally different position (the seventh TMH) than that of A- and B-type HCOs.^{14,15}

Sequence and structural analyses revealed that the catalytic subunit of nitric oxide reductases (NORs) is homologous to that of the HCOs.^{6,16} NOR catalytic subunit also has 12 core TMHs sharing the same topology of HCO catalytic subunit and binds two hemes and a non-heme iron (instead of copper in HCOs) in a similar fashion by using six conserved histidines. The HCO superfamily thus includes both HCOs and NORs. NORs catalyze the reduction of nitric oxide (NO) to nitrous oxide (N₂O) with the help of the heme groups and the non-heme iron in the denitrification pathway of the nitrogen cycle.^{6,17,18} As NO is a toxic reactive agent, NORs in some pathogenic bacteria also play important roles in detoxifying exogenous NO generated by hosts.¹⁹ Besides substrate preference, NORs differ from HCOs in that they do not translocate protons across

the membrane and do not have a catalytic tyrosine due to fewer electrons required in one catalytic cycle. Two major subgroups of NORs have been described: the cytochrome *c*-dependent cNOR and the quinol-dependent qNOR.²⁰ NORs appear to be more closely related to C-type HCOs than A- and B-type HCOs in terms of sequence similarity and subunit composition.²¹

Although NORs and the three types of HCOs each form well-separated clades in the phylogeny reconstructed for the HCO superfamily, the position of the root remains controversial. Different evolutionary scenarios have been proposed for the origin and evolutionary order of HCOs and NORs. Several studies^{18,22–25} suggested that NORs may be more ancient than HCOs, consistent with the assumption that aerobic respiration evolved from denitrification after the emergence of atmospheric oxygen. Other researchers have proposed that the widely distributed A-type HCOs were present before the split of bacteria and archaea and are ancestors to B- and C-type HCOs and NORs.¹² These hypotheses, still in debate, explain how the 12 core TMHs developed into various types of oxidases. However, the origin of this pseudosymmetric helical architecture, an ancient event, is rarely discussed.

In this study, we used sensitive sequence similarity search methods such as transitive PSI-BLAST²⁶ searches and HHpred²⁷ to detect proteins homologous to the catalytic subunits of the HCO superfamily members. We called the newly found homologs HCO homology (HCOH) proteins. Interestingly, we discovered HCOH proteins with only four TMHs. These four-TMH proteins exhibit the highest similarity to the last four TMHs of HCOs (TMHs 9–12). They are considered to correspond to one evolutionary unit (EU) and are called single-EU HCOH proteins. Single-EU HCOH proteins may form homotrimers or heterotrimers to maintain the general structure and the ligand-binding sites defined by the fold of HCO/NOR catalytic subunits. HCO/NOR catalytic subunits are proposed to contain three homologous EUs made of TMHs 1–4, TMHs 5–8, and TMHs 9–12. We also discovered several groups of 12-TMH HCOH proteins that, like HCOs/NORs, contain three EUs. The majority of these three-EU HCOH proteins possess two conserved histidines that are predicted to bind a single heme. Most of the newly found remote homologs of HCOs/NORs are hypothetical proteins without experimental characterization. Only two of the seven major groups of HCOH proteins have been defined in current domain databases (DUF2871 and NnrS). Limited experimental studies and genomic context analysis suggest that they could function in the denitrification pathway and in the detoxification of reactive agents such as NO. Remarkably, the structural core of the three-EU assembly of HCOs/NORs resembles

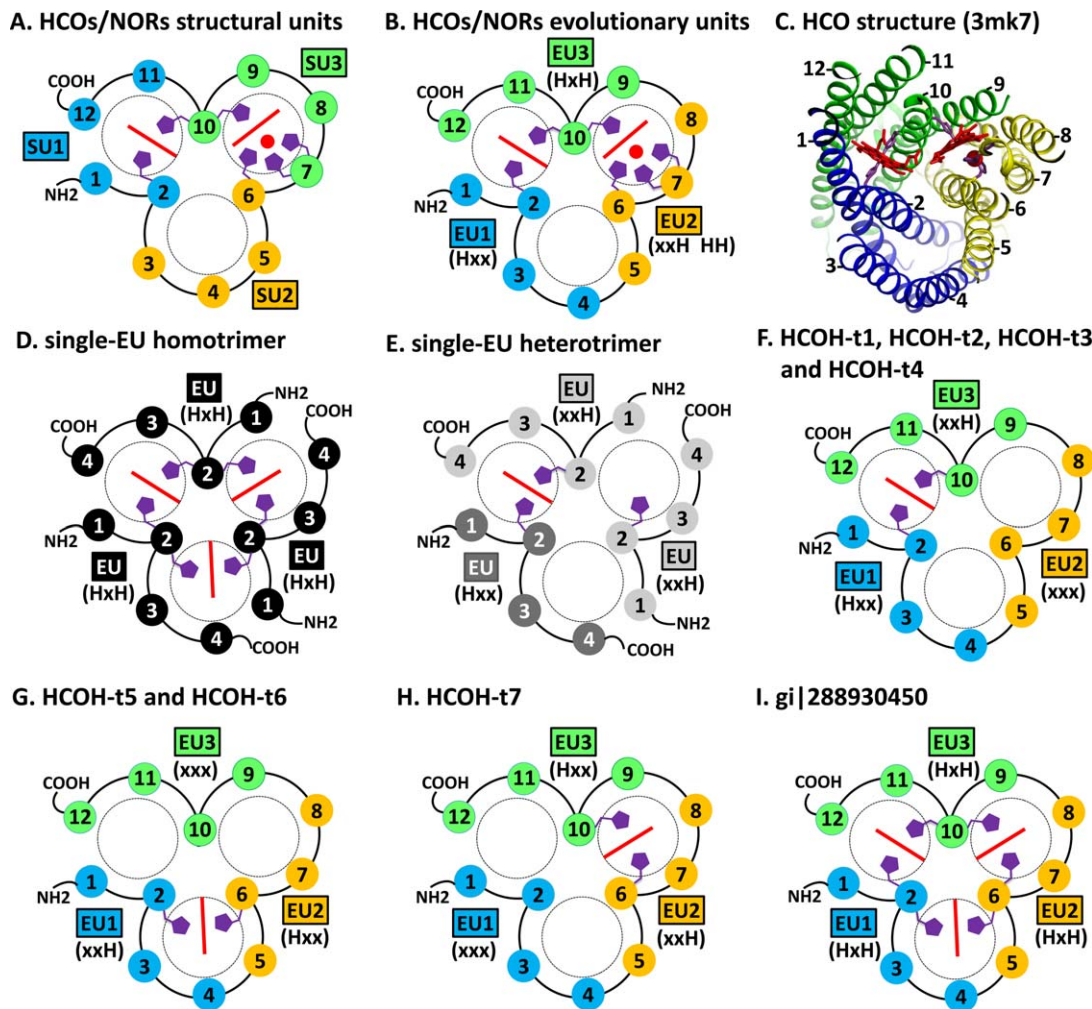


Figure 1. TMH topology diagrams of HCOs/NORs and HCOH proteins. TMHs are shown in numbered circles. TMHs in the same structure unit (marked by SU) or the same evolutionary unit (marked by EU) are filled with the same color. The histidine patterns of the EUs are shown in parentheses. Hemes are shown as red lines. Copper or non-heme irons in HCOs/NORs are shown as red spheres. Conserved histidine sidechains are shown in purple. N- and C-termini of each modeled protein are marked by NH2 and COOH, respectively. **(A).** Previously defined structural units of HCOs/NORs. **(B).** Newly defined EUs of HCOs/NORs. **(C).** Structure of a C-type HCO catalytic subunit (pdb: 3mk7). **(D).** Model of homotrimer for single-EU proteins with the HxH motif. **(E).** Model of heterotrimer for single-EU proteins with xxH and Hxx motifs. **(F).** Model for HCOH-t1, HCOH-t2, HCOH-t3 and HCOH-t4. **(G).** Model for HCOH-t5 and HCOH-t6. **(H).** Model for HCOH-t7. **(I).** Model for the protein (gi|288930450) with the HxH motif in all three EUs.

that of a diverse family of trimeric membrane-associated proteins in eicosanoid and glutathione metabolism (MAPEG).^{28,29} We propose the potential evolutionary scenarios linking existing families, as well as the early evolutionary events of HCOs/NORs in aerobic respiration.

Results and Discussion

Transitive PSI-BLAST²⁶ searches (see Materials and Methods) and HHpred²⁷ were used to detect proteins homologous to the catalytic subunits of the HCO superfamily members. A number of remote homologs of HCOs/NORs with different patterns of conserved histidines were discovered. We called these newly found superfamily members HCOH proteins and divided them into groups based on the number of

TMHs in the homologous regions, patterns of conserved histidines, and the CLANS³⁰ sequence clustering results.

Four-TMH proteins homologous to HCOs/NORs help define EUs

The catalytic subunits of HCOs/NORs exhibit an approximate three-fold structural symmetry and are considered as a result of duplications of four-TMH units.^{9,10} In previous structure studies, three pseudosymmetric structural units (SUs) have been defined as TMHs 11/12/1/2, TMHs 3/4/5/6, and TMHs 7/8/9/10 [Fig. 1(A)] (TMHs 1–12 correspond to previously defined TMHs I–XII).^{9,10} We detected a set of four-TMH proteins homologous to the catalytic subunits of HCOs/NORs. Most of these four-

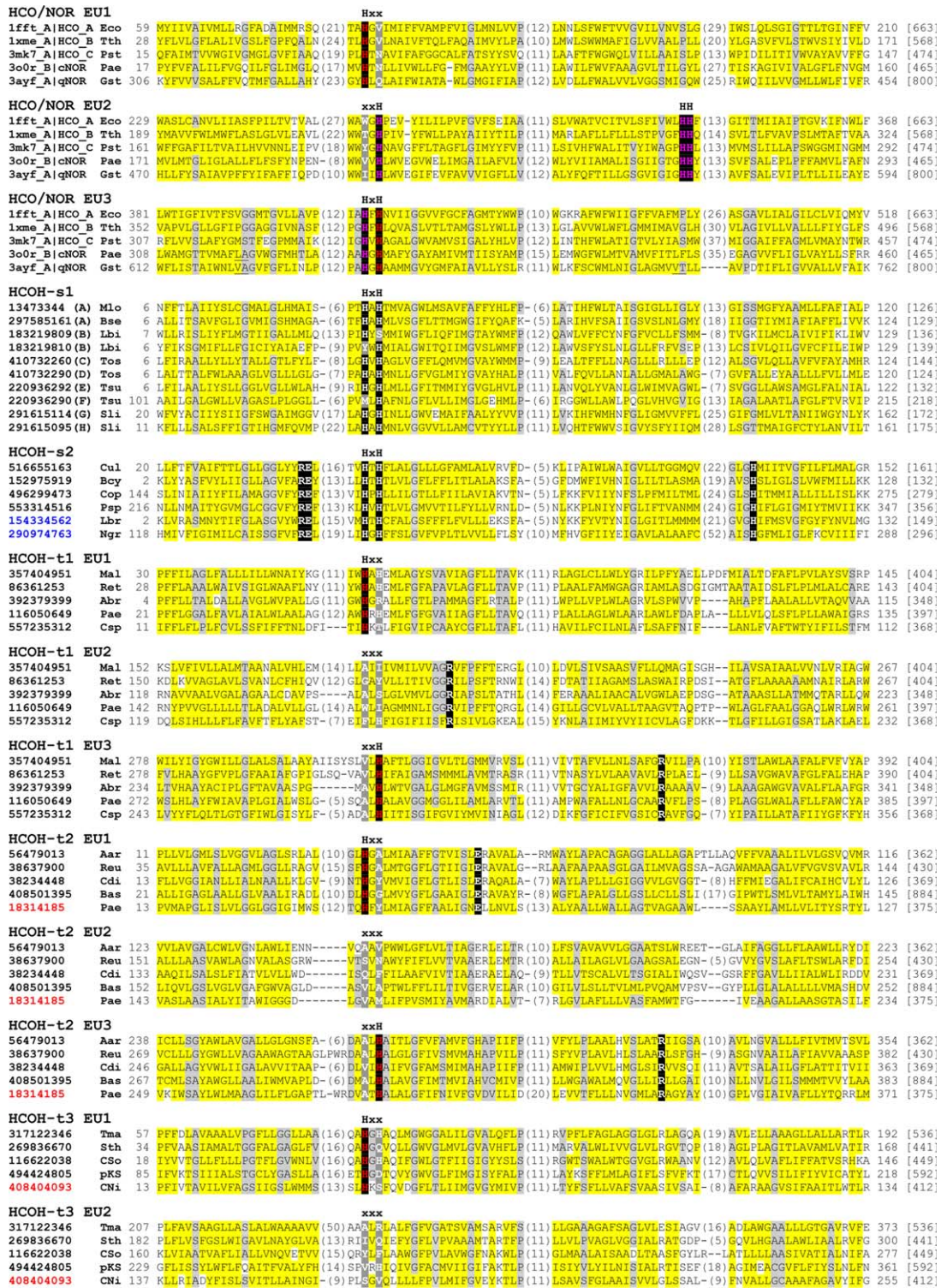


Figure 2.

TMH proteins possess the HxH motif at the beginning of the second TMH. These proteins exhibit the highest sequence similarity to the last four TMHs of HCOs/NORs (TMHs 9/10/11/12). The HxH motif of these four-TMH proteins aligns to the HxH motif in the tenth TMH of the catalytic subunits of HCOs/NORs. We consider that these four TMHs correspond to an evolutionarily conserved unit and

define them as one EU. Each of the four-TMH HCOH proteins possesses one EU and is thus called a single-EU HCOH protein. The other hand, HCOs/NORs contain three EUs: TMHs 1/2/3/4 (EU1), TMHs 5/6/7/8 (EU2), and TMHs 9/10/11/12 (EU3) [Fig. 1(B,C)].

The second TMH in each of the three EUs in HCOs/NORs harbors conserved histidine(s) for heme

HCOH-t3 EU3		xxH	
317122346	Tma 393	VAALLAYAWALVAALVLRGLH(13)DAVAVAGGFMTELLIMAVAFTEP(11)WVVAAYVLAALAAYRVTEPG(16)ALAGAAAGAAIVLLVVTERR	522 [536];
26983670	Sch 318	LRLLSAYNWLRAAGVSMGLTGLH(11)DMFAVAGAGFATLLILLVSGSHMP(11)LPWVTLRAGMLVAVLRVAVGL(9)GVAGVGLVAVLWVFMVNLHG	438 [441];
116622038	Cse 293	SPVPSAYWLLTAAAGCVRAAD-(7)GASRAALTGVFLSMTVFAGCVLP(12)IMTSLTPTAMGCLPRAASE(14)FVSAATELTAFTVATATGI	415 [449];
494424805	pRS 377	KETRAALVFLVFSYTALTYTY(13)GASRAALFIFGFSNMILCCASSMP(11)LLNAFTLIVNGCLPRAVVSQ(13)GISGFIIVAMFCFGINAWK	503 [592];
408404093	CNI 278	QHSRLAFLVFLYAGTAVATAFNV-(6)DAVAVTAIGFVGTITIALYLDEP(11)FNSVPLVLLTALAVRAGDM(15)MSSGWLVAALSTVFVAMVHR	399 [412];
HCOH-t4 EU1		HxH	
343082909	Cma 3	SWLLISLINFIAALMLGLRAA(14)HGSRAVLLGWLYLGLVFLIHGRFL(11)RLFWLQFSVVGMMLSFFLQ(6)FFSSHLIVLSYFVVYVWVKD	123 [404];
390943549	Fba 23	TMTWVAFFVFCACAIPLGLARVYF(14)HRSRAVLLGWGVLVLTGLITFVSYV(10)KILITLIIANLGMMSFFPQ(6)AFSTHIVMNSVYVFAFFVD	142 [607];
255536242	Fba 4	FWLKFSVFNELLVALLGVLRYK(14)EASRAFAFYGWIHTVLYLVLTNLYL(12)LIIISNLIASFAMLTGFIVGG(6)AASTHALLFVFCVFFVRD	125 [408];
337749376	Pmu 19	SWRTPFLLYEWTAAATGVLRAM(14)HRSRAVLLGWVATLMLLAAARFL(11)VNNGLTQAAVLMFAAFVQ(6)ALSTVHLIILSYFVYVWVRW	139 [439];
495910933	Bar 3	KLWLTCLNLFIAAINGLITRYS(13)HRSRAVHAGVLRVAVLITLITLFFAIFYV-(8)KLFWITQLAVIGMMVSFFPQ(6)SFSTHLLSYFVYVWVKD	119 [403];
284036739	Sli 11	NAMWTALGWVVAAGVGLRVO(14)HRSRAVLLGWVNFGLFLALVSAG-(9)KINLGFQASIFGLMVFFPIQ(6)VASTVHVFYSYVWMCNLRD	129 [411];
HCOH-t4 EU2		xxxx	
343082909	Cma 132	LLLRRTALLPFLISTCGVAVAV(15)QYVLEFQFNWGLLFGVLLALINDFK(10)FYFLMCLSQVTLFGLILFWAY-(7)VNFIVGLGALLGALLSR	253 [404];
390943549	Fba 152	NLLKLSIIWMISSIGLWAIAPV(15)QWLELQINGWVVFVFLGILLGYME(11)TLFALHLSLFTLYALSVAWST-(7)INLGVLLQVIAYFIIIKPI	274 [607];
255536242	Fba 134	IFWVAGLFFFAISSAGVFNLAYM(15)YVLEFQYNGFFIFSCIGLILYSIK(13)MFWLMFPGCLIGFLSVLWMK-(7)LIVVASVSCTVGSMLNEV	258 [408];
337749376	Pmu 150	RFQAALACMALSSAGPWAALW(15)YVLEFQYNGWLTGLFVAVLYALE(13)HLLYLGMLPPAPFLSMLMWAG-(7)TAAAGLMLQAAQAFAVLLW	274 [439];
495910933	Bar 128	KLWLASLLEMVISTIGVNLGPA(15)QWLELQFNWGLFVAVLTAHQIQO-(9)FFYLIISTILSFALPIQWA-(7)INGFALPLAVLAVVLRIT	248 [403];
284036739	Sli 138	RLVFWLFFALSTLGPVAVGLI(15)YVLEFQYNGWFMFQCLALLVRLW(13)FVTALSAFGLTALSALWTD-(7)VGGIALLCAGAGGWLLWL	262 [411];
HCOH-t4 EU3		xxH	
343082909	Cma 267	SFLLLAIFVGVIRVVAQVLLIFP(13)IGYIYDNNMLGLFSAYLILYSPFKNO(10)VPIFFIFGIFGSELILFFQ(6)LAFSGFIPLGIAYVIYVFLK	395 [404];
390943549	Fba 290	SLLWIGITSLAKALQATLIIP(13)IGYIYDNNMLGLFSAYLILYSPFKNO(10)AVTFISFGFISELLILGQGT(16)FISSIFFPVGSLIIIAQWT	417 [607];
255536242	Fba 274	FVLIFVGFPAFAKATLQGSNIP(13)IAYELVLLMCIATFVLSQILATNY(9)GKLVLLIGFTINEAIIIGMT(15)LLPSSLILFVGLISLVFNKMI	400 [408];
337749376	Pmu 290	RLLRLSLLWAAKMGLEAASALP(13)VGYLEVLLGFVSLPFLACWLQQW(10)GVLLLAAGLGLENELLLFNGM(16)AAASLLFVSLVILGGAGK	418 [439];
495910933	Bar 264	MYOFALPCFIIKIVQFLSIVP(13)IGYIYDNNMLGLFSAYLILYSPFKNO(10)GKLVLLIGFTINEAIIIGMT(15)LLPSSLILFVGLISLVFNKMI	391 [403];
284036739	Sli 278	GVRLAFLSFAKLIILQVSSVP(13)IAYELVLLMCIATFVLSQILATNY(9)GKLVLLIGFTINEAIIIGMT(15)LLPSSLILFVGLISLVFNKMI	401 [411];
HCOH-t5 EU1		xxH	
39935138	Fpa 22	RFFVAAGVHVLLANGVLIAND(11)LAIVNLLTGVLITFVAVGAVQLP(11)PILASWLLIPIGAVVLIAGHA-(7)IQAASSTIGLIIFALAD	140 [427];
38637924	Reu 18	GCLMAPWLGAAGLLALFAGV(10)LAIVNLLTGVLITFVAVGAVQLP(11)PILASWLLIPIGAVVLIAGHA-(7)IQAASSTIGLIIFALAD	135 [445];
530610160	Gap 16	SFIMFAVLAFAVSOQLMLWAVPA(11)WAAAVHGLLGFALMTAMGAMQYQLVP(11)LGFQWVAFATGIVAFVSLA-(6)FLPGLLILGVLIVFQWAM	133 [413];
118474250	Cfe 16	AFYVIVFAFVSFLYKFSQD-(7)ATELHFLVNGVFTIIVKISYLQSS(11)GAVINVFAITLALIFLNGME(7)YLSLILFLLSILCYLL	330 [393];
389847617	Hme 20	RHLVVALGSLFVAIGLGVLD-(7)LAIVNLLTGVLITFVAVGAVQLP(11)PILASWLLIPIGAVVLIAGHA-(7)IQAASSTIGLIIFALAD	134 [455];
288930450	Fpl 5	KFLYSALVYLVAATFGLNLPGFYKPV(11)HRSRAVHAGVLRVAVLITLITLFFAIFYV-(8)KLFWITQLAVIGMMVSFFPQ(6)SFSTHLLSYFVYVWVKD	114 [379];
HCOH-t5 EU2		Hxxx	
39935138	Reu 152	AYGVAFASLIALVALGLALS(13)RANMGLGGFPMGLALGFVSHVLP(11)MAWGFAAAAAAVAGLTIGAL-(7)SAACVGVVAVGLHGLMHW	272 [427];
38637924	Fpa 149	TLAWIGAPLLLTAACGGVLAGL(12)ALHPLWGLGVLGVAALVPTAMLQIVP(11)LRWVPLAHVPLLAFTAVV(10)WTLLPVSVAVAVGFGLAG	271 [445];
530610160	Gap 142	LEVATSLFLLTAAGGMLAH(14)GLWVFLGCGWFTLLIMGFSYKMAP(13)RYVYALYGGIAVAFASLSP-(5)AVGAILMGFVAFAVHWIRA	266 [413];
118474250	Cfe 145	LCLFVSAIMFLMCLGMLLV(13)RHFHLEVE-GFIETVIGASSVLLP(11)LFYASFTLYISGFLIFWLLD----GALKVLLSAAVLAAILQWL	253 [393];
389847617	Hme 147	RHFALAGFELLVTFPGLLA(17)ESRAVAVGAVVTTIVGALYQLP(13)LQRVEVGYVGVLLLATGRL-(7)RVGGVFLVSLVLFVGVFR	272 [455];
288930450	Fpl 127	WFFVAALYYLAGITYALLAFLG-(6)RVAHPLAAGWIGLITFGGLYLEP(11)LAIVNLLTGVLITFVAVGAVQLP(11)PILASWLLIPIGAVVLIAGHA	239 [379];
HCOH-t5 EU3		xxxx	
39935138	Reu 286	VLRSSWVMTLTPLAALAAQG-(7)TLGGMALGGWMLTLVVLRIVP(24)PLKHAVCHLIAIPALGLAV-(7)AGAAVIGAVGALAFAYAL	413 [427];
38637924	Fpa 283	PLWPVACASLAAALAAAMPV-(8)WNLGSLALIGGVLPVPTAMLQIVP(27)RQRQAMLLAAYLMLLALPV-(6)PAGGLAFAANLWGLVQITL	413 [445];
530610160	Gap 262	SNTAVAIAGLAAHAAVIAIAGSGRWLGLHMLSTIGWIFGSIYGLYKIVP(27)AITIQCRLLWLAALAAVALA-(7)ATAQVLLAGLSAVAFATILA	409 [413];
118474250	Cfe 267	LVLSIFEFMGFSIFAYSLFNLN----LAVTFLPFGFIPFVAVIYKIM(27)YAFGLFINLAIFFIFFKFE----YLAQVFMVLSVILVNVMIN	382 [393];
389847617	Hme 287	RYVAAAPALAVWLLTFLPSWAN(10)PAIVNLLTGVLITFVAVGAVQLP(11)PILASWLLIPIGAVVLIAGHA-(7)IQAASSTIGLIIFALAD	420 [455];
288930450	Fpl 255	FVVALFVGLAGITLAPLNIFDL-TFHSRLLAGWIALTIVGAEYHIP(27)IGKFLYVSPGLVILLIPIQTS----TAGVLTFAFLAVFADWFA	375 [379];
HCOH-t6 EU1		xxH	
25026829	Cef 402	VIAPFLVGVVAATLGYRGRGVTV---VDVHPFFITGALTTIAYSTHFA(11)GVGLRVAIVNLMGLLIDRA-(8)DVSATAVAVLWQIAVNSVK	511 [784];
74318194	Tde 4	ALVLLALTVAFAALLASPHGR---AAVLEVPVFGALVLTGAVMTHFIP(11)LAATPLAAATAGALVVGSLVD-(7)HAAAGLGMATLAFVNSRG	110 [380];
257068095	Mgi 37	VVLNVALVTVLHIIQWIPQR---WLVVWVTLGILTISSIMWGQFAE(13)RQVRIWLLTAGLAVTLGML-(7)VLGGLIVSAAVVAALGA	146 [969];
145224073	Bfa 12	VVLGWLVAWMAATHPVIALSN---WLVVWVTLGILTISSIMWGQFAE(13)RQVRIWLLTAGLAVTLGML-(7)VLGGLIVSAAVVAALGA	120 [857];
317126498	Ica 12	FVWLVAIAFVSVLHPPFAYR---WLVVWVTLGILTISSIMWGQFAE(13)RQVRIWLLTAGLAVTLGML-(7)VLGGLIVSAAVVAALGA	121 [909];
HCOH-t6 EU2		Hxxx	
25026829	Cef 525	PFYLTAAAGFLVAVLLATLATRV-(7)AAVLEVPVFGALVLTGAVMTHFIP(11)LAATPLAAATAGALVVGSLVD-(7)HAAAGLGMATLAFVNSRG	641 [784];
74318194	Tde 123	AWYDAALVCLTLAALAAILLAGS(8)RNVVHNTIIGVNTIAGTAVLWLP(12)WLRADPLALGTTLLTAIGAA-(6)WLGALLMGVAVLHGRWRAR	238 [380];
257068095	Bfa 160	RAYGAACTLPAGLGAVMAMS(10)LAIVNLLTGVLITFVAVGAVQLP(11)PILASWLLIPIGAVVLIAGHA-(7)IQAASSTIGLIIFALAD	279 [969];
145224073	Mgi 134	HYVVAACLAIVAGLGAVMAMS(11)TABAVNLLFVWGLVVTGTLVTLMP(13)ASRRCLPALVLSVAVAGAI-(7)GVGLSFTAVGVFLVHWID	254 [857];
317126498	Ica 135	RYYVAAFLPVGATFVLLASG(16)VAIVNLLGWLITLGLTGLVTLMP(13)ASRRCLPALVLSVAVAGAI-(7)GVGLSFTAVGVFLVHWID	260 [909];
HCOH-t6 EU3		xxxx	
25026829	Cef 654	VSVAAGLLMVAATDAVILLV-(8)LPLFALLIGSGLLQLVTVLHLLP(19)AGGARVTLINLGAALLLDDVTGPARSAGLLILIGLIGHVITTR	774 [784];
74318194	Tde 251	PLGAAGWLTLSLGVAVADVP-(6)AQLFAFVAGFLPLVSGAASQLP(21)YGGARGLFVGGIAGGVGL----AVGLVLAVALVLLFLQLV	366 [380];
257068095	Bfa 298	LSISAGVAWFTVGLGLVVWR(15)QEVVFPVAGFLVLTGVLGVAYLLP(19)FAVPSVAVNVLVLAFLVLAGM(27)VLLSLLFAVLSVFEVLLV	448 [969];
145224073	Mgi 266	LSQLGVVWLAGSLVLSGLIT(11)AAQFALLAGFLVQLVGLSAYLIP(18)GAPSRVAVNVLVLAFLVLAGM(27)VLLSLLFAVLSVFEVLLV	388 [857];
317126498	Ica 272	LSQLGVVWLAGSLVLSGLIT(11)AAQFALLAGFLVQLVGLSAYLIP(18)GAPSRVAVNVLVLAFLVLAGM(27)VLLSLLFAVLSVFEVLLV	395 [909];
HCOH-t7 EU1		xxxx	
15605807	Aae 9	FFFALALLNLLSIFTRLSQDT----SLFVGLVGFVGLTIMGAMYIIP-(8)IPKVSYLFFLIVSFFEFTGN-TESSLFLFLGSLIIFHHLL	108 [378];
195953645	Hsp 9	LFFLLSILMLTIDILIKAFPTNN---INVPLIALYGFNLVIGAMYQIIP-(9)PKLSFTLGLSVLNGFIMFLM(10)CISNTLITFLIHLTK	118 [376];
389848681	Hme 11	RFLISAAFLVVRFGALVIGSR---RVEVTIALGPFVHVHVGAMLYLVE-(8)DRVPAALHVAVSQTLLTA(15)SAGVLLWCAGVAFVAAALLW	123 [408];
397775133	Nsp 15	AFVAVGIFVAVQAVAVDAGR---AAVLEVPVFGALVLTGAVMTHFIP(11)LAATPLAAATAGALVVGSLVD-(7)HAAAGLGMATLAFVNSRG	123 [405];
313116973	Hbo 7	RVVLSAVFLTGWQAVVVGIPR---RTEVWLGIRGFVHVIFGKAYLVE-(8)FPVAVQVFPVAVGTGLVA(11)AAGAILMCLIGITGIFFLAW	115 [388];
HCOH-t7 EU2		xxH	
15605807	Aae 118	RFLASMTYLVLSALFLYLFHFKGLFVQLSVRHLTVGSMLNAYGVELWIP(10)KGEKLFYAKVSTLALLLAFV-(7)ALAGLLEFGVALVFLVLYNE	228 [378];
195953645	Hsp 125	KFLSASLFTITNLEMLFFANFKINIFFTHEPFTIGVFINAVIGVFALVP(10)LANLWFWHTSASTILIASFY-(7)YFAGLEFEVIGFVFLYLV	235 [376];
389848681	Hme 151	RELPVSVFAYLAAGSYELARVSP-(9)PRVHMLAAGFVLLVTVLGLRFLAP-(8)KVTGTIVLITGAVGPALIFG-(7)WAGLLEAVAVFVGTGAVL	264 [408];
397775133	Nsp 151	VAVPFLVAYLLGSSALPLAAAG(10)PATHTLLAAGTAAVLLVFAICRLIP-(8)PILVSVVAAGIAGPALLAAG-(7)RVGALCATLAVAVFAVAVD	265 [405];
313116973	Hbo 143	GFVVALYLILGAYCTIAYVA(9)QVYVHLLAAGTAGVMMGLAFLMPE-(8)RWVGVVLTGAVGVPVLLAAG-(7)QLGALVAVAVGATNAT	256 [388];
HCOH-t7 EU3		Hxxx	
15605807	Aae 243	KIFLFAFLLPPLGMLLGFPSASH-(8)RFDLIIYGFCAIVGGLMLHLL(23)ERLQTFLEYSALLYALFLAV(10)STVYVLYWMLFLKIKHAF	373 [378];
195953645	Hsp 250	RFFPLGLGFLVGLVNSVASSKLEPPHDDWVYIGFLTVLGGIFHFM(25)QNVWKLFPGLFGLIFVVE-(5)LKPVGLVAFVAVSYLYAF	373 [376];
389848681	Hme 276	GSIILVCTVSQVGVAVGVAVG(13)QHEICNLGFLGLVLTGAVVQVFP(10)ERTARIVIGTLAVGLAVG(10)ERLARIVIGTLAVGLAVG(10)ERLARIVIGTLAVGLAVG	398 [408];
397775133	Nsp 279	LISAGCGALLINGLCFAFAPA-(8)DAVAVTAIGFVGTITIALYLDEP(11)FNSVPLVLLTALAVRAGDM(15)MSSGWLVAALSTVFVAMVHR	396 [405];
313116973	Hbo 268	YVLAGAGFGLVATTLGLLAFS-(7)RAVHPLAAGWIGLITFGGLYLEP(11)LAIVNLLTGVLITFVAVGAVQLP(11)PILASWLLIPIGAVVLIAGHA	384 [388];

Figure 2. Continued

or metal-binding, with characteristic two-residue motifs of Hxx, xxH, and HxH in EU1, EU2, and EU3, respectively (x: a variable residue) (Fig. 2). These motifs are homologous and occupy structurally equivalent positions in the superposition of EU1, EU2, and EU3. The third TMH of HCO/NOR EU2 additionally harbors a conserved HH motif

(Fig. 2). The histidine in the Hxx motif of EU1 and the second histidine in the HxH motif of EU3 coordinate the low-spin heme in the ligand-binding pocket between EU1 and EU3 [Fig. 1(B,C)]. The histidines in the xxH motif and the HH motif of EU2 as well as the first histidine in the HxH motif of EU3 contribute to the binding of high-spin heme and copper/

non-heme iron in the pocket between EU2 and EU3 [Fig. 1(B,C)].

Single-EU proteins with four TMHs may form homotrimers or heterotrimers

The single-EU proteins can be roughly divided into two major groups: HCOH-s1 (HCO homology proteins with a single EU, group 1) and HCOH-s2, according to CLANS sequence clustering results (Fig. 3) and sequence conservation.

HCOH-s1 (Fig. 3, red up triangles) consists of a main cluster (marked by A in Fig. 3) and several nearby small clusters (marked by letters B, C, D, E, F, G, and H in Fig. 3). HCOH-s1 proteins of the main cluster (cluster A) are mostly from Proteobacteria and Firmicutes (Supporting Information Fig. S1). To maintain the structural compactness in the general fold of HCOs/NORs, these single-domain proteins most likely assemble as trimers. They could form homotrimers, since genomes of the main cluster of HCOH-s1 proteins contain only one single-EU protein. As this cluster of HCOH-s1 contains the HxH motif, three symmetric heme-binding sites at the interfaces of EUs can be inferred for the homotrimers [Fig. 1(D)], similar to the fashion of coordination of the low-spin heme by HCOs/NORs.

The B cluster (Fig. 3) of HCOH-s1 consists of closely related proteins encoded by gene pairs that are chromosomal neighbors [see three gene structure examples in Fig. 4(A–C)]. Interestingly, two proteins from the same species have the xxH and Hxx motifs, respectively (e.g., gi|183219809 and gi|183219810 in Fig. 2, also see Supporting Information Fig. S1

for other B cluster proteins). These neighboring gene products could form heterotrimers, with the histidines contributing to the heme-binding site similar to the way the low-spin heme is coordinated in HCOs/NORs. Figure 1(E) depicts one possible way of forming a heterotrimer consisting of two proteins with the xxH motif and one protein with the Hxx motif. A single heme-binding site can be inferred for such a heterotrimer.

The remaining six small clusters (C–H) of the HCOH-s1 each have limited species distribution. HCOH-s1 proteins of clusters C and D contain the HxH motif and are from the Thermales order of the Deinococcus-Thermus phylum. In several species of the *Thermus* genus, a cluster C member and a cluster D member are products of genes not far away from each other, separated by gene clusters containing denitrification enzymes [see one example of Fig. 4(D)]. These proteins may form homotrimers or heterotrimers for genomes containing both C and D cluster members. On the other hand, two species (*Meiothermus silvanus* DSM 9946 and *Oceanithermus profundus* DSM 14977) have cluster C members and do not have cluster D members [see the example of Fig. 4(E)], suggesting that cluster C members can form homotrimers. HCOH-s1 proteins of clusters E and F, mostly from the *Thioalkalivibrio* genus, are also largely encoded by neighboring gene pairs [one example shown in Fig. 4(F)]. Although cluster E proteins have the HxH motif, cluster F proteins are characterized by the xxH motif (Supporting Information Fig. S1). Products of these neighboring gene pairs may form heterotrimers. HCOH-s1 proteins of

Figure 2. Multiple sequence alignment of EUs of HCOs/NORs and HCOH proteins. For HCOs/NORs, the sequences are denoted by their pdb id and chain id, followed by type name. For HCOH proteins, each sequence is denoted by its NCBI gene identification (gi) number. The cluster type for HCOH-s1 is shown in parentheses after gi numbers. Starting and ending residue numbers are shown before and after the sequences, respectively. Protein lengths are shown in brackets at the end. Positions with conserved motifs in the second TMHs of these EUs are marked with HxH, xxH, Hxx, or xxx. The conserved HH motif in the second EU of HCOs/NORs is also marked. Conserved histidines in these positions are in black background. For three-EU proteins, conserved histidines at the interface between EU1 and EU3 are colored red, those at the interface between EU2 and EU3 are colored orange, and those at the interface between EU1 and EU2 are colored magenta. Other family-specific conserved residues are also highlighted in black background. Non-charged residues in positions with mainly hydrophobic residues are shaded in yellow. Small residues (G,A,S,C,T,P, and V) in positions with mainly small residues are shaded in grey. Insertion regions are replaced by the number of inserted residues in parentheses or omitted in between underscored letters. Three-letter species name abbreviations shown before the residue starting numbers are as follows: Aae, *Aquifex aeolicus*; Aar, *Aromatoleum aromaticum*; Abr, *Azospirillum brasilense*; Bcy, *Bacillus cytotoxicus*; Bse, *Bacillus selenitireducens*; Bba, *Belliella baltica*; Bas, *Bifidobacterium asteroides*; Bar, *Bizionia argentinensis*; Bfa, *Brachybacterium faecium*; Cfe, *Campylobacter fetus*; Csp, *Campylobacter* sp.; Cni, *Candidatus Nitrososphaera gargensis*; Cso, *Candidatus Solibacter usitatus*; Cop, *Coprobacillus*; Cdi, *Corynebacterium diphtheriae*; Cef, *Corynebacterium efficiens*; Cul, *Corynebacterium ulceribovis*; Cma, *Cyclobacterium marinum*; Eco, *Escherichia coli*; Fpl, *Ferroglobus placidus*; Fba, *Flavobacteriaceae bacterium*; Gsp, *Geobacillus* sp.; Gst, *Geobacillus stearothermophilus*; Hme, *Haloferax mediterranei*; Hbo, *Halogeometricum borinquense*; Hsp, *Hydrogenobaculum* sp.; Ica, *Intrasporangium calvum*; Lbr, *Leishmania braziliensis*; Lbi, *Leptospira biflexa*; Mlo, *Mesorhizobium loti*; Mal, *Methylomicrobium alcaliphilum*; Mgi, *Mycobacterium gilvum*; Ngr, *Naegleria gruberi*; Nsp, *Natrinema* sp.; pKS, *planctomycete* KSU-1; Pmu, *Paenibacillus mucilaginosus*; Psp, *Peptoniphilus* sp.; Pae, *Pseudomonas aeruginosa*; Pst, *Pseudomonas stutzeri*; Pae, *Pyrobaculum aerophilum*; Reu, *Ralstonia eutropha*; Ret, *Rhizobium etli*; Rpa, *Rhodopseudomonas palustris*; Sli, *Sideroxydans lithotrophicus*; Sth, *Sphaerobacter thermophilus*; Sli, *Spirosoma linguale*; Tma, *Thermaerobacter marianensis*; Tos, *Thermus oshimai*; Tth, *Thermus thermophilus*; Tsu, *Thioalkalivibrio sulfidophilus*; Tde, *Thiobacillus denitrificans*. Gi numbers of bacterial, archaeal, and eukaryotic proteins are in black, red, and blue colors, respectively.

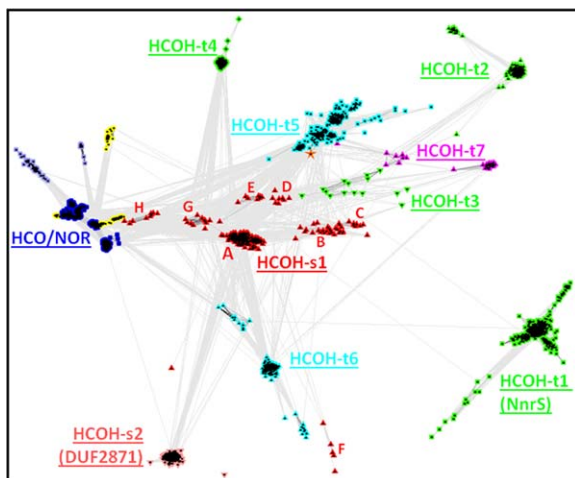


Figure 3. CLANS diagram of HCOs/NORs and HCOH proteins. Connections between proteins indicate BLAST P-values less than $1e-10$. Single-EU HCOH proteins are shown as red up triangles (HCOH-s1) and pink low down triangles (HCOH-s2). Members of HCOH-t1, HCOH-t2, HCOH-t3, and HCOH-t4, the four groups with the Hxx.xxx.xxH motif pattern, are shown in green squares, green up triangles, green down triangles, and green diamonds, respectively. HCOH-t5 and HCOH-t6 members, both with the xxH.Hxx.xxx motif pattern, are shown in cyan square and cyan up triangles, respectively. The sequence in the HCOH-t5 group with the HxH.HxH.HxH motif pattern is marked by an orange star. HCOH-t7 members with the xxx.xxH.Hxx motif pattern are shown in magenta. Underlined group names are shown. For the HCOH-s1 group, the A to H clusters are marked in red letters. HCOs/NORs are shown as blue dots. Two small groups of proteins shown as light blue dots are closely related to HCOs/NORs. They do not contain the HH motif in the third helix of EU2, while maintaining all the other conserved histidines (Hxx, xxH, and HxH in EU1, EU2, and EU3, respectively). Some HCO-related sequences with all conserved histidines deteriorated are shown as yellow dots.

clusters G and H all possess the HxH motif (Supporting Information Fig. S1) and are encoded by pairs of genes in chromosomal vicinity [see two examples in Fig. 4(G,H)].

Members of the HCOH-s1 group have not been classified in publicly available domain databases, while the HCOH-s2 group corresponds to Pfam family DUF2871 and consists of proteins with unknown function. HCOH-s2 members have the conserved HxH motif in their second TMH. They likely form homotrimers with three symmetric sites that can coordinate three hemes [Fig. 1(D)]. Compared to HCOH-s1 proteins, HCOH-s2 proteins additionally possess a conserved “RE” motif at the end of the first TMH and a conserved histidine in the fourth TMH (Fig. 2 and Supporting Information Fig. S2). HCOH-s2 proteins are mostly from bacterial phyla Firmicutes and Actinobacteria (Supporting Information Fig. S2). Interestingly, several single-cell eukaryotes also possess HCOH-s2 proteins, including some species in the order of Trypanosomatida such

as those of the *Leishmania* genus, *Angomonas deanei*, and *Strigomonas culicis*. Manual inspection of weak PSI-BLAST hits also revealed a divergent HCOH-s2 protein from *Naegleria gruberi* (gi|290974763, Fig. 2), a free-living single-cell eukaryotic species of the Heterolobosea class. HCOH-s2 may be present in the ancestor of eukaryotes, and its patchy phylogenetic distribution suggests that it may be lost independently in most eukaryotic lineages. *Leishmania* species are parasites for leishmaniasis, a disease that causes skin sores and visceral failure.^{31,32} *Leishmania* species only include the last three enzymes in the heme biosynthetic pathway.³³ Although *Leishmania* may be able to synthesize heme from heme precursors, it is thought to transport heme with an unknown mechanism^{34,35} and is uniquely dependent on the acquisition of exogenous heme for survival.^{36,37} Considering the membrane localization and potential heme-binding capability of HCOH-s2 proteins, *Leishmania* HCOH-s2 proteins might be involved in the maintenance of heme homeostasis in these parasites. One hypothesis about their function is that *Leishmania* HCOH-s2 proteins sequester hemes in the membrane to reduce heme toxicity³⁸ to the cell and increase heme accessibility to other membrane proteins.

The majority of HCOH-s1 and HCOH-s2 proteins consist of a single domain corresponding to one EU of four TMHs. As exceptions, one HCOH-s1 protein has an N-terminal divergent cupin_2 domain as suggested by HHpred (gi|91786010, Fig. 5), and all F-cluster HCOH-s1 proteins contain an N-terminal thioredoxin domain (e.g., gi|220936290, Fig. 5). A small number of HCOH-s2 members have several additional TMHs (four or seven) (Fig. 5) in their N-termini that do not show detectable sequence similarity to the EUs of HCOs/NORs and HCOH proteins (based on PSI-BLAST and HHpred results).

HCOH groups with three EUs

We found a number of HCOH proteins with 12 TMHs (about 1500 proteins in the nre90 database). These HCOH proteins, like HCOs/NORs, consist of three EUs. However, they usually have fewer conserved histidines than HCOs/NORs. They form several clusters in the CLANS protein clustering diagram (Fig. 3). We divided them into seven groups: HCOH-t1 (HCO homology proteins with three EUs, group 1), HCOH-t2, HCOH-t3, HCOH-t4, HCOH-t5, HCOH-t6, and HCOH-t7 (see Supporting Information Figs. S3–S9 for their sequence information and alignments).

HCOH-t1, HCOH-t2, and HCOH-t3 proteins possess two conserved histidine residues in motifs Hxx of EU1 and xxH of EU3 and do not have conserved histidines in EU2 [Figs. 1(F) and 2]. These proteins also possess a conserved arginine in the third TMH

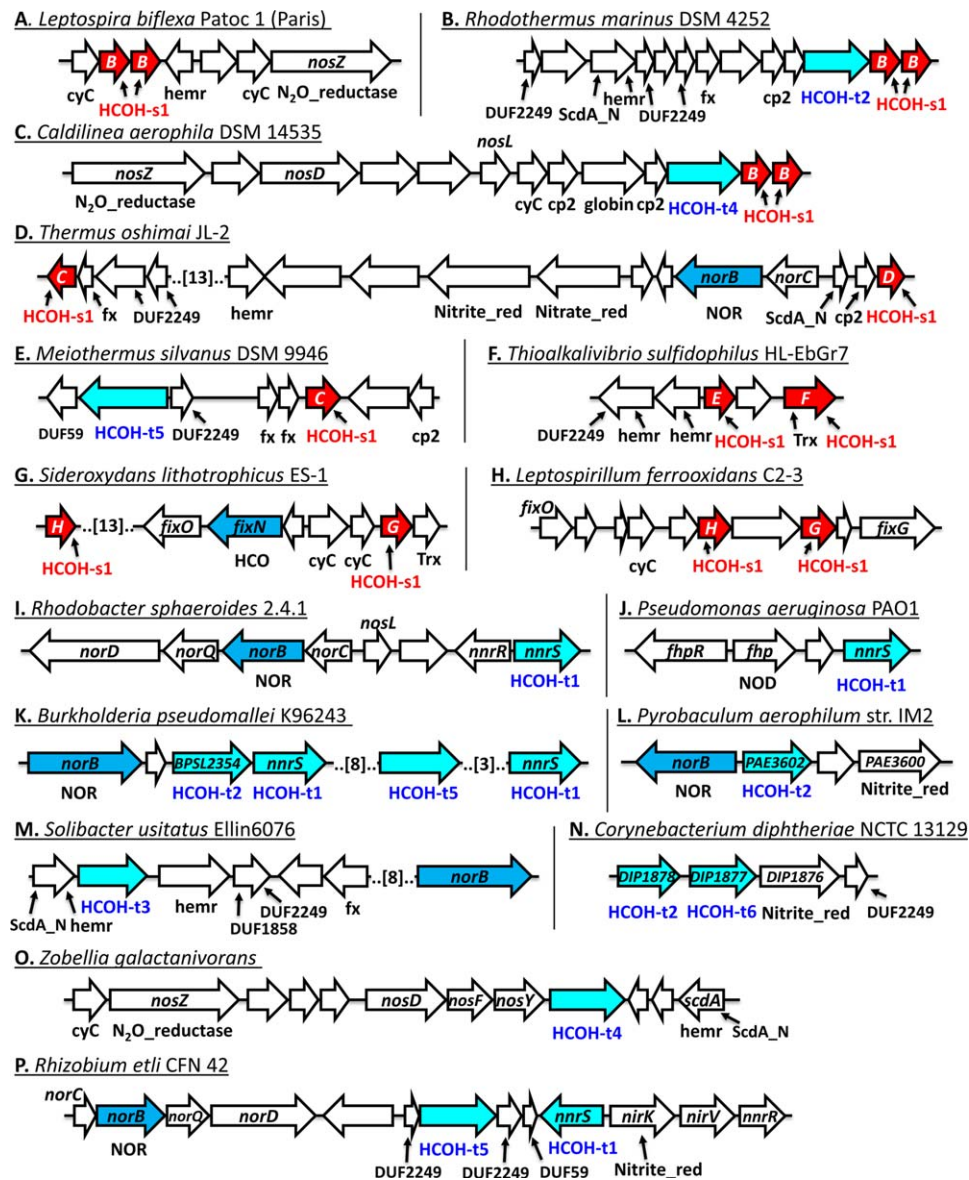


Figure 4. Gene structure diagrams of selected HCOH proteins. Genes are shown as arrows. HCOH-s1 genes, three-EU HCOH genes, and HCO/NOR genes are shown as red arrows, cyan arrows, and blue arrows, respectively. Names of frequently occurring domains are shown below the genes. Numbers of omitted genes are shown in brackets. Domain name abbreviations are as follows: cp2, cupin_2 domain; cyC, cytochrome c; fx, iron-sulfur ferredoxin domain; HCO: heme-copper oxidase; hemr, hemerythrin domain; N₂O_reductase, nitrous oxide reductase; Nitrate_red, nitrate reductase; Nitrite_red, nitrite reductase; NOD, nitric oxide dioxygenase; NOR, nitric oxide reductase; and Trx, thioredoxin.

of EU3 (Fig. 2). The two conserved histidines correspond to the two residues in HCOs/NORs that bind the low-spin heme. HCOH-t1 proteins, mainly from Proteobacteria (Supporting Information Fig. S3), correspond to the previously classified NnrS family (PF05940) in the Pfam database. The *nnrS* gene was identified as the neighboring gene of *nnrR* in *Rhodobacter sphaeroides* 2.4.1 and *R. sphaeroides* 2.4.3.³⁹ The *nnrR* gene encodes a transcriptional regulator that responds to nitric oxide (NO) to activate the expression of the NOR gene *norB*. The expression of the *nnrS* gene is also dependent on the *nnrR* gene.³⁹ The STRING functional association server⁴⁰ suggests

that the *nnrR* and *nnrS* genes co-occur with the *norB*-containing *nor* operon in various bacterial genomes, and they are often chromosomal neighbors, such as in the genome of *R. sphaeroides* 2.4.1 [Fig. 4(I)]. The purified NnrS protein appears to contain heme and copper.⁴¹ Disruption of the *nnrS* gene affected taxis towards nitrate and nitrite, suggesting a role of NnrS in the denitrification process.⁴¹

Recent studies showed that NnrS contributes to NO resistance in the bacterial pathogen *Vibrio cholerae*.⁴² NO is a host-generated reactive nitrogen species toxic to many bacterial pathogens such as *V. cholerae*. The *nnrS* gene in *V. cholerae* is up-

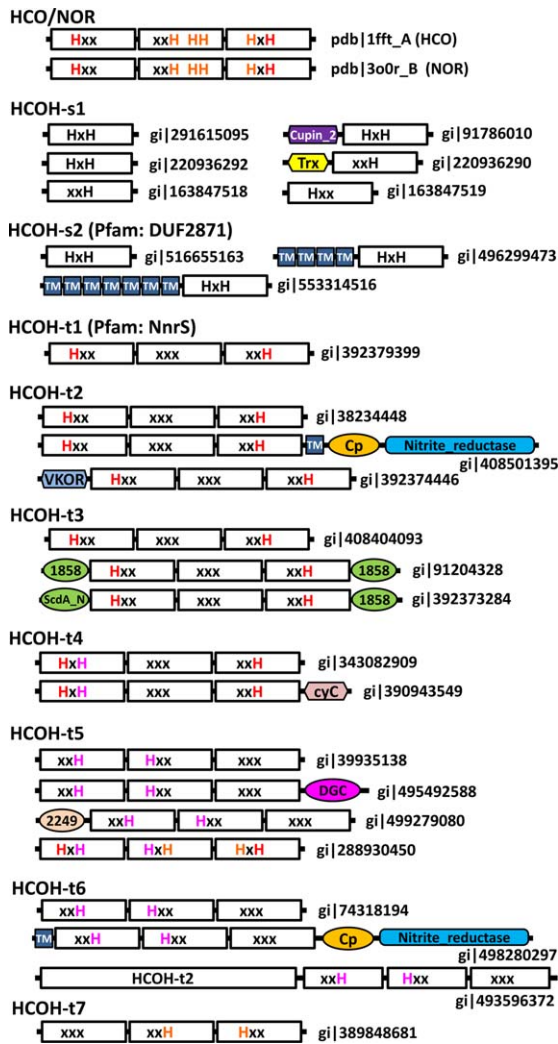


Figure 5. Domain structure diagrams of selected HCO/NOR and HCOH proteins. NCBI gi number or pdb/chain id is shown for each protein. EUs are shown in white rectangular boxes with motifs (HxH, xxH, Hxx, xxx, and HH). For three-EU proteins, conserved histidines at the interface of EU1 and EU3 are shown in red letters, those at the interface of EU2 and EU3 are shown in orange letters, and those at the interface of EU1 and EU2 are shown in magenta letters. Domain or module name abbreviations are as follows: cupin_2, cupin 2 domain; cyC, cytochrome c domain; Trx: thioredoxin domain; Cp, cupredoxin domain; 1858, DUF1858; 2249, DUF2249; TM, predicted transmembrane helix; and VKOR, vitamin K epoxide reductase domain.

regulated by the NorR transcriptional regulator in response to NO.⁴² NorR also activates the expression of the *hmpA* gene that encodes a protein with nitric oxide dioxygenase (NOD) activity that turns NO to less toxic nitrogen oxides.⁴² Gene disruption experiments suggest that *nnrS* and *hmpA* are important for *V. cholerae* colonization of intestines under the NO conditions, suggesting their roles in NO detoxification.⁴² Unlike HmpA, *V. cholerae* NnrS does not remove NO, but it may protect cellular iron pool from NO damage.⁴³ The STRING server revealed that genes encoding NorR, HmpA, and NnrS are

chromosomal neighbors in some bacteria, such as the opportunistic pathogen *Pseudomonas aeruginosa* [Fig. 4(J), *fhpR* and *fhp* encoding orthologous genes of *norR* and *hmpA* in *V. cholerae*, respectively), further supporting their functional association. The *nnrS* gene was also identified in a transposon mutagenesis screen to be important in host colonization of *Neisseria meningitidis*, a bacterial pathogen that causes meningitis.⁴⁴

The majority of HCOH-t2 proteins are from Actinobacteria and Proteobacteria (Supporting Information Fig. S4). They also include some archaeal members mainly from the Crenarchaeota phylum. Most of the HCOH-t2 proteins are annotated as hypothetical proteins. A few HCOH-t2 proteins harbor additional domains. For example, the protein gi|408501395 has an additional TMH, a cupredoxin domain, and a copper-containing nitrite reductase domain C-terminal to the HCOH domain (Fig. 5). Such a domain composition suggests that HCOH-t2 may function in the denitrification process. Another protein, gi|392374446, has a different oxidoreductase domain (VKOR, vitamin K epoxide reductase) located N-terminally to the HCOH domain (Fig. 3). Genes encoding HCOH-t2 proteins are frequently found as neighbors of reductases in the denitrification process. For example, the HCOH-t2 gene is the neighbor to a NOR gene (*norB*) in *Burkholderia pseudomallei* K96243 [Fig. 4(K)] and to a copper-containing nitrite reductase gene in *Corynebacterium diphtheriae* NCTC 13129 [Fig. 4(N)]. In the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum* str. IM2, the HCOH-t2 gene (*PAE3602*) is adjacent to a NOR gene (*norB*) and a nitrite reductase subunit (cytochrome D1) [Fig. 4(L)]. These three genes were all up-regulated after induction with nitrate.⁴⁵ The association with these reductase genes also suggests that HCOH-t2 may function in the denitrification process.

The HCOH-t3 group consists of 14 bacterial sequences and one archaeal sequence forming a loosely connected cluster in the CLANS diagram (Fig. 3). A few HCOH-t3 proteins also possess DUF1858 (Pfam: PF08984) and ScdA_N (Pfam: PF04405) domains (Fig. 5). HHpred results suggest that DUF1858, a domain of unknown function, is distantly related to the ScdA_N domain. ScdA_N domain is named after the N-terminal domain of *Staphylococcus aureus* protein ScdA, which also contains the hemerythrin domain (Pfam: PF01814) that binds non-heme diiron.⁴⁶ The bacterial hemerythrin domain-containing RIC (repair of iron centers) family proteins, including ScdA from *S. aureus*, DnrN from *Neisseria gonorrhoeae* and YtfE from *Escherichia coli*, confer resistance to reactive nitrogen and oxygen molecules such as NO and H₂O₂ by repairing their damages to iron-sulfur centers.⁴⁷ The presence of the ScdA_N domain in many RIC proteins

suggests that ScdA_N could aid in oxidative or nitrosative stress response. Such a domain in a few HCOH-t3 proteins indicates that they may also be involved in resistance to reactive nitrogen molecules such as NO, like some HCOH-t1 (NnrS) members. In *Solibacter usitatus* Ellin6076, the HCOH-t3 gene (*Acid_2923*) is in the neighborhood of genes of the RIC family and genes containing ScdA_N and DUF1858 domains [Fig. 4(M)], further supporting their functional associations.

HCOH-t4 proteins exhibit the histidine patterns of HxH, xxx, xxH in EU1, EU2, and EU3, respectively (Fig. 2). Similar to HCOH-t1, HCOH-t2, and HCOH-t3, such a pattern allows coordination of a heme group at the interface of EU1 and EU3. HCOH-t4 proteins are mainly from the Bacteroidetes phylum (Supporting Information Fig. S6). A few of them contain a cytochrome c domain at the C-terminus (Fig. 5), suggesting that they may be involved in cytochrome c-dependent electron transfer. The HCOH-t4 gene is often located near the *nos* operon [one example shown in Fig. 4(O)] that encodes nitrous oxide (N₂O) reductase, which catalyzes the final step of the denitrification pathway: conversion of N₂O to dinitrogen. A gene encoding the RIC family protein ScdA is also frequently found to be close to the HCOH-t4 gene [e.g., Fig. 4(O)], suggesting that HCOH-t4 could be involved in denitrification and detoxification of reactive molecules.

The majority of HCOH-t5 and HCOH-t6 proteins possess the xxH and Hxx motifs in the second TMH of EU1 and the second TMH of EU2, respectively, and lack conserved histidines in EU3 (Fig. 2 and Supporting Information Figs. S7 and S8). Such a pattern allows coordination of one heme group at the interface between EU1 and EU2 [Fig. 1(G)]. HCOH-t5 members are present in both bacteria and archaea. The bacterial members of HCOH-t5 are mainly from Proteobacteria and Firmicutes, while the archaeal members are all from the Halobacteria class of the Euryarchaeota phylum (Supporting Information Fig. S7). The majority of HCOH-t5 proteins are annotated as hypothetical proteins and do not contain additional domains. As exceptions, a few HCOH-t5 proteins possess domains of unknown function such as DUF2249 (Pfam: PF10006) and DGC (Pfam: PF08859, a domain with four conserved cysteines that likely coordinate zinc). The STRING server revealed strong association of HCOH-t5 proteins with proteins containing DUF2249 domain, proteins with DUF59 domain, and RIC proteins based on evidence of gene neighborhood, gene fusion (in the case of DUF2249 domain), and gene co-occurrence. For example, the HCOH-t5 gene from *Rhizobium etli* CFN 42 (*RHE_PF00521*) is predicted to be functionally associated with two genes containing DUF2249 domains (*RHE_PF00520* and *RHE_PF00522*) and a gene with DUF59 domain

(*RHE_PF00523*) [Fig. 4(P)]. These genes are also neighbors to *nnrS* (a HCOH-s1 gene), *nnrR*, and gene clusters encoding denitrification enzymes such as NOR and nitrite reductase [Fig. 4(O)].

HCOH-t6 proteins are mainly from Actinobacteria (Supporting Information Fig. S8). Some HCOH-t6 proteins are annotated as “multicopper oxidase” or “nitrite reductase,” as they also possess a cupredoxin domain and a copper-containing nitrite reductase domain. Such a domain composition is similar to a few HCOH-t2 proteins described above (Fig. 5). Interestingly, HCOH-t6 and HCOH-t2 genes are often chromosomal neighbors, such as *DIP1877* (a HCOH-t6 gene) and *DIP1878* (a HCOH-t2 gene) of *Corynebacterium diphtheriae* NCTC 13129 [Fig. 4(N)]. In one case, HCOH-t2 and HCOH-t6 are fused together in one open reading frame (gi|493596372 from *Actinomyces urogenitalis*, Fig. 5). These observations suggest that HCOH-t2 and HCOH-t6 may have related functions.

HCOH-t7 proteins exhibit yet another pattern of histidine motifs with xxH in EU2 and Hxx in EU3, while lacking conserved histidines in EU1 (Fig. 2 and Supporting Information Fig. S9). Such a pattern would allow coordination of one heme group at the interface between EU2 and EU3 [Fig. 1(H)]. HCOH-t7 proteins form two small clusters in the CLANS diagram (Fig. 3). Each cluster has restricted phylogenetic distribution. They are from the bacterial phylum Aquificae and from the archaeal class Halobacteria of the Euryarchaeota phylum, respectively (Supporting Information Fig. S9). These proteins are annotated as hypothetical proteins and do not have additional domains.

In summary, three-EU HCOH genes are often neighbors to genes involved in the denitrification process and/or detoxification of small reactive molecules such as NO (Fig. 4). Many of these neighboring genes encode various denitrification enzymes such as nitrate reductase, nitrite reductase, NOR, and nitrous oxide reductase (Fig. 4). The detoxification genes include those that encode NOD and the RIC family proteins with hemerythrin domains. In addition, genes encoding a few domains of unknown functions, such as ScdA_N, DUF2249, DUF1858, and DUF59, are also frequently found in the neighborhood of three-EU HCOH genes. Three-EU HCOH proteins from different groups are sometimes gene neighbors (Fig. 4), suggesting that they are involved in the same biological process. The gene neighborhood associations are consistent with the domain contents of limited multi-domain three-EU HCOH proteins, as some of these proteins contain nitrite reductase domain, DUF2249 domain and DUF1858 domain (Fig. 5).

For some single-EU HCOH proteins such as HCOH-s2 (DUF2871) proteins and cluster A HCOH-s1 proteins, we did not find strong functional

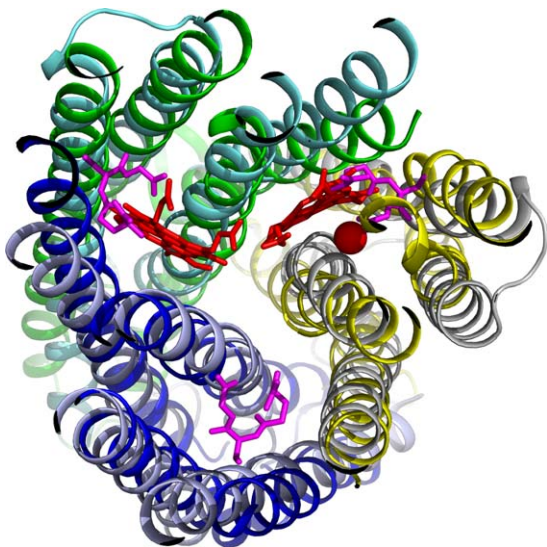


Figure 6. Structure superposition of an HCO superfamily member (pdb id: 3mk7) and a MAPEG member (pdb id: 4al0). EU1, EU2, and EU3 of 3mk7 are shown in blue, yellow, and green, respectively. The heme groups and the non-heme iron are shown as red sticks and a red sphere, respectively. The three chains of 4al0 aligned to EU1, EU2, and EU3 of 3mk7 are shown in light blue, grey, and cyan, respectively. The ligands of 4al0 are shown as magenta sticks.

associations to denitrification/detoxification genes according to the results of the STRING server. On the other hand, genes encoding HCOH-s1 proteins of clusters B, C, D, E, and F are frequently neighbors to denitrification genes and potential detoxification genes such as those with the hemerythrin domain and the globin domain⁴⁸ (Fig. 4(A–F)). Like many three-EU HCOH genes, these single-EU HCOH genes often have neighbors with domains of unknown function such as ScdA_N, DUF2249 and DUF59 [Fig. 4(A–F)]. Other genes frequently in the vicinity of HCOH-s1 genes include those encoding proteins with cytochrome c domain (cyC), ferredoxin domain (fx), thioredoxin domain (Trx), and cupin_2 domain (cp2) [Fig. 4(A–F)]. Some three-EU HCOH genes are also found to be neighbors of HCOH-s1 genes [e.g., Fig. 4(B,C,E)].

Limited experimental studies on NnrS proteins, gene context analysis and domain content analysis indicate the involvement of many HCOH proteins in denitrification and detoxification. The molecular mechanisms of their functions remain unknown and await further experimental investigations. As putative heme-binding proteins, HCOH proteins could contribute to denitrification and detoxification in several ways, such as maintenance of cellular iron and heme homeostasis, being part of the electron transfer pathways in various denitrification enzymes, binding/sequestering/exporting reactive nitrogen species, and possessing enzymatic activities that convert reactive nitrogen species to less toxic molecules.

Structure similarity between the catalytic subunits of HCOs/NORs and the MAPEG family proteins

After establishing the EUs for the HCO superfamily, we sought to explore its relationship to known structures. We queried the structures of HCO/NOR EUs against the SCOP⁴⁹ database using the HorA server,⁵⁰ which not only reports structural similarity, but also evaluates if the similarity represents homology or analogy. HorA identified a MAPEG structure as the top hit following the structures from the HCO superfamily, with relatively good scores for both structure comparison (Dali Z-score 7.6) and sequence comparison (HHpred probability 0.52), resulting in an overall score (5.042) that is consistent with scores derived for distant homologs.⁵¹ MAPEG proteins are a group of membrane-bound enzymes with diverse functions, including glutathione transferase activity that provides protection from oxidative stress in the membrane.²⁸ The MAPEG fold, consisting of four TMHs, forms a homotrimer that binds three substrates in a symmetric manner. The four TMHs of a MAPEG subunit adopt the same topology as the four TMHs in each EU of HCOs/NORs, showing right-handed connections between them. In addition, an unexpected structural similarity lies in the same arrangement of the three subunits of the MAPEG homotrimer compared to the three EUs in HCOs/NORs, as noticed before.²⁹ A superposition of a trimeric MAPEG structure (pdb id: 4al0) onto an HCO structure (pdb id: 3mk7) (Fig. 6) reveals a striking similarity (Dali Z-score: 20.6, RMSD: 3.3Å) covering all 12 TMHs. Each of the three EUs of HCO structure corresponds to one monomer of the MAPEG homotrimer (Fig. 6). Interestingly, the hemes from the HCO structure (red sticks, Fig. 6) overlap with the MAPEG substrate (glutathione, magenta sticks, Fig. 6), which is located at the interfaces of the monomers. Combined with the trimeric state required for MAPEG enzyme function,^{52,53} the structural similarity and similar active site position suggest that MAPEG proteins and HCOs/NORs are evolutionarily related.²⁹

Evolutionary scenarios of HCOs/NORs and the HCOH proteins

Gene duplication, divergence, fusion, and fission are the main driving forces in the evolution of proteins with novel functions.^{54–56} It is estimated that a large fraction of proteins form oligomers with functional importance.⁵⁷ Evolution of oligomeric protein complexes has drawn interest in both theoretic and experimental studies.^{58,59} Gene duplication is considered as the cause to generate heteromers (hetero-oligomers) from homomers (homo-oligomers), for example, in the evolution of proteasomes⁶⁰ and chaperonins.⁶¹ Subunits in a heteromer have

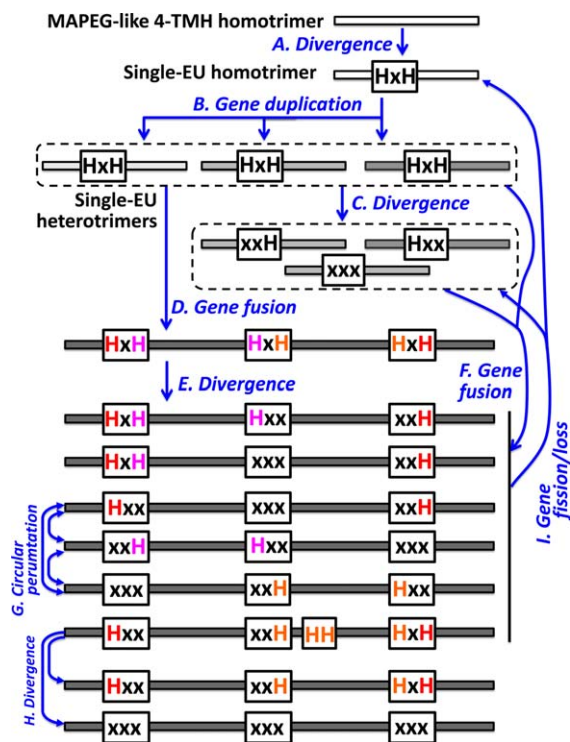


Figure 7. Possible evolutionary events in HCOs/NORs and HCOH proteins.

different sequences and structures that break the symmetry of homomers and allow more versatile functions.⁶² It is estimated that a large fraction of membrane proteins form oligomers or are internally pseudosymmetric,^{63,64} like the catalytic subunits of HCOs/NORs.

The structural similarity between the catalytic subunits of HCOs/NORs and the MAPEG trimers suggests that HCOs/NORs could have evolved from a four-TMH ancestral protein that forms MAPEG-like homotrimers (Fig. 7). It is likely that a fortuitous sequence divergence event of this ancestor [Fig. 7(A)] gave rise to the HxH motif in the second TMH that enabled binding of heme groups. The discovery of single-EU HCOH proteins with four TMHs and the HxH motif supports this hypothesis. The homotrimers of four-TMH single-EU proteins with the HxH motif would have three symmetric heme-binding sites located at the interfaces of the monomers, similar to those of the MAPEG proteins. Single-EU HCOH genes could have undergone gene duplications [Fig. 7(B)] followed by sequence divergence to generate multiple copies of single-EU HCOH genes encoding proteins that form heterotrimers. If the HxH motif is kept in all duplicated genes, a heterotrimer with three heme-binding sites could be formed. In case some histidines are mutated [Fig. 7(C)], like the HCOH-s1 cluster B proteins, heterotrimers that bind less than three hemes could evolve. The fusion of three single-EU genes would result in an open reading frame with three EUs [Fig. 7(D,F)]. Three-EU proteins could bind

three hemes with six histidines if the HxH motif is kept in all three EUs [Fig. 7(D)]. Interestingly, we identified one archaeal protein in the HCOH-t5 group (gi|288930450 in Fig. 2, marked by a star in Fig. 3) that has three HxH motifs and could bind three hemes. Deterioration of some histidines, either in the stage of single-EU ancestors [Fig. 7(C)] or three-EU ancestors [Fig. 7(E)], could lead to three-EU HCOH proteins with fewer than six conserved histidines and thus less than three heme-binding sites.

Most of extant three-EU HCOH proteins possess two conserved histidines and are inferred to bind a single heme. They mostly follow one of three histidine patterns: Hxx.xxx.xxH (HCOH-t1, HCOH-t2, HCOH-t3, and HCOH-t4), xxH.Hxx.xxx (HCOH-t5 and HCOH-t6) and xxx.xxH.Hxx (HCOH-t7) (Fig. 7). These three patterns can also be related by circular permutations of the three EUs [Fig. 7(G)]. We also identified a small number of archaeal HCOH-t5 proteins (e.g., gi|389847617 in Fig. 2) that contain a combination of patterns of HCOH-t1 and HCOH-t5 and have four histidines in the motif pattern of HxH.Hxx.xxH (Fig. 7). These proteins could thus bind two hemes.

HCOs/NORs follow the histidine pattern of Hxx.xxH.HH.HxH. The HH motif in the seventh TMH (the third TMH in EU2) is a unique feature of HCOs/NORs that is not present in HCOH proteins. Such an addition allows HCOs/NORs to coordinate a copper or non-heme iron in addition to binding two hemes. We also identified several small groups of proteins closely related to HCOs/NORs that have some or all of the conserved histidines deteriorated. Two small groups (light blue dots in Fig. 3 and also see the close-up view in Supporting Information Fig. S10) have the HH motif deteriorated (see Supporting Information Figs. S14 and S15 for their alignments). Four small groups (yellow dots in Fig. 3 and Supporting Information Fig. S10) have all of the six histidines deteriorated (see Supporting Information Figs. S16–S19 for their alignments). These groups could have evolved from the HCOs/NORs by sequence divergence [Fig. 7(H)].

Although it is likely that three-EU HCOH proteins have evolved by gene duplication and fusion of single-EU HCOH ancestors, the opposite evolutionary scenario, gene split (or fission) of three-EU HCOH genes to generate single-EU HCOH open reading frames [Fig. 7(I)], is also plausible. For example, HCOH-s1 proteins of clusters H and G, especially cluster H proteins, are close to the HCOs/NORs in the CLANS diagram (Fig. 3). Their top BLAST hits include C-type HCOs. In the genome of *Sideroxydans lithotrophicus* ES-1, one cluster G HCOH-s1 gene and one cluster H HCOH-s1 gene are close to each other and neighboring to a C-type HCO (Fig. 4). It is likely that HCOH-s1 proteins of

clusters G and H are derived from C-type HCO proteins by gene fission.

Conclusions

Comparative sequence-structure analysis revealed novel homology between a number of HCOH proteins and the catalytic subunits of HCOs/NORs. HCOH proteins form groups of four-TMH single-EU proteins (HCOH-s1 and HCOH-s2) and groups of 12-TMH three-EU proteins (HCOH-t1, HCOH-t2, HCOH-t3, HCOH-t4, HCOH-t5, HCOH-t6, and HCOH-t7). Among these groups, only HCOH-s2 and HCOH-t1 correspond to known domains (DUF2871 and NnrS, respectively), while the majority of other HCOH members are currently annotated as hypothetical proteins without known domains. Gene context and domain content analyses, coupled with limited experimental studies of NnrS, suggest that most HCOH proteins are involved in the denitrification process and/or detoxification of reactive small molecules. Based on the structures of HCOs/NORs, single-EU HCOH proteins could form homotrimers or heterotrimers with active sites located at the interfaces between monomers. Conserved histidines in HCOH proteins indicate that they can bind heme. Strong structural similarity was observed between the homotrimers of the MAPEG family membrane enzymes and the catalytic subunits of HCOs/NORs. Such a similarity, together with the discovery of single-EU HCOH proteins, suggests that HCOs/NORs and three-EU HCOH proteins could have evolved from four-TMH ancestors that form homotrimers similar to MAPEG proteins. Gene duplication, sequence divergence, and gene fusion of ancestral single-EU HCOH proteins could give rise to three-EU HCOH proteins and HCOs/NORs. Conversely, gene fission of three-EU HCOH proteins or HCOs/NORs may have produced some extant single-EU HCOH proteins.

Materials and Methods

Sequence similarity searches

PSI-BLAST²⁶ iterations were conducted to search for homologs of the HCO superfamily proteins starting from one representative with known structure (protein databank (PDB⁶⁵) id: 3o0r, chain B)¹⁶ against a database composed of NCBI non-redundant proteins and environmental sequences with maximal 90% identity (nre90) protein database (e-value inclusion cutoff: 1e-4). To perform transitive searches, PSI-BLAST hits were grouped by BLASTCLUST (with the score coverage threshold [-S, defined as the bit score divided by alignment length] set to 1, length coverage threshold (-L) set to 0.5, and no requirement of length coverage on both sequences (-b F)], and a representative sequence from each group was used to initiate new PSI-BLAST searches. Such an

iterative procedure was repeated until convergence. HHpred²⁷ was used for profile-profile-based similarity searches to identify distant homologous relationships for (1) HCOs/NORs (gi|315583520), (2) the HCOH-s1 group (gi|499132825), (3) the HCOH-s2 group (gi|316941303), (4) the HCOH-t1 group (gi|110347088), (5) the HCOH-t5 group (gi|292656262), and (6) the HCOH-t6 group (gi|256378768) (profile databases used: Pfam,⁶⁶ PDB,⁶⁵ and CDD⁶⁷). Detections of conserved domains are performed by the CDD server⁶⁷ and the HMMER3 package.⁶⁸ We also employed the HorA server⁵⁰ to detect structural homologs for the pseudosymmetric units of HCO proteins, using the C terminus of a NOR structure (pdb: 3o0r, chain B, residue 302–458) as input.

Sequence clustering and multiple sequence alignment

Sequence clustering was performed and visualized by the CLANS program.³⁰ Several cutoffs of *P*-values were tried. The *P*-value cutoff 1e-10 was chosen since it gave the best separation between clusters according to manual inspections. We extracted the sequences in each manually defined group of CLANS results and performed multiple sequence alignments by PROMALS3D⁶⁹ for each group. These alignments as well as information about their sequences such as species and domain ranges are available in the Supporting Information. Representative sequences for HCOs/NORs and the newly defined HCOH groups were selected and split into individual EUs of four TMHs. A multiple sequence alignment was constructed for the EUs of these representatives by PROMALS3D. This alignment was manually adjusted by taking into account sequence conservation, structural superposition of known structures of HCOs/NORs by DaliLite⁷⁰ and MUSTANG,⁷¹ hydrophobicity and small residue patterns, and transmembrane regions predictions by Phobius⁷² and TMHMM.⁷³

References

1. Catling DC, Glein CR, Zahnle KJ, McKay CP (2005) Why O₂ is required by complex life on habitable planets and the concept of planetary "oxygenation time". *Astrobiology* 5:415–438.
2. Payne JL, Boyer AG, Brown JH, Finnegan S, Kowalewski M, Krause RA Jr, Lyons SK, McClain CR, McShea DW, Novack-Gottshall PM, Smith FA, Stempien JA, Wang SC (2009) Two-phase increase in the maximum size of life over 3.5 billion years reflects biological innovation and environmental opportunity. *Proc Natl Acad Sci USA* 106:24–27.
3. Dahl TW, Hammarlund EU, Anbar AD, Bond DP, Gill BC, Gordon GW, Knoll AH, Nielsen AT, Schovsbo NH, Canfield DE (2010) Devonian rise in atmospheric oxygen correlated to the radiations of terrestrial plants and large predatory fish. *Proc Natl Acad Sci USA* 107: 17911–17915.

4. Gribaldo S, Talla E, Brochier-Armanet C (2009) Evolution of the haem copper oxidases superfamily: a rooting tale. *Trends Biochem Sci* 34:375–381.
5. Pereira MM, Santana M, Teixeira M (2001) A novel scenario for the evolution of haem-copper oxygen reductases. *Biochim Biophys Acta* 1505:185–208.
6. van der Oost J, de Boer AP, de Gier JW, Zumft WG, Stouthamer AH, van Spanning RJ. The heme-copper oxidase family consists of three distinct types of terminal oxidases and is related to nitric oxide reductase. *FEMS Microbiol Lett* 121:1–9.
7. Richter OM, Ludwig B (2003) Cytochrome c oxidase—structure, function, and physiology of a redox-driven molecular machine. *Rev Physiol Biochem Pharmacol* 147:47–74.
8. Ferguson-Miller S, Babcock GT (1996) Heme/copper terminal oxidases. *Chem Rev* 96:2889–2908.
9. Michel H, Behr J, Harrenga A, Kannt A (1998) Cytochrome c oxidase: structure and spectroscopy. *Annu Rev Biophys Biomol Struct* 27:329–356.
10. Iwata S, Ostermeier C, Ludwig B, Michel H (1995) Structure at 2.8 Å resolution of cytochrome c oxidase from *Paracoccus denitrificans*. *Nature* 376:660–669.
11. Sousa FL, Alves RJ, Ribeiro MA, Pereira-Leal JB, Teixeira M, Pereira MM (2012) The superfamily of heme-copper oxygen reductases: types and evolutionary considerations. *Biochim Biophys Acta* 1817:629–637.
12. Brochier-Armanet C, Talla E, Gribaldo S (2009) The multiple evolutionary histories of dioxygen reductases: Implications for the origin and evolution of aerobic respiration. *Mol Biol Evol* 26:285–297.
13. Gennis RB (1998) Multiple proton-conducting pathways in cytochrome oxidase and a proposed role for the active-site tyrosine. *Biochim Biophys Acta* 1365:241–248.
14. Rauhamaki V, Baumann M, Soliymani R, Puustinen A, Wikstrom M (2006) Identification of a histidine-tyrosine cross-link in the active site of the cbb3-type cytochrome c oxidase from *Rhodobacter sphaeroides*. *Proc Natl Acad Sci USA* 103:16135–16140.
15. Hemp J, Christian C, Barquera B, Gennis RB, Martinez TJ (2005) Helix switching of a key active-site residue in the cytochrome cbb3 oxidases. *Biochemistry* 44:10766–10775.
16. Hino T, Matsumoto Y, Nagano S, Sugimoto H, Fukumori Y, Murata T, Iwata S, Shiro Y (2010) Structural basis of biological N₂O generation by bacterial nitric oxide reductase. *Science* 330:1666–1670.
17. Moenne-Loccoz P, Fee JA (2010) Biochemistry. Catalyzing NO to N₂O in the nitrogen cycle. *Science* 330:1632–1633.
18. Hendriks J, Gohlke U, Saraste M (1998) From NO to O₂: nitric oxide and dioxygen in bacterial respiration. *J Bioenerg Biomembr* 30:15–24.
19. Stevanin TM, Laver JR, Poole RK, Moir JW, Read RC (2007) Metabolism of nitric oxide by *Neisseria meningitidis* modifies release of NO-regulated cytokines and chemokines by human macrophages. *Microbes Infect* 9:981–987.
20. Hendriks J, Oubrie A, Castresana J, Urbani A, Gemeinhardt S, Saraste M (2000) Nitric oxide reductases in bacteria. *Biochim Biophys Acta* 1459:266–273.
21. Sousa FL, Alves RJ, Pereira-Leal JB, Teixeira M, Pereira MM (2011) A bioinformatics classifier and database for heme-copper oxygen reductases. *PLoS One* 6:e19117.
22. Castresana J, Lubben M, Saraste M, Higgins DG (1994) Evolution of cytochrome oxidase, an enzyme older than atmospheric oxygen. *EMBO J* 13:2516–2525.
23. de Vries S, Schroder I (2002) Comparison between the nitric oxide reductase family and its aerobic relatives, the cytochrome oxidases. *Biochem Soc Trans* 30:662–667.
24. Ducluzeau AL, van Lis R, Duval S, Schoepp-Cothenet B, Russell MJ, Nitschke W (2009) Was nitric oxide the first deep electron sink? *Trends Biochem Sci* 34:9–15.
25. Castresana J, Saraste M (1995) Evolution of energetic metabolism: the respiration-early hypothesis. *Trends Biochem Sci* 20:443–448.
26. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
27. Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33:W244–248.
28. Jakobsson PJ, Morgenstern R, Mancini J, Ford-Hutchinson A, Persson B (1999) Common structural features of MAPEG—a widespread superfamily of membrane associated proteins with highly divergent functions in eicosanoid and glutathione metabolism. *Protein Sci* 8:689–692.
29. Holm PJ, Bhakat P, Jegerschold C, Gyobu N, Mitsuoka K, Fujiyoshi Y, Morgenstern R, Hebert H (2006) Structural basis for detoxification and oxidative stress protection in membranes. *J Mol Biol* 360:934–945.
30. Frickey T, Lupas A (2004) CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20:3702–3704.
31. Oliveira F, de Carvalho AM, de Oliveira CI (2013) Sand-fly saliva—man: the trigger trio. *Frontiers Immunol* 4:375.
32. Vannier-Santos MA, Martiny A, de Souza W (2002) Cell biology of *Leishmania* spp.: invading and evading. *Curr Pharma Des* 8:297–318.
33. Koreny L, Lukes J, Obornik M (2010) Evolution of the haem synthetic pathway in kinetoplastid flagellates: an essential pathway that is not essential after all? *Intl J Parasitol* 40:149–156.
34. Campos-Salinas J, Cabello-Donayre M, Garcia-Hernandez R, Perez-Victoria I, Castany S, Gamarro F, Perez-Victoria JM (2011) A new ATP-binding cassette protein is involved in intracellular haem trafficking in *Leishmania*. *Mol Microbiol* 79:1430–1444.
35. Huynh C, Yuan X, Miguel DC, Renberg RL, Protchenko O, Philpott CC, Hamza I, Andrews NW (2012) Heme uptake by *Leishmania amazonensis* is mediated by the transmembrane protein LHR1. *PLoS Pathog* 8:e1002795.
36. Dutta S, Furuyama K, Sassa S, Chang KP (2008) *Leishmania* spp.: delta-aminolevulinic acid-inducible neogenesis of porphyria by genetic complementation of incomplete heme biosynthesis pathway. *Exp Parasitol* 118:629–636.
37. Chang CS, Chang KP (1985) Heme requirement and acquisition by extracellular and intracellular stages of *Leishmania mexicana amazonensis*. *Mol Biochem Parasitol* 16:267–276.
38. Anzaldi LL, Skaar EP (2010) Overcoming the heme paradox: heme toxicity and tolerance in bacterial pathogens. *Infect Immun* 78:4977–4989.
39. Kwiatkowski AV, Laratta WP, Toffanin A, Shapleigh JP (1997) Analysis of the role of the nrrR gene product in the response of *Rhodobacter sphaeroides* 2.4.1 to exogenous nitric oxide. *J Bacteriol* 179:5618–5620.
40. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 41:D808–D815.
41. Bartnikas TB, Wang Y, Bobo T, Veselov A, Scholes CP, Shapleigh JP (2002) Characterization of a member of

- the NnrR regulon in *Rhodobacter sphaeroides* 2.4.3 encoding a haem-copper protein. *Microbiology* 148:825–833.
42. Stern AM, Hay AJ, Liu Z, Desland FA, Zhang J, Zhong Z, Zhu J (2012) The NorR regulon is critical for *Vibrio cholerae* resistance to nitric oxide and sustained colonization of the intestines. *MBio* 3:e00013–00012.
 43. Stern AM, Liu B, Bakken LR, Shapleigh JP, Zhu J (2013) A novel protein protects bacterial iron-dependent metabolism from nitric oxide. *J Bacteriol* 195:4702–4708.
 44. Jamet A, Euphrasie D, Martin P, Nassif X (2013) Identification of genes involved in *Neisseria meningitidis* colonization. *Infect Immun* 81:3375–3381.
 45. Cozen AE, Weirauch MT, Pollard KS, Bernick DL, Stuart JM, Lowe TM (2009) Transcriptional map of respiratory versatility in the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *J Bacteriol* 191:782–794.
 46. French CE, Bell JM, Ward FB (2008) Diversity and distribution of hemerythrin-like proteins in prokaryotes. *FEMS Microbiol Lett* 279:131–145.
 47. Overton TW, Justino MC, Li Y, Baptista JM, Melo AM, Cole JA, Saraiva LM (2008) Widespread distribution in pathogenic bacteria of di-iron proteins that repair oxidative and nitrosative damage to iron-sulfur centers. *J Bacteriol* 190:2004–2013.
 48. Frey AD, Kallio PT (2005) Nitric oxide detoxification—a new era for bacterial globins in biotechnology? *Trends Biotechnol* 23:69–73.
 49. Lo Conte L, Ailey B, Hubbard TJ, Brenner SE, Murzin AG, Chothia C (2000) SCOP: a structural classification of proteins database. *Nucleic Acids Res* 28:257–259.
 50. Kim BH, Cheng H, Grishin NV (2009) HorA web server to infer homology between proteins using sequence and structural similarity. *Nucleic Acids Res* 37:W532–W538.
 51. Cheng H, Kim BH, Grishin NV (2008) Discrimination between distant homologs and structural analogs: lessons from manually constructed, reliable data sets. *J Mol Biol* 377:1265–1278.
 52. Martinez Molina D, Wetterholm A, Kohl A, McCarthy AA, Niegowski D, Ohlson E, Hammarberg T, Eshaghi S, Haeggstrom JZ, Nordlund P (2007) Structural basis for synthesis of inflammatory mediators by human leukotriene C4 synthase. *Nature* 448:613–616.
 53. Martinez Molina D, Eshaghi S, Nordlund P (2008) Catalysis within the lipid bilayer-structure and mechanism of the MAPEG family of integral membrane proteins. *Curr Opin Struct Biol* 18:442–449.
 54. Ohno S (1970) *Evolution by Gene Duplication*. Springer-Verlag, New York.
 55. Kummerfeld SK, Teichmann SA (2005) Relative rates of gene fusion and fission in multi-domain proteins. *Trends Genet* 21:25–30.
 56. Pasek S, Risler JL, Brezellec P (2006) Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. *Bioinformatics* 22:1418–1423.
 57. Goodsell DS, Olson AJ (2000) Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct* 29:105–153.
 58. Levy ED, Boeri Erba E, Robinson CV, Teichmann SA (2008) Assembly reflects evolution of protein complexes. *Nature* 453:1262–1265.
 59. Lynch M (2012) The evolution of multimeric protein assemblages. *Mol Biol Evol* 29:1353–1366.
 60. Bouzat JL, McNeil LK, Robertson HM, Solter LF, Nixon JE, Beever JE, Gaskins HR, Olsen G, Subramaniam S, Sogin ML, Lewin HA (2000) Phylogenomic analysis of the alpha proteasome gene family from early-diverging eukaryotes. *J Mol Evol* 51:532–543.
 61. Archibald JM, Logsdon JM Jr, Doolittle WF. Origin and evolution of eukaryotic chaperonins: phylogenetic evidence for ancient duplications in CCT genes (2000) *Mol Biol Evol* 17:1456–1466.
 62. Pereira-Leal JB, Levy ED, Kamp C, Teichmann SA (2007) Evolution of protein complexes by duplication of homomeric interactions. *Genome Biol* 8:R51.
 63. Choi S, Jeon J, Yang JS, Kim S (2008) Common occurrence of internal repeat symmetry in membrane proteins. *Proteins* 71:68–80.
 64. Duran AM, Meiler J (2013) Inverted topologies in membrane proteins: a mini-review. *Comput Struct Biotechnol J* 8:e201308004.
 65. Rose PW, Bi C, Bluhm WF, Christie CH, Dimitropoulos D, Dutta S, Green RK, Goodsell DS, Prlic A, Quesada M, Quinn GB, Ramos AG, Westbrook JD, Young J, Zardecki C, Berman HM, Bourne PE (2013) The RCSB Protein Data Bank: new resources for research and education. *Nucleic Acids Res* 41:D475–482.
 66. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M (2014) Pfam: the protein families database. *Nucleic Acids Res* 42:D222–D230.
 67. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH (2013) CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res* 41:D348–D352.
 68. Eddy SR (2011) Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195.
 69. Pei J, Kim BH, Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* 36:2295–2300.
 70. Holm L, Park J (2000) DaliLite workbench for protein structure comparison. *Bioinformatics* 16:566–567.
 71. Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM (2006) MUSTANG: a multiple structural alignment algorithm. *Proteins* 64:559–574.
 72. Kall L, Krogh A, Sonnhammer EL (2007) Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res* 35:W429–432.
 73. Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Intl Conf Intellig Syst Mol Biol* 6:175–182.