Structure and Sequence Conservation of hao Cluster Genes of Autotrophic Ammonia-Oxidizing Bacteria: Evidence for Their Evolutionary History

David J. Bergmann, 1 Alan B. Hooper, 2 and Martin G. Klotz 3*

Black Hills State University, Spearfish, South Dakota 57783; University of Minnesota, St. Paul, Minnesota 55108; and Department of Biology, University of Louisville, Louisville, Kentucky 40292

Received 27 December 2004/Accepted 21 April 2005

Comparison of the organization and sequence of the hao (hydroxylamine oxidoreductase) gene clusters from the gammaproteobacterial autotrophic ammonia-oxidizing bacterium (aAOB) Nitrososphaera virosa and the betaproteobacterial aAOB Nitrosospira multiformis and Nitrosomonas europaea revealed a highly conserved gene cluster encoding the following proteins: hao, hydroxylamine oxidoreductase; orf2, a putative protein; cycA, cytochrome c554; and cycB, cytochrome c552. The deduced protein sequences of HAO, c554, and c552 were highly similar in all aAOB despite their differences in species evolution and codon usage. Phylogenetic inference revealed a broad family of multi-c-heme proteins, including HAO, the pentaheme nitrite reductase, and tetrahionate reductase. The c-hemes of this group also have a nearly identical geometry of heme orientation, which has remained conserved during divergent evolution of function. High sequence similarity is also seen within a protein family, including cytochromes c552, NrfII/B, and NapC/NirT. It is proposed that the hydroxylamine oxidation pathway evolved from a nitrite reduction pathway involved in anaerobic respiration (denitrification) during the radiation of the Proteobacteria. Conservation of the hydroxylamine oxidation module was maintained by functional pressure, and the module expanded into two separate narrow taxa after a lateral gene transfer event between gamma- and betaproteobacterial ancestors of extant aAOB. HAO-encoding genes were also found in six non-aAOB, either singly or tandemly arranged with an orf2 gene, whereas a c554 gene was lacking. The conservation of the hao gene cluster in general and the uniqueness of the c554 gene in particular make it a suitable target for the design of primers and probes useful for molecular ecology approaches to detect aAOB.

Autotrophic ammonia-oxidizing bacteria (aAOB) obtain all energy for growth from the oxidation of ammonia to nitrite. Ammonia is first oxidized to hydroxylamine by the membrane-bound hetero-trimeric copper enzyme ammonia monooxygenase (AMO) (4, 22, 40). Hydroxylamine is oxidized to nitrite by the periplasmic enzyme hydroxylamine oxidoreductase (HAO) (61). HAO binds seven c-type hemes and an active-site heme, known as heme P460 (3, 57). A small amount of hydroxylamine is also oxidized by the periplasmic, monomeric copper nitrite oxidase P460 protein, proposed to scavenge hydroxylamine not oxidized by HAO (9, 15, 42). aAOB occur in two phylogenetic lineages of the Proteobacteria: the closely related genera Nitrospira and Nitrosospira within the Betaproteobacteria and several strains in the gammaproteobacterial genus Nitrospira, including Nitrospira virosaae (18, 43, 54). An unrelated autotrophic nitrifying bacterium, canidatus Brocadia anammoxidans, phylum Planctomycetes (29), employs a hydrazine-oxidizing HAO for part of the anaerobic conversion of ammonia and nitrite to dinitrogen.

Based on studies of Nitrosomonas europaea, the hydroxylamine oxidation in extant aAOB is assumed to proceed through catalysis by the octaheme HAO followed by electron transfer to membrane ubiquinone mediated by the tetraheme cytochrome c554 and the membrane-associated tetraheme cytochrome c552 (22, 24). The latter passes two electrons to AMO and channels two electrons into the electron transport chain to cytochrome oxidase. The close functional connection of these cytochromes is supported by the clustering of their encoding genes in the genome of N. europaea ATCC 19718 (Schmidt): HAO (hao)- and cytochrome c554 (cycA and hcy)-encoding genes are part of a cluster that exists in three nearly identical copies (8, 11, 20, 21, 36, 47). In two of the three hao gene clusters, the cycA gene is followed by cycB, which encodes the membrane-anchored tetraheme cytochrome c552 (8). The hao and cycA genes are separated by a hypothetical open reading frame (orf2) in all three copies of the N. europaea hao gene cluster (11). This conservation of Orf2 in the N. europaea genome was peculiar, because earlier primer extension and Northern blot studies identified a terminator for the hao transcript and a functional promoter of the cycA gene between the translational controls of Orf2 (8, 21, 47).

Biochemically, HAOs have been characterized from three aAOB: the aerobic ammonia oxidizers Nitrosospira multiformis and Nitrospira virosaae (23) and the anaerobic ammonia oxidizer B. anammoxidans (52). Amino acid sequence and X-ray crystal structure information are available only for HAO from N. europaea (26, 47). All three enzymes share similar spectral properties, including the unique P460 chromophore. However, partial protein sequences from proteolytic digests of HAO from B. anammoxidans indicate that there may be only limited sequence similarity between HAO from B. anammoxid-
sequences of the entire degenerate primers) was used to obtain a partial sequence for the hao gene of this organism (50).

Here we present the sequence of the entire hao gene cluster that exists as a single copy in the genome of N. oceani and compare it to the hao gene clusters of N. europaea and N. multiformis. Our analyses of gene cluster organization and phylogenetic analyses of the protein sequences encoded by genes in the hao gene cluster support the hypothesis that the hao gene cluster evolved in the gammaproteobacterial ancestors of extant aAOB from genes involved in denitrification and resides in extant betaproteobacterial aAOB as a result of horizontal gene transfer events.

MATERIALS AND METHODS

Bacterial strains, culture maintenance, and DNA isolation. The nucleotide sequence and organization of the hao gene cluster of two strains of Nitrosococcus oceani was determined using different methods. In the case of N. oceani strain C-107WH, a reverse genetics approach (isolation and purification of the HAO protein, amino acid sequencing, and PCR amplification of the hao gene using degenerate primers) was used to obtain a partial sequence for the hao gene. The sequences of the entire hao gene cluster from N. oceani strain ATCC 19707 and N. multiformis C-71 have been obtained by using whole-genome shotgun sequencing.

N. oceani strain C-107WH was obtained as a gift from the collection of Bess B. Ward at Princeton University. Media were prepared using a mixture of 90% artificial seawater (Tropic Marin, Wartenburg, Germany) and 10% natural seawater (pH 8.0) with the following additions before autoclaving: (NH4)2SO4 (10 mM), MgSO4 (1.45 mM), and CaCl2 (0.14 mM). After autoclaving, K2HPO4 was added (final concentration, 0.65 mM), along with 0.5 ml each of two filter-sterilized trace element solutions, A and B (solution A, FeSO4 [4.66 mM], EDTA [4.66 mM], and CuSO4 [0.16 mM]; solution B, Na2MoO4 [8.3 μM], MgCl2 [20 μM], CoCl2 [17 μM], and ZnSO4 [700 μM]). Cultures were grown in 1-liter batches of medium in Fernbach flasks and titrated to pH 8.0 daily with K2CO3 or MgCl2. This cell extract was centrifuged at 150,000 g for 2 min, and the suspension was adjusted with EDTA to a final concentration of 5 mM. This cell extract was centrifuged at 150,000 × g for 1 h at 5°C. The supernatant was decanted, and the pellet was resuspended in 10 ml of 1 M KCl in phosphate buffer and incubated at 5°C for 1 h. After a repeated centrifugation of the suspension followed by decanting of the supernatant (KCl membrane wash), phosphate buffer was added to the KCl membrane wash to adjust the KCl concentration to 0.2 M and the liquid was concentrated to 1 ml on an Amicon stirred cell and a Centricon 10 spin filter. This HAO preparation could be partially reduced by hydroxyamine and exhibited a characteristic absorbance at 460 nm when reduced by diithionite (23).

Sequence comparison and analysis. Sequence similarities were investigated initially using the National Center for Biotechnology Information BLAST program (1). Protein sequences deduced from experimentally determined nucleotide sequences were analyzed with the PSORT (39) or Top-Pred (12) programs to predict protein subcellular location. The hydrophilic properties of the proteins in the NapC-TorC-NirT-NrfB/H protein family. To this end, a total of 5372 BERGMANN ET AL. APPL. ENVIRON. MICROBIOL. available at this stage of the project have been provided freely by the U.S. DOE-JGI for use in this publication/correspondence only, the five nitrifier genome sequences currently in progress will be made available in their entirety through the JGI website following established federal data release policies (http://genome.jgi-psf.org/mic_home.html).

Purification of HAO from N. oceani strain C-107WH and PCR amplification of an hao gene fragment. After harvesting 4.8 g of N. oceani cells by centrifugation, cells were resuspended in 20 ml of 50 mM phosphate buffer (pH 7.5). Cells were broken using three passes through a French press and sonicated for 2 min, and the suspension was adjusted with EDTA to a final concentration of 5 mM. This cell extract was centrifuged at 150,000 × g for 1 h at 5°C. The supernatant was decanted, and the pellet was resuspended in 10 ml of 1 M KCl in phosphate buffer and incubated at 5°C for 1 h. After a repeated centrifugation of the suspension followed by decanting of the supernatant (KCl membrane wash), phosphate buffer was added to the KCl membrane wash to adjust the KCl concentration to 0.2 M and the liquid was concentrated to 1 ml on an Amicon stirred cell with a YM 10 membrane and Centricon 10 spin filters (Amicon-Grace, Beverly, MA). The KCl membrane wash was applied to a 1- to 105-cm column of Sephadex G100 (Sigma Co., St. Louis, MO) equilibrated with 0.2 M KCl in phosphate buffer. The cytochrome (HAO) eluted in the void volume. This was concentrated with an Amicon stirred cell and a Centricon 10 spin filter. This HAO preparation could be partially reduced by hydroxyamine and exhibited a characteristic absorbance at 460 nm when reduced by diithionite (23).

Downloaded from http://aem.asm.org/ on October 30, 2017 by guest
functions. A maximum parsimony tree was built from the ClustalX alignment by performing a full heuristic search with the PAUP* program and the following in effect: 50% majority consensus, random order taxon-addition replicates with tree-bisection-reconnection branch-swapping, mulspar, and steepest decent functions in effect. The quality of the branching patterns was assessed by bootstrap resampling of the data sets using 100 replications. Because inclusion or exclusion of a few characters can highly affect subtree proportions of maximum parsimony trees derived from limited data sets, we also conducted a maximum likelihood inference by subjecting the alignment to a Bayesian inference of phylogeny by using the program MrBayes (v. 3.0b4; written by Huelsenbeck and Ronquist; http://mrbayes.ebc.uu.se/mrbayes). In the latter, the protein sequence alignment was subjected to Metropolis-coupled Monte Carlo Markov chain sampling over 100,000 generations. Four equally heated Markov chains were used to build a sufficient number of reliable trees after the likelihoods of the trees had converged on a stable value and to allow successful swapping between chains. Three independent runs led to convergence on stable likelihood values after 30,000 generations (data not shown). The searches were conducted assuming an equal or a gamma distribution of rates across sites and using the WAG empirical amino acid substitution model (60). In a postrun analysis, MrBayes summarized the results concerning tree topology and branch lengths. By ignoring the trees generated before the convergence on stable likelihood values (removed as “burn-in”), a 50% majority rule consensus phylogram was constructed that displayed the mean branch lengths and posterior probability values of the observed clades. Multiplication by 100 made these probability values comparable to the bootstrap proportions calculated for the clade pattern in the maximum parsimony consensus tree.

Nucleotide sequence accession number. The sequence of the 1,410-bp PCR-amplified fragment of hao from N. oceani strain C-107WH determined using the degenerate reverse primer and other custom-synthesized primers was deposited in GenBank under accession number AY858555.

RESULTS

Sequence of the hao gene cluster. The hao gene cluster of N. oceani strain ATCC 19707, obtained by whole-genome shotgun sequencing (Fig. 1), contained four genes, hao, orf2 (encoding a putative membrane protein), cycA (encoding cytochrome e553), and cycB (encoding cytochrome e552), and closely resembled two of the three copies of the hao gene cluster in the genome of N. europaea (8, 11, 47). A putative promoter with a consensus for rRNA polymerase was identified upstream of a Shine-Dalgarno sequence that precedes the hao gene. A putative 3′-independent transcriptional termination sequence was identified downstream of cycB.

PCR amplification of an hao gene fragment. The N-terminal sequence of the Hao protein from N. oceani strain C-107WH was determined, by Edman degradation at the UMN Microchemical Facility by using a HP241 gas-phase sequenator, to be DIPDLELYEALGVDXYXAXPXELYEAATERY. Based on this sequence, the following degenerate forward primer for PCR amplification was synthesized (1 denotes inosine): 5′-GA(C/T)ATICCIGA(A/G)(C/T)TITA(C/T)GA-3′. The reverse primer, 5′-GTTCACTTGGGGCCTCATCTA-3′, was based on residues 440 to 449 of Hao from N. europaea (47). PCR was performed using a standard program (denaturation for 5 min at 94°C and 25 cycles consisting of 30-s denaturation at 94°C, 30-s annealing at 54°C, and 60-s extension at 72°C, followed by a final 7-min extension step at 72°C).

The partial protein sequence of Hao from N. oceani strain C-107WH (not shown) was identical to that of the first 471 residues of Hao from N. oceani strain ATCC 19707 (Fig. 2), except at positions 22 (Leu in ATCC 19707, Phe in C-107WH), 68 (Ala in ATCC 19707, Asp in C-107WH), 168 (Val in ATCC 19707, Ala in C-107WH), 171 (Thr in ATCC 19707, Ile in C-107WH), and 298 (Trp in ATCC 19707, Leu in C-107WH).

![Fig. 1. Cartoon showing ORFs of the hao gene clusters in Nitrosococcus oceani (Ne_oce), Nitrosomonas europaea (Nm_eur), and Nitrosospira multiformis (Ns_mul).](http://aem.asm.org/)

The PCR product containing a fragment of the N. oceani hao gene was labeled with 32P and used to probe a blot of restriction digests of genomic DNA from N. oceani strain C-107WH. Only one restriction fragment hybridized to the probe in four of the five restriction digests, indicating that only one copy of hao was present in the genome of N. oceani strain C-107WH (Fig. 3). A probe based on a 458-bp fragment of the cycA gene from N. europaea hybridized weakly to genomic DNA from N. oceani strain C-107WH (data not shown), indicating that the sequence identity between the cycA genes from N. europaea and N. oceani was higher than for genes in the amo operon, for which cross-hybridization did not occur (2). In the cycA gene Southern blot assay, the two restriction digests revealing more than one hybridizing restriction fragment were accounted for by the presence of the respective restriction sites seen in the nucleotide sequence within the fragment. The genome sequence of N. oceani strain ATCC 19707 contained, indeed, only one copy of the cycA gene.

Comparison of the hao gene clusters in aerobic ammonia-oxidizing bacteria. The sequences of the Hao proteins from N. europaea (Schmidt), N. multiformis (Surinam), and N. oceani (ATCC 19707) were highly similar with a high degree of conservation in the eight heme-binding regions (Fig. 2). It is thus likely that the functional HAOs have similar secondary, tertiary, and quaternary structures. The tyrosine residue Tyr467, which is cross-linked to heme P460 in N. europaea (3, 22, 26),
FIG. 2. ClustalW alignment of Hao protein sequences from Nitrosoccus oceanii (NO0160), Nitrosospira multiformis (NMU1996), Nitrosomonas europaea (NE0962), Methylococcus capsulatus (MCA0956), and Silicibacter pomeroyi (SPOA0201). Also included are the deduced protein sequences of putative Hao proteins from Geobacter metallireducens (Gmet02001372), Desulfovibrio desulfuricans (Ddes02001487), Methanococcus burtonii (Mbur03000734), and the likely nonfunctional Hao-like sequence from Magnetococcus sp. strain Mc-1 (Mmc10221065). Amino acid residues conserved in the majority of sequences are shaded. The secondary structure of HAO from N. europaea (26) is shown beneath the alignments, with the letters H designating /H9251-helical regions and B designating /H9252-turns. Heme-binding regions are underlined.
as well as Asp267 and His268 (but not Tyr334), which are hypothesized to deprotonate or H-bond with substrate (22), were also conserved in all Hao protein sequences, as were the residues nearby. It is interesting that the sequence of amino acid residues of α-helix 1 appears to be repeated in α-helix 2 in all three Hao proteins, perhaps the result of an intragenic nucleotide sequence duplication event within hao. The least-conserved region of Hao from N. oceani, N. multiformis, and N. europaea was the region between hemes 2 and 3, which was similar in length but dissimilar in sequence, as well as the loop between α-helices 10 and 11, which was 11 amino acids longer in N. oceani than N. europaea. A relatively hydrophobic C-terminal region of Hao from N. europaea, which was not visualized in the X-ray crystal structure of this protein (26), was conserved in the Hao proteins from N. multiformis and N. oceani and has been identified by P-SORT and TopPred analyses as a membrane-spanning domain. The N. multiformis Hao had a longer C terminus by 11 residues. In the modeled HAO trimer (26), the three copies of this domain appear to be close and can be imagined to interact together in the membrane. Despite the overall sequence similarity of the Hao proteins from the three αAOB, the two betaproteobacterial Hao proteins lacked a nine-amino-acid-long fragment found between α-helices 10 and 11 in N. oceani.

The orf2 gene was found in all hao gene clusters of the three αAOB; however, the predicted protein sequence of Orf2 was the least conserved among the gene cluster products (Fig. 4A; Table 1). Orf2 was predicted by PSORT to be located in the plasma membrane.

The protein sequences of cytochrome c₅₅₄ from all three αAOB contained several conserved regions evenly distributed throughout the sequence (Fig. 4B), indicating that the overall secondary and tertiary structures of the cytochrome c₅₅₄ from N. oceani are likely similar to those from N. europaea (28). Heme-binding regions and the region between α-helix 1 and β-strand 2 were most conserved, whereas the loops joining α-helices 4 and 5 and α-helices 7 and 8 were extended in cytochrome c₅₅₄ from N. oceani compared to those from N. multiformis and N. europaea (Fig. 4B).

The cycB gene, which encodes tetraheme c cytochrome c₅₅₅₂, was present as Orf4 in the hao gene cluster of all three αAOB (Fig. 4C). As with the other genes involved in ammonia catabolism in αAOB, the genome of N. oceani contained only one copy, whereas the betaproteobacterial αAOB contained multiple copies. The N termini of the c₅₅₅₂ proteins differed in sequence in all three αAOB; however, all three were predicted by PSORT to contain a membrane anchor region. As expected, the central portion of the c₅₅₅₂ proteins, containing the heme-binding regions, was highly conserved. The C-terminal domains of c₅₅₅₂ proteins from beta-αAOB were rich in Glu and Asp and hence negatively charged, a sequence feature absent from the c₅₅₅₂ protein in N. oceani.

Comparison of the hao gene cluster proteins from αAOB with similar proteins from non-αAOB. To our surprise, genes encoding Hao-like proteins were identified in the genomes of six bacteria that do not utilize ammonia oxidation to gain energy and reductant for growth. Due to significant sequence similarity (Fig. 2; Table 1), the Hao protein sequences from Nitrosomonas, Nitrosospira, and Nitroscococcus aligned well with proteins annotated as Hao in the genomes of Methylococcus capsulatus Bath (on the chromosome) and Silicibacter pomeroyi DSS-3 (on a megaplasmid). Although the S. pomeroyi Hao protein lacked the hydrophobic C-terminal domain that is present in the Hao proteins in αAOB and M. capsulatus (which may anchor functional HAO in the plasma membrane), the presence of the critical heme and ligand-coordinating residues (Tyr467, Asp267, and His268) suggests that these proteins can assemble to bona fide trimeric HAOs, while genes encoding proteins with significant sequence similarities to other genes in the hao gene cluster were either missing (cytochrome c₅₅₄) or not near the hao gene (cytochrome c₅₅₅₂ homologues). The hao genes in M. capsulatus and S. pomeroyi were tandem arranged with orf2 genes (Fig. 4A). Hao stop and Orf2 start codons overlapped by 1 nucleotide in the M. capsulatus genome and were preceded by a purine-rich region. It is thus likely that Hao and Orf2 are coexpressed in M. capsulatus. Hao stop and Orf2 start codons were separated by intergenic sequence on the S. pomeroyi megaplasmid. The conservation of orf2 genes and the predicted association of its expression product with the plasma membrane suggest that the Orf2 protein may play a role in the biogenesis, arrangement, or stabilization of HAO.

We also identified genes in the genomes of Magnetococcus sp. strain Mc-1, Desulfovibrio desulfuricans G20, Geobacter metallireducens GS15, and Methanococcoides burtoni whose Hao protein-like expression products had lesser but significant sequence similarities to Hao (Fig. 2). These putative Hao proteins were different from the other Hao proteins in that they lacked helices 1, 2, 3, 6, and 7 and had different sixth α-helix histidines for hemes 1, 6, and 7. In addition, they contained long insertions between hemes P460 and 5 and between helix 19 and heme 8. The genes encoding the putative Hao proteins in G. metallireducens and M. burtoni were located adjacent to genes similar in sequence to one another; however, these genes had no significant sequence identity with either cycA, cycB, or orf2 genes. The putative hao gene in D. desulfuricans was found next to a gene encoding a putative tetraheme c cytochrome in the c₅₅₅₂-NrfB/H family. The Hao-like protein-encoding gene found in the genome of Magnetococcus
can very likely not assemble into a functional HAO, since its protein sequence lacks critical residues, including Tyr467, Asp267, and the histidine ligated to heme 3, all of which were present in the bona fide and putative Hao proteins.

Discussion. The high degree of conserved gene organization (hao, orf2, cycA, and cycB) and sequence in the hao gene cluster from both gamma- and beta-aAOB reported here suggests that all four genes are essential for hydroxylamine oxidation and electron transfer in aerobic aAOB. Whereas beta-aAOB, such as N. europaea and N. multiformis, contain multiple copies of the hao gene cluster, N. oceani has a single copy. Correspondingly, N. oceani has a single copy of operons encoding AMO (2) and urease (32), whereas beta-aAOB contain two copies (N. europaea) (36) or three copies (Nitrosospira spp.) (30, 40, 41) of their amo operons. Preliminary results indicate that other gamma-aAOB have single copies of their catalytic genes (2; M. G. Klotz, unpublished data). It has been argued that the presence of a single copy rather than multiple copies of ammonia catabolic genes can be an adaptation to a low but constant ammonia concentration in the environment in contrast to the highly varied concentrations of ammonia encountered by other nitrifying bacteria (40). Whereas the near identity in sequence of multiple copies of amo and hao gene clusters in a given organism may be explained by the operation of a rectification mechanism (31), the synteny of these gene clusters between organisms that do not belong to closely related taxa (i.e., the Gamma- and Betaproteobacteria) suggests that all their member genes encode proteins that interact physically (25, 33).

AMO of aAOB exhibits considerable similarity in known
aspects of structure, amino acid sequence, and enzymatic properties to the membrane-bound particulate methane monoxygenase (pMMO) of methanotrophs (7, 19, 40), which are found in the \( \alpha \)- and \( \gamma \)-subdivisions of the Proteobacteria (17). It is notable that the amino acid sequences of the Amo proteins from \( N. oceani \) are considerably more similar to the sequences of pMmo proteins of gammaproteobacterial methanotrophs than to Amo protein sequences from betaproteobacterial aAOB (40, 43). Unlike AMO from \( N. europaea \), which has a much lower affinity for methane than for ammonia, AMO from \( N. oceani \) has nearly equal affinities for methane and ammonia (58). Because \( N. oceani \) is a member of the oldest gammaproteobacterial lineage, the Chromatiales (purple sulfur bacteria), extant genes encoding AMO and pMMO may have evolved from an ancestral gene cluster that was present in the proteobacterial ancestor before the divergence of the \( \alpha \)-proteobacteria from the \( \gamma \)- and \( \beta \)-proteobacteria. Such an ancestral monoxygenase may have been able to oxidize both methane and ammonia at low rates, thereby producing low levels of the diffusible intermediate, methanol, and the poten-
tial toxin and mutagen hydroxylamine (4). This monooxygenase required oxygen and reductant to incorporate oxygen into methane or ammonia. Therefore, if this ancestral monooxygenase evolved as a copper enzyme during the radiation of *Proteobacteria*, it was useful only in locally oxic microhabitats of the globally anoxic environment. Production of hydroxylamine would have been wasteful of reductant and may have driven evolution towards a more complex and energy-conserving hydroxylamine-dehydrogenase system than the mono-heme cytochrome P460, a hydroxylamine dehydrogenase (9, 10) that may that is active during this prooxic evolutionary stage.

With the exception of cytochrome \( \epsilon_{m552}/\text{NirT} \) (8, 61), no homologues for the four proteins (HAO, Orf2, cytochrome \( \epsilon_{554} \), and cytochrome \( \epsilon_{m552} \)) have been reported in bacteria other than chemolithoautotrophic nitrifiers. In addition to reporting the sequence and structural conservation of the *hao* gene cluster in aAOB, we report here the presence of *hao*-like genes in the genomes of six prokaryotes that do not use catalytic ammonia oxidation. It has been noted that the operonic order of orthologous genes (synten) in different genomes is less preserved than their presence and that synteny is low when protein sequence identity between orthologous proteins in different genomes is lower than 50% (25, 33). It was thus significant to find that the sequences of *Hao* proteins from *N. europaea* and *N. multiformis* were 68% identical, a value similar to the identity of homologous Amo proteins (40). As also reported for Amo proteins (40), identity of the *Hao* proteins from *N. oceani* and beta-aAOB was lower and ranged between 50 and 60% (Table 1). Protein sequence similarity between *Hao* proteins from beta-aAOB was near 80%, and similarities among *Hao* proteins from *N. oceani*, *M. capsulatus*, and *S. pomeroyi* were in the lower-70 percentile (Table 1). Comparisons of all other bona fide *Hao* protein sequences yielded values in the upper-60 percentile (Table 1). As seen by analysis of the region stretching from heme-binding sites 3 to 8, the *N. oceani* *Hao* protein (NO0160) had low similarity to putative *Hao* proteins from three other non-aAOB (35%, 38%, and 41% for *G. metallireducens* [Gmet02001372], *D. sulfuricans* [Ddes02001487], and *M. burtoni* [Mbur03000734], respectively). This was largely due to numerous and large gaps. The low level of sequence identity/similarity correlated with the loss of synteny in that the *hao* genes in these organisms were not arranged with any that are found in the four-gene cluster of aAOB.

In two of the non-aAOB that contain *Hao*, *M. capsulatus* and *S. pomeroyi*, the *hao* genes were tandem arranged in an operon with an *orf2* gene. These sequences are remarkable finds, as they contribute to our understanding of several aspects of the evolutionary history of aerobic ammonia oxidation:

First, an *hao-orf2* gene tandem was identified on a megaplasmid in the marine sulfur-oxidizing alphaproteobacterium *S. pomeroyi* that appears to lack further inventory suited for *ammonia* or methane oxidation (37). Existence of a vector-borne *hao* gene strengthens the hypothesis of horizontal exchange of the *hao* gene among the *Proteobacteria*. Potential donors include the ancestors of extant marine ammonia-oxidizing purple sulfur bacteria, such as *Nitrosococcus*.

Second, the genomes of *Alphaproteobacteria* (*Magnetococcus*), iron- and sulfate-reducing *Deltaproteobacteria* (*Geobacter* and *Desulfovibrio*), and the cold-adapted and methanogenic *euryarchaeon Methanococcosoides burtoni* contained a gene that encodes a potentially functional *Hao* protein, reinforcing the hypothesis that horizontal transfer of the *hao* gene has likely occurred more than once between obligate or facultative lithrophic prokaryotes.

Third, an *hao-orf2* gene tandem was identified in the *Methylococcus capsulatus* genome. Thus, HAO may be the agent that ameliorates the potential toxicity of hydroxylamine produced by a substrate-promiscuous (ammonia-oxidizing) pMMO of this and related type X and type I methanotrophs. We note that the *Hao* proteins of the non-aAOB lack N- and C-terminal regions, which are thought to be involved in docking for electron transfer to cytochrome *c554* (27) or membrane association in the vicinity of \( \epsilon_{m552} \) and ubiquinone. The putative absence of the electron transfer to ubiquinone would disable the system of reductant for pMMO and force electrons (from hydroxylamine) to cytochrome oxidase, bypassing the proton motive force-generating cytochrome *bc_1* complex. These factors may account for the inability of *M. capsulatus* to grow on ammonia and CO_2 despite having a functional Calvin cycle (6).

Fourth, the inability of *M. capsulatus* to function as an aAOB, the presence of a diverse catabolic genomic inventory in the *M. capsulatus* genome (59), and because pMMO from *M. capsulatus* is more closely related to AMO from gamma-aAOB than to pMMO from alphaproteobacterial type II methanotrophs (40) suggest that *hao-orf2* and other genes otherwise found only in aAOB (e.g., the operon containing a multi-copper oxidase-encoding gene and *orf5*) that flanks the *amo* operon in aAOB [Klotz, unpublished] likely arrived in the ancestor of the *Methylococcaceae* by transfer from donors such as gamma-aAOB.

Cytochrome \( \epsilon_{m552} \), the product of the fourth gene in the *hao* gene cluster, has been grouped within a large family of membrane-associated tetraheme and pentaheme cytochromes that includes NirT of *Pseudomonas stutzeri* and NapC of *Paracoccus denitrificans*, several of which have been shown to carry electrons from a pool of quinols in the plasma membrane to nitrite reductase, nitrate reductase, or other periplasmic reductases during anaerobic respiration (22). It is thus not surprising that the NapC-TorC-NirT proteins have considerable overall sequence similarity except for their C-terminal regions, perhaps reflecting differences in their periplasmic electron acceptors (44), which include oxidoreductases containing heme, flavin, or molybdenopterin (22). In contrast to the NapC-TorC-NirT proteins, \( \epsilon_{m552} \) cytochromes apparently relay electrons from HAO via cytochrome *c554* to the membrane quinone/quinol pool. Our initial sequence analysis showed that cytochromes \( \epsilon_{m552} \) from aAOB have much greater sequence similarity to one another than to any member of the NapC-TorC-NirT protein family. The octaheme HAO of *N. europaea* and the formate-dependent pentaheme *c*-cytochrome nitrite reductase (NrfA) have been shown to have congruent heme stacking arrangements and secondary and tertiary structures (13, 14, 22) as well as common N-oxide intermediates. Although hydroxylamine is a bound intermediate during the reduction of nitrite to ammonia by extant NrfA proteins (13), free hydroxylamine also reacts with the enzyme. The Nrf systems were of interest to us because they operate in numerous bacteria, including the *Proteobacteria* and *Bacteroidetes* (51). Of particular interest was the Nrf system from the *Delta- and Epsilonproteobacteria*, such
as *Wolinella succinogenes* and *Sulfurospirillum deleyianum*, since they are likely more ancient and appear to be structurally simpler. Delta- and epsilonproteobacterial NrfAH systems include only two interacting proteins, whereas the gammaproteobacterial system consists of four proteins, NrfABCD (51). Hence, we hypothesized that the HAO- and NrfA-NrfH redox chains are evolutionarily related. Noting that the conserved heme stacking in HAO/NrfA was also seen in the octaheme tetrathionate reductase (Ttr) (38), we hypothesized that HAO evolved in parallel with tetrathionate reductase from a common ancestor of extant delta- and epsilonproteobacterial NrfA cytochromes. This hypothesis meant in general terms that the inventory involved in nitrification evolved partly from nitrogen and sulfur-based anaerobic respiration. There is evidence that nitrite respiration has preceded oxygen respiration (46), and it has been suggested that the genomic inventory needed for denitrification was established long before molecular oxygen was available to allow nitrification (34). 

Our analysis revealed homology between Hao, Ttr, and NrfA proteins (Fig. 5). Based on the phylogenetic tree (Fig. 5), we propose here that the octaheme Hao proteins have evolved in parallel with octaheme tetrathionate reductases from a common ancestor octaheme cytochrome c reductase (+) that emerged from the common ancestor of extant delta-epsilonproteobacterial HAO systems. Bootstrap values for the clades are shown at the branch points. Protein accession numbers are indicated.

FIG. 5. Most parsimonious phylogenetic tree constructed from 36 ClustalX-aligned sequences of multi-heme cytochrome proteins in the proposed tetrathionate reductase (Ttr), hydroxylamine oxidoreductase (Hao), and formate-dependent nitrite reductase (Nrf) protein family. The tetrathionate reductase and hydroxylamine oxidoreductase protein subfamilies likely evolved from a common ancestral octaheme cytochrome c reductase (+) that emerged from the common ancestor of extant delta-epsilonproteobacterial HAO systems. Bootstrap values for the clades are shown at the branch points. Protein accession numbers are indicated.
between HAO and NrfA, we decided to include all known NrfH sequences (and gammaproteobacterial NrfB homologues) in our phylogenetic analysis of the \textit{cm552} proteins from aAOB. By way of verifying our hypothesis, the analysis revealed homology between \textit{cm552} and NrfH/B proteins (Fig. 6), and we propose here that the gene cluster including \textit{hao} and \textit{cm552}-encoding genes in extant aAOB evolved from ancestors of the \textit{nrfA}-\textit{nrfH} operon, found in extant Delta- and Epsilonproteobacteria. The data depicted in Fig. 6 also suggest that the ancestral \textit{cm552}/NrfH/NrfB protein may itself have a common ancestor with other members of the NapC-TorC-NirT protein family before being recruited into an operon with NrfA.

The present work established for the first time a molecular evolutionary basis for the congruent heme stacking in HAO, cytochrome \textit{c} nitrite reductase, and tetrathionate reductase. Despite conserved heme stacking, secondary and tertiary structure, HAO and cytochrome \textit{c}_{552} of \textit{Nitrosomonas} (27, 28) and the tetrathione portion of fumarate reductase (49) did not display high enough sequence similarity with Hao to justify their inclusion into a phylogenetic analysis. In marked contrast to the observed distribution of \textit{hao-orf2} gene tandems in aAOB and non-aAOB, cytochrome \textit{c}_{552} (\textit{cycA}) has been found only in aAOB. The ancestry of the \textit{cycA} (cytochrome \textit{c}_{552}) gene of the \textit{hao} gene cluster and the tetra-\textit{c}-heme domain of the periplasmic fumarate reductase remain unknown.

Our data and their interpretation suggest that the \textit{hao} gene cluster has evolved as a structural and functional unit from pentaheme nitrite reductase under pressure of increasing hydroxylamine concentrations associated with more efficient ammonia catabolism. Factors including the intersubunit covariant cross-linking of a tyrosine to the catalytic heme and the expansion of the pentaheme to the octaheme configuration likely contributed to the conversion of an ancient electron disposal system into a mechanism capable of using hydroxylamine as an electron donor at the ubiquinone level. We propose that expansion of the ancient \textit{hao-cycB} tandem into a gene cluster with four genes (\textit{hao-orf2-cycA-cycB}) took place rather early in purple sulfur bacterial ancestors of extant gamma-aAOB in the family \textit{Chromatiaceae}. The conserved organization and high sequence similarity of the \textit{hao} gene cluster in a single family each of the \textit{Gamma-} and \textit{Betaproteobacteria} suggests that the \textit{hao} gene cluster arrived in the common ancestor of the beta-aAOB likely by a lateral gene transfer event before extensive speciation of these nitrifiers. Because one of three copies each

---

**FIG. 6.** Unrooted most parsimonious phylogenetic tree constructed from 56 ClustalX-aligned protein sequences of the membrane-associated tetraheme cytochromes in the proposed \textit{cm552}/NrfB/H\&NapC/TorC/NirT protein family. The clade with members of the Nap/C/TorC/NirT protein subfamily was collapsed to focus on the clade with the members of the \textit{cm552}/NrfB/H protein subfamily in the phylogram. Bootstrap values for the clades are shown at the branch points. See Materials and Methods for accession numbers not provided in the figure.
of the hao and amo gene clusters are adjacent in the genome of *N. multiformis*, it is possible that amo and hao gene clusters were jointly transferred as a structural and functional metabolic unit before operon duplication led to multiple copies in betaproteobacterial aAOB. The observed differential regulation of hao and cycA genes in *N. europaea* (8, 21, 47), the unique presence of orf3 in the hao gene cluster in *N. multiformis*, and the low sequence identity (below 50%) between Orf2 protein homologues (Table 1) suggest that the region between the hao and cycA genes in the hao gene cluster has experienced change at higher rates since the hao cluster evolved independently in the gamma- and betaproteobacterial lineages of the aAOB.

In addition to improving our understanding of the molecular evolution of nitrification, our data have significance for molecular ecology. The presence of unique and highly conserved regions within the Hao and cytochrome *c* subunit proteins of *N. europaea*, *N. oceani*, and *N. multiformis* suggests that cycA genes could be used to design new PCR primers and molecular probes to amplify and detect this gene in order to better assess and characterize the diversity of aerobic aAOB in environmental samples. To date, the presence of aAOB within both the Betaproteobacteria and Gammaproteobacteria precludes that a single 16S rRNA sequence-based set of primers (or probe) can be employed for detection/identification of both beta- and gamma-proteobacterial aAOB. Further, the rather low variability of 16S rRNA genes within the betaproteobacterial aAOB makes it difficult to separate the sequences of PCR products amplified from different strains of aAOB (40). A partial solution to this problem has been the design of PCR primers to amplify that part of the amoA gene which encodes the 27-kDa subunit of AMO in aAOB. However, the amoA gene sequences of gammaproteobacterial aAOB are more similar to gammaproteobacterial *pmoA* sequences than they are to amoA sequences of Betaproteobacteria (43); hence, different PCR primer sets must be used to obtain amoA-based amplicons from Betaproteobacteria and Gammaproteobacteria. The present demonstration that cycA is unique to aAOB suggests that PCR primer sets and probes based on cycA gene sequences may be very useful tools for estimating aAOB diversity in the environment.

ACKNOWLEDGMENTS

We thank Mike Morton for culturing and Vilasak Thamvongsa for preparation of genomic DNA from *N. oceani* C-107SW, Brian Jackson for help in primer design, David Arciero for advice on the protein isolation, and Eric Ecleston for preparing protein samples for sequencing. Our thanks also go to Amal El-Shiekh for culturing of and genomic DNA isolation from *N. oceani* ATCC 19707.

This project was supported, in part, by DOE’s G2L initiative, incentive funds provided to M.G.K. by the University of Louisville VPR office, and funds to A.B.H. provided by DOE (DE-FG02-95ER20191.A004) and NSF (MCB-0093447).

ADDENDUM IN PROOF

Genes of putative Hao proteins not affiliated with orf2 genes have been identified in three additional genomes: *Rhodofex reriireducens* (IMG5004557640), *Anaeromyxobacter dehalogenans* (IMG4000925800), and *Desulfuromonas acetoxidans* (IMG400047730) (M. G. Klotz, unpublished data).

REFERENCES

27. Iverson, T. M., D. M. Arciero, A. B. Hooper, and D. C. Rees. 2001. High-
resolution structures of the oxidized and reduced states of cytochrome c\textsubscript{554} from \textit{Nitrosomonas europaea}. J. Biol. Inorg. Chem. 6:390–397.


