Research Article

A Novel Pathway for the Biosynthesis of Heme in Archaea: Genome-Based Bioinformatic Predictions and Experimental Evidence

Sonja Storbeck,1 Sarah Rolfes,1 Evelyne Raux-Deery,2 Martin J. Warren,2 Dieter Jahn,1 and Gunhild Layer1

1 Institute for Microbiology, Technical University of Braunschweig, Spielmannstraße 7, 38106 Braunschweig, Germany
2 School of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, UK

Correspondence should be addressed to Gunhild Layer, g.layer@tu-braunschweig.de

Received 1 September 2010; Accepted 12 October 2010

Academic Editor: Stefan Spring

Copyright © 2010 Sonja Storbeck et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Heme is an essential prosthetic group for many proteins involved in fundamental biological processes in all three domains of life. In Eukaryota and Bacteria heme is formed via a conserved and well-studied biosynthetic pathway. Surprisingly, in Archaea heme biosynthesis proceeds via an alternative route which is poorly understood. In order to formulate a working hypothesis for this novel pathway, we searched 59 completely sequenced archaeal genomes for the presence of gene clusters consisting of established heme biosynthetic genes and colocalized conserved candidate genes. Within the majority of archaeal genomes it was possible to identify such heme biosynthesis gene clusters. From this analysis we have been able to identify several novel heme biosynthesis genes that are restricted to archaea. Intriguingly, several of the encoded proteins display similarity to enzymes involved in heme d1 biosynthesis. To initiate an experimental verification of our proposals two Methanosarcina barkeri proteins predicted to catalyze the initial steps of archaeal heme biosynthesis were recombinantly produced, purified, and their predicted enzymatic functions verified.

1. Introduction

Heme, a modified tetapyrrole, acts as an essential prosthetic group in many enzymes, sensory, and regulatory proteins. Hemes are also essential components of electron transport chains driving aerobic and anaerobic respiration and photosynthesis in almost all living organisms. Consequently, heme-containing proteins are found in all three domains of life, the Eukaryota, the Bacteria, and the Archaea. The biosynthesis of this important and ubiquitously distributed molecule has been intensively studied in eukaryotic and bacterial organisms, but little is known about heme biosynthesis in archaea. It is now well established for bacteria and eukarya that heme biosynthesis proceeds along a conserved pathway with highly related enzymes and identical biosynthetic intermediates (Figure 1(a)) [1]. Heme synthesis represents just one component of a larger, branched tetapyrrole biosynthesis pathway, which is also responsible for the synthesis of chlorophylls, bacteriochlorophylls, cobalamin, siroheme, heme d1 and coenzyme F430 (Figure 1(b)) [2].

The common precursor for the formation of heme and all other tetapyrroles is 5-aminolevulinic acid (ALA). Depending on the organism this molecule is either synthesized through the condensation of glycine and succinyl-CoA (Shemin pathway) by ALA synthase (HemA) or in a two-step enzymatic process from glutamyl-tRNA via the intermediate glutamate-1-semialdehyde (GSA) by glutamyl-tRNA reductase (HemA) and GSA-2,1-aminomutase (HemL) (C3-pathway) [3, 4]. Eight molecules of ALA are then converted into uroporphyrinogen III (UROGEN), the first cyclic tetapyrrole of the pathway, in three consecutive enzymatic steps. First, two ALA molecules are condensed by porphobilinogen synthase (HemB) to the pyrrole derivative porphobilinogen (PBG) [5].
Figure 1: Tetrapyrrole biosynthesis pathways. (a) Heme biosynthesis in most bacteria and the Eukaryota. The first common precursor in the classical heme biosynthesis pathway is ALA of which eight molecules are converted into UROGEN in three consecutive enzymatic steps. UROGEN is then further converted into heme through successive modifications of the macrocycle side chains and finally iron insertion. The enzymes involved in the classical heme biosynthesis are glutamyl-tRNA reductase (GluTR), glutamate-1-semialdehyde-2,1-aminomutase (GSAM), 5-aminolevulinic acid synthase (ALAS), porphobilinogen synthase (PBGS), porphobilinogen deaminase (PBGD), uroporphyrinogen III synthase (UROS), uroporphyrinogen III decarboxylase (UROD), oxygen-dependent coproporphyrinogen III oxidase (CPO), coproporphyrinogen III dehydrogenase (CPDH), oxygen-dependent and oxygen-independent protoporphyrinogen IX oxidase (PPO), and ferrochelatase (FC). The corresponding bacterial gene names are denoted in brackets below the enzyme names. (b) Overview of the different branches of the tetrapyrrole biosynthesis pathway. The last common precursor for the formation of all tetrapyrroles is UROGEN. Hemes and (bacterial)chlorophylls share PROTO as their last common intermediate. Siroheme, cobalamin, coenzyme F₄₃₀, and heme d₁ are all biosynthesized via precorrin-2. In the Archaea and some bacteria an alternative heme biosynthesis pathway exists in which the heme is biosynthesized from precorrin-2 via yet unknown intermediates.
uroporphyrinogen III (UROGEN) is formed by cyclization of the pre-uroporphyrinogen by UROGEN synthase (HemD) [6]. The intermediate uroporphyrinogen III represents the last common precursor for all tetrapyrroles and is therefore an important branchpoint of the pathway. One of the diverting biosynthetic routes leads to the formation of hemes and (bacterio)chlorophylls via the intermediate coproporphyrinogen III (COPROGEN) and the other represents the first step of cobalamin, siroheme, heme \( d_1 \), and coenzyme F\(_{430} \) biosyntheses via the common intermediate precorrin-2. COPROGEN and precorrin-2 are formed from UROGEN by the key branchpoint enzymes uroporphyrinogen III decarboxylase (HemE) and S-adenosyl-L-methionine-dependent uroporphyrinogen III methyltransferase (SUMT), respectively. Eukaryotic and bacterial heme biosynthesis further proceeds via the conversion of COPROGEN into protoporphyrinogen IX (PROTOGEN) by coproporphyrinogen III oxidase (HemF) or dehydrogenase (HemN) and the subsequent oxidation by protoporphyrinogen IX oxidase (HemY, HemG) to protoporphyrin IX (PROTO) [7]. Finally, the insertion of ferrous iron into PROTO by ferrochelatase (HemH) yields the end product heme [8]. All heme biosynthetic enzymes have been purified from many different eukaryotic and bacterial organisms and biochemically characterized [1]. The corresponding genes (\( \text{hemA}, \text{hemB}, \text{hemC}, \text{hemD}, \text{hemE}, \text{hemF}, \text{hemG} \)) all have all been cloned and sequenced [9]. In Figure 1(a) the commonly used gene designations for all bacterial heme biosynthesis genes are given together with their corresponding enzyme names.

In two independent bioinformatics studies in 2002 and 2008 the distribution of heme biosynthetic genes in prokaryotic organisms was investigated by analysis of the currently available sequenced microbial genomes [10, 11]. It was found that almost all bacteria which synthesize heme \( \text{hem} \) possess the complete set of \( \text{hem} \) genes (i.e., \( \text{hemA}^A \) or \( \text{hemA}^B \), \( \text{hemL} \), \( \text{hemB} \), \( \text{hemC} \), \( \text{hemD} \), \( \text{hemE} \), \( \text{hemF} \) and/or \( \text{hemN} \), \( \text{hemY} \) or \( \text{hemG} \), \( \text{hemH} \)). In contrast, some bacteria (e.g., \text{Clostridia} and \text{Desulfovibrio} species) and almost all archaea were found to possess only the genes encoding the enzymes required for UROGEN formation (\( \text{hemA}^B \), \( \text{hemB} \), \( \text{hemC} \), \( \text{hemD} \)) and lacked the genes encoding the enzymes necessary for the conversion of UROGEN into heme [10, 11]. Such a finding can be explained by (i) the possibility that these organisms have no need for heme and require the initial genes for cobalamin, siroheme, coenzyme F\(_{430} \), or heme \( d_1 \) formation, (ii) the possibility that they take up heme from the environment, or (iii) the existence of an alternative, yet unknown, heme biosynthesis pathway. For many free-living bacteria and archaea heme uptake is not very likely since heme is simply not available in their environment. However, a few examples of such cases exist in archaea and bacteria [12, 13]. Moreover, several pathogenic bacteria take up heme from their host and use it as an iron source. However, these bacteria often possess an intact heme biosynthetic apparatus [11].

It has been known for some time that \text{Desulfovibrio} \( \text{D. vulgaris} \) and \text{Methanosarcina barkeri} it was shown experimentally that an alternative heme biosynthesis pathway must exist. In these cases \textit{in vivo} labeling studies demonstrated that their hemes contain methyl groups on rings A and B that are derived from methionine (via S-adenosyl-L-methionine) and not from ALA as is the case for hemes synthesized \textit{via} the classical pathway [23, 24]. Further, in \textit{D. vulgaris} sirohydrochlorin (the oxidized form of precorrin-2), 12,18-didecarboxysirohydrochlorin, coproporphyrin III, and PROTO were isolated as potential heme biosynthesis intermediates [25]. Thus, the alternative heme biosynthesis pathway seems to branch off the classical pathway at the stage of UROGEN. In the first step of the alternative route UROGEN is methylated at rings A and B by a SUMT-like enzyme to yield precorrin-2 (Figure 1(b)). This SUMT-dependent methylation of UROGEN is also required for the biosyntheses of cobalamin, siroheme, heme \( d_1 \), and coenzyme F\(_{430} \) (Figure 1(b)). Recently, in \textit{D. vulgaris} a bifunctional enzyme carrying both UROGEN synthase and SUMT activities and a precorrin-2 dehydrogenase (PC2-DH) catalyzing precorrin-2 oxidation to sirohydrochlorin were biochemically characterized [26]. Both enzymes are probably involved in the alternative heme biosynthesis pathway in this organism.

Besides the \textit{in vivo} labeling study in \textit{M. barkeri} the alternative heme biosynthesis pathway has not been investigated in archaea, so far. In the last few years the number of completely sequenced archaeal genomes has greatly increased and therefore we decided to start our investigation of archaeal heme biosynthesis with the search for potential heme biosynthesis gene clusters within these genomes. We found that many archaea indeed contain gene clusters consisting of the known early heme biosynthesis genes (\( \text{hemA}^B \), \( \text{hemL} \), \( \text{hemB} \), \( \text{hemC} \), \( \text{hemD} \)) and of “\text{irn-like}” genes which encode proteins homologous to proteins involved in heme \( d_1 \) biosynthesis in denitrifying bacteria. Moreover, very often the genes encoding a putative SUMT and a potential PC2-DH were found localized in these archaeal heme biosynthesis gene clusters. Here, the predicted SUMT and PC2-DH from \textit{M. barkeri} were recombinantly produced, purified and shown \textit{in vitro} to carry SUMT and PC2-DH activity, respectively.

2. Materials and Methods

2.1. Chemicals. All chemicals, reagents, and antibiotics were obtained from Sigma-Aldrich (Taufkirchen, Germany) or Merck (Darmstadt, Germany). DNA polymerase, restriction endonucleases, and PCR requisites were purchased from New England Biolabs (Frankfurt a.M., Germany). Oligonucleotide primers were obtained from metagenom international AG (Martinsried, Germany). PCR purification and gel extraction Kits were purchased from Qiagen GmbH (Hilden, Germany). Ni Sepharose 6 Fast Flow was purchased from GE Healthcare (München, Germany). Uroporphyrin III was obtained from Frontier Scientific Europe (Carnforth, UK).
2.2. Construction of Vectors for Recombinant Protein Production. The gene mbaA1461 encoding a potential PC2-DH from M. barkeri was PCR amplified using the primers 01CySGN_Mba_BamHIFw (GAA GGG ATC CAA ATA ATT TTC) and 02CySGN_Mba_Not1Lev (GAA GGC GCC CGG TTA ACG GTT GCT GTT CAC) containing BamHI and NotI restriction sites (underlined) and cloned into appropriately cut pET-Duet1 (Novagen, Darmstadt, Germany) to generate pET-Duet_mbaA1461. The plasmid pMA_mbaA1461 (GeneART, Regensburg, Germany), which contains a synthetic copy of the mbaA1461 gene, codon-optimized for expression in *Escherichia coli*, was used as the DNA template for PCR.

The gene mbaA1791 encoding a putative SUMT from *M. barkeri* was PCR amplified using the primers MbarcobA-ATG (CAC ATA TG M. barkeri) and MbarcobA-Stop (AGG ATC CAA ATA ATT TTC) and MbarcobA-Stop (AGG ATC CAA ATA ATT TTC) and MbarcobA-Stop (AGG ATC CAA ATA ATT TTC) BamHI and NotI restriction sites (underlined) from genomic *M. barkeri* DNA. The resulting PCR fragments and the vector pET14b were subsequently digested with Ndel and BamHI and ligated to generate pET14b_mbaA1791.

2.3. Bacterial Strains and Growth Conditions. *E. coli* DH10B was used as the host for cloning. For production of recombinant proteins the *E. coli* strains BL21 (DE3) and BL21 Star (DE3) pLysS were used, respectively. The expression vector pET14b_mbaA1791 was transformed into *E. coli* BL21 Star (DE3) pLysS. The vector pET-Duet_mbaA1461 was transformed into *E. coli* BL21 (DE3). For recombinant protein production the *E. coli* strains carrying the corresponding vectors were grown at 37°C in LB-medium containing appropriate antibiotics. Protein production was induced by adding 50 μM isopropyl isopropyl-β-D-thiogalactopyranosid (IPTG) to the cultures at an optical density at 578 nm of 0.6. The *E. coli* BL21 (DE3) strain containing pET-Duet_mbaA1461 was further cultivated at 37°C for 4h. The *E. coli* BL21 Star (DE3) pLysS strain containing pET14b_mbaA1791 was further cultivated at 17°C for 18h. The cells were harvested by centrifugation and stored at −20°C.

2.4. Purification of Enzymes and Tetrapyrroles. The affinity chromatographic purification of the recombinant enzymes was performed as described before [27] with minor changes. Briefly, for resuspension of the *E. coli* cells, harbouring the produced recombinant protein, buffer A (50 mM Tris/HCl (pH 7.5), 300 mM NaCl, 10% (w/v) glycerol) was used containing 1 mM phenylmethanesulfonyl fluoride. The cells were disrupted using a French press (1000 p.s.i.) and the soluble protein fraction was obtained by ultracentrifugation (60 min, 175000 × g, 4°C). The supernatant was applied to 1 mL of Ni Sepharose 6 Fast Flow (GE Healthcare). The flow-through containing the tetrapyrroles accumulated during *in vivo* protein production was applied to a 1 mL silica gel 100 G18-reversed phase column (Sigma- Aldrich) and the tetrapyrroles were extracted as described before [27]. The Ni-resin with bound proteins was washed extensively with buffer A. After a preelution step with buffer A containing 20 mM imidazole the recombinant protein was eluted with buffer A containing 300 mM imidazole. Immediately after elution a buffer exchange was performed in an anaerobic chamber (Coy Laboratories, Grass Lake, MI, USA) by passing the protein solution through a NAP-25 column (GE Healthcare) that had been equilibrated with degassed buffer A containing 5 mM dithiothreitol. The protein was stored at −20°C until required.

2.5. Determination of Protein Concentration. The Bradford Reagent (Sigma-Aldrich) was used to determine protein concentrations, according to the manufacturer’s instructions, using BSA as a standard.

2.6. Molecular Mass Determination. In order to determine the oligomeric state of proteins a gel permeation chromatography was performed using a Superdex 200 10/30 GL column with an AKTA Purifier system (GE Healthcare). The column was equilibrated with buffer A containing 5 mM dithiothreitol and calibrated using four standard proteins: cytochrome c, conalbumin, alcohol dehydrogenase, β-amylase (Kit Gel Filtration molecular weight markers (Sigma-Aldrich)). Protein samples (1 mg mL⁻¹) were applied to the column and the elution of proteins was monitored by determination of the absorption of the eluate at 280 nm as described before [27].

2.7. In Vitro Enzyme Activity Assays. *In vitro* enzyme activities of the recombinantly produced and purified MbaA1791 and MbaA1461 proteins were measured using a coupled enzyme assay as described before [27]. The assay was performed in an anaerobic chamber (Coy Laboratories) under strictly anaerobic conditions (O₂ = 0 ppm). The substrate uroporphyrinogen III was generated enzymatically from 1 mM ALA using purified HemB (0.14 μM) from *Psuedomonas aeruginosa*, HemC (0.15 μM), and HemD (0.17 μM) both from *Bacillus megaterium* in a final volume of 1 mL of degassed buffer B containing 50 mM Tris/HCl (pH 8.0), 100 mM KCl, 5 mM MgCl₂, and 50 mM NaCl. In order to investigate the activity of MbaA1791 the enzyme was added to a final concentration of 1.5 μM, and SAM as methyl donor was added to a final concentration of 200 μM. In order to determine the activity of MbaA1461, precorrin-2 was generated using *P. aeruginosa* NirE as SUMT [27] at a concentration of 1.5 μM. The MbaA1461 was added to a final concentration of 1.5 μM with 100 μM NAD⁺. The reaction mixtures were incubated overnight at 37°C in the dark. UV-visible spectra of the assay mixtures were recorded on a V-650 spectrophotometer (Jasco, Gross-Umstadt, Germany).

2.8. Bioinformatics Analysis. For the analysis and comparison of archaeal genomes the “Microbial Genome Database for Comparative Analysis” (http://mbgd.genome.ad.jp/) was used [28–30]. This database contains a total of 68 completely sequenced archaeal genomes. Of these 68 genomes we initially chose one for each species analyzed, that is, different strains within one species were not included, which left
us with 59 genomes. The archaebial species whose genomes were chosen are listed in Table 1. We also included the genomes of E. coli, P. aeruginosa, and D. vulgaris as positive and negative controls for our search. First, the database was searched for the known early heme biosynthesis genes of E. coli (hemA\(^B\), hemL, hemB, hemC, hemD) and then the database was used to find “orthologous clusters” in the archaebial genomes. With the “orthologous cluster” tool all homologous hem genes in the chosen genomes were displayed and a multiple genome map comparison could be viewed. For the clustering parameters we chose the default values of the database. Using the “multiple genome map comparison” tool we identified gene clusters with similar gene organization in the neighborhood of the known early heme biosynthesis genes in the archaebial genomes.

### 3. Results and Discussion

3.1. The Late Heme Biosynthesis Genes Are Missing in Archaeal Genomes

In order to identify potential heme biosynthesis gene clusters in the 59 archaebial genomes analyzed we first checked for the presence and genomic localization of the early heme biosynthesis genes hem\(^A\), hemL, hemB, hemC, and hemD. Next, we inspected the genes located in their direct neighborhood using the MBGD database. Out of the 59 archaebial genomes included in this study we found 12 genomes which do not contain any obvious hem gene (Table 1). These organisms apparently do not synthesize tetrapyrroles de novo unless via a completely novel pathway. Alternatively, these members of the Archaea do not need heme and other tetrapyrroles, respectively, or they are able to take up these compounds from their environment as described previously [12]. For example, it was reported recently that many archaebial species possess genes that encode putative homologs of the prokaryotic BtuFCD system for cobalamin uptake [31]. Here, in the 12 genomes that lack the hem genes we also found btuFCD homologs with the exception of the Korarchaeum cryptofilum and the Nanoarchaeum equitans genomes (not shown).

However, in the majority (47) of the studied genomes we found all five hem genes (hemA\(^B\), hemL, hemB, hemC, hemD) whose encoded protein products are known to be responsible for formation of the tetrapyrrole precursor UROGEN. One exception is the genome of Aeropyrum pernix which is missing a recognizable hemD gene. As already observed before [10, 11] we failed to detect the genes hemE, hemF/N, hemG/Y, hemH encoding the known late heme biosynthesis enzymes catalyzing the conversion of UROGEN into heme. Exceptions from this rule came from the analysis of the Picrophilus torridus, Thermoplasma acidophilum, and Thermoplasma volcanium genomes. The genomes of these three species contain hemE and hemH genes encoding UROGEN decarboxylase and ferrochelatase, respectively, as highlighted in earlier studies [10, 11]. However, no genes encoding recognizable COPROGEN oxidases/dehydrogenases (hemF/N) or PROTOGEN oxidases (hemG/Y) were found. Thus, most archaebial possess the genetic potential for synthesizing UROGEN from glutamyl-tRNA via the intermediates GSA, ALA, PBG, and pre-uroporphyrinogen and appear to have genes for heme-containing proteins. Therefore, to make heme they must transform the UROGEN by a novel pathway that differs from the known heme biosynthesis route. This is consistent with the observation that the methanogenic archaebion M. barkeri synthesizes its heme via the intermediate precorrin-2 [24].

3.2. Archaeal Hem Genes Are Clustered with SUMT and PC2-DH Genes

Upon closer bioinformatical inspection of the chromosomal organization of the detected archaebial hem genes we found that they are often located within gene clusters comprising two or more hem genes (Table 1 and Figure 2). Interestingly, within these hem gene clusters we also detected genes potentially encoding a SUMT and a precorrin-2 dehydrogenase (PC2-DH). SUMT proteins catalyze the S-adenosyl-L-methionine-dependent methylation of UROGEN on rings A and B at positions 2 and 7 to give precorrin-2 (Figure 1(b)). The PC2-DH proteins in turn oxidize the precorrin-2 to sirohydrochlorin in a NAD\(^+\)-dependent reaction. Unfortunately, the nomenclature for these two genes in the MBGD database is quite inconsistent as the SUMT encoding gene is sometimes named colbA, cysG-I, cysG-2, cysG, uroM, or hemX, and the gene encoding PC2-DH is referred to as sirC, hemX, cysG, or cysG1. In the following we will refer to the genes encoding the methyltransferase and the dehydrogenase simply as the SUMT and PC2-DH genes, respectively. In archaebial a SUMT is probably required for the synthesis of all tetrapyrroles including heme, cobalamin, siroheme, and coenzyme F\(_{430}\) [24, 32]. The PC2-DH has been shown conclusively to be involved in siroheme and anaerobic cobalamin formation in bacteria [33, 34] and will probably fulfill this function also in archaebial. Its involvement in heme and coenzyme F\(_{430}\) biosyntheses has not yet been demonstrated and requires further experimental evidence.

Although both enzymes are required for the formation of all these different tetrapyrroles in archaebial, it is interesting to note that almost all archaebial species possess only one SUMT and one PC2-DH gene. The only exception from this rule is Archaeoglobus fulgidus which possesses two SUMT genes. As outlined above, the SUMT and PC2-DH genes are often clustered on the genomes with the early hem genes. This clustering of genes encoding the enzymes responsible for the transformation of ALA into precorrin-2 or sirohydrochlorin provides the organisms with the possibility of coordinated gene expression and production of enzymes catalyzing consecutive biosynthetic steps. However, such heme biosynthesis gene clusters were not found in all archaebial genomes. For some of the investigated species, like Ignicoccus hospitalis and Caldivirga maquilingensis, the heme biosynthesis genes were found scattered randomly throughout the genome (Table 1).

3.3. Potential Involvement of nir-Like Genes in Archaeal Heme Biosynthesis

Interestingly, in 32 of the archaebial genomes that contain the early hem genes we also found so-called nir genes (nirD, nirH, nirF) co-localized in large gene clusters...
| Organism | henA<sup><i>b</i></sup> | henL | henB | henC | henD | SUMT | PC2-DH | nirD | nirH | nifJ | nifJ2 | heme<sup><i>b</i></sup> | B12<sup><i>c</i></sup> | siroheme<sup>d</sup> | F<sub>430</sub> |
|----------|----------------|------|------|------|------|------|--------|------|------|------|------|----------|--------|-----------|-------|
| 1 Aeropyrum pernix K1 | APL_2296<sup>c</sup> | 2299.1 | 3001.1 | 2298.1 | 2236.1 | 1491.1 | 1497.1 | 1497.1 | 1497.1 | 1655 | 1652 | + | — | — | 1 |
| 2 Desulfovibrio kamchatkensis 1221n | | | | | | | | | | | | | | | |
| 3 Ignicoccus hospitalis KIN4/1 | IGNL_0512 | 0.083 | 0.0324 | 0.0684 | 0.0909 | 0.089 | 0.0843 | 0.0839 | 0.0839 | 0.0397 | 0.0630 | — | — | — | 3 |
| 4 Staphylothermus marinus F1 | | | | | | | | | | | | | | | |
| 5 Hyperthermus butylicus DSM 5456 | HBU01_0206 | 0.0837 | 0.1386 | 0.0836 | 0.0834 | 0.0835 | 0.0510 | 0.0036 | 0.0036 | 0.0035 | 0.0037 | — | — | — | 5 |
| 6 Metallosphaera sedula DSM 5348 | MSED_0214 | 0.0216 | 0.0217 | 0.0218 | 0.0608 | 0.0213 | 0.0013 | 0.0013 | 0.0512 | 0.0511 | + | + | — | 6 |
| 7 Sulfolobus acidocaldarius DSM 639 | SACL_0777 | 0.0779 | 0.0778 | 0.0780 | 0.0781 | 0.0914 | 0.0776 | 0.0711 | 0.0711 | — | — | + | + | + | 7 |
| 8 Sulfolobus islandicus M.14.25 | M1425_1955 | 1.953 | 1.954 | 1.952 | 1.951 | 0.0226 | 1.956 | 1.894 | 1.894 | 1.048 | 1.047 | — | + | — | 8 |
| 9 Sulfolobus solfataricus P2 | SSO01_180 | 0.0182 | 0.0181 | 0.0183 | 0.0184 | 2435 | — | 0.0245 | 0.0245 | 1631 | 1632 | 1840 | 1839 | + | + | 9 |
| 10 Sulfolobus tokaloai 7 | ST0212 | 0.0215 | 0.0214 | 0.0217 | 0.0218 | 0.0563 | 0.0211 | 0.0293 | 0.0293 | 0.0127 | 0.0126 | + | + | + | 10 |
| 11 Thermofilum pendens Hrk 5 | | | | | | | | | | | | | | | |
| 12 Caldicivibrio maquilensis IC-167 | CMAQ_1730 | 1.413 | 1.740 | 1.121 | 1.001 | 1.313 | 1.731 | 0.0150 | 0.150 | 1.900 | 1.347 | — | — | — | 12 |
| 13 Pyrobaculum aerophilum IM2 | PAE0601 | 0.0594 | 0.0583 | 0.0580 | 0.0589 | 0.0590 | 0.0585 | 0.0581 | 0.0581 | 0.0596 | 0.0579 | + | + | + | 13 |
| 14 Pyrobaculum aerophilum DSM 13514 | PAR2261 | 2.254 | 2.247 | 2.245 | 2.250 | 2.251 | 2.248 | 2.246 | 2.246 | 2.255 | 2.244 | + | + | + | 14 |
| 15 Pyrobaculum calidiphilum ICM 11548 | PICAL_0234 | 1.177 | 1.179 | 1.170 | 1.172 | 1.173 | 1.170 | 1.1708 | 1.1708 | 1.1708 | 1.1706 | + | + | + | 15 |
| 16 Pyrobaculum islandicum DSM 4184 | PISL_0096 | 0.0114 | 0.0042 | 0.0050 | 0.0116 | 0.0115 | 0.0119 | 0.0044 | 0.0044 | 0.0113 | 0.0051 | — | — | — | 16 |
| 17 Thermotoga neapolitana V24Sta | TNEU_1917 | 1.900 | 0.957 | 0.968 | 1.898 | 1.899 | 1.896 | 0.0962 | 0.0962 | 1.901 | 0.0969 | — | — | — | 17 |
| 18 Archaeoglobus fulgidus DSM 4304 | AF1975 | 1241 | 1974 | 1242 | 0.116 | 1243 | 1592 | 1594M | 1594M | 2413 | 1125 | + | — | — | 18 |
| 19 Haloarcula marismortui ATCC 43049 | RRNAC1708 | 2628 | 2601 | 3086 | 3086 | 3087 | 1709 | 1711 | 1711 | 1711 | 3489 | 1363 | + | + | + | 19 |
| 20 Halobacterium sp. NRC-1 | VNG174G | 2326G | 2322G | 2330G | 2332G | 2331G | 1775C | 1776G | 1776G | 1185G | 184G | + | + | — | 20 |
| 21 Halococcus halococcoides DSM 12286 | HMUK_1612 | 0.0914 | 0.0925 | 0.0935 | 0.1105 | 0.1103 | 0.1104 | 0.1613 | 0.1614 | 0.1614 | 0.1679 | 0.1984 | + | + | + | 21 |
| 22 Haloquadatum walsbyi DSM 16790:HBQ001 | HQ3336A | 3447A | 3443A | 3450A | 3452A | 3451A | 3335A | 3334A | 3334A | — | — | + | + | + | 22 |
| 23 Halorhabdus utahensis DSM 12940 | HUTA_2825 | 0.1973 | 0.1362 | 0.0927 | 0.0928 | 0.1761 | 0.1755 | — | — | 0.0031 | 0.0734 | + | + | — | 23 |
| 24 Halorubrum lacusprofundi ATCC 49239 | HLAC_2132 | 0.2620 | 0.0015 | 0.2252 | 0.2254 | 0.2253 | 0.2313 | 0.2310 | 0.2310 | 0.1215 | 0.2081 | + | — | — | 24 |
| 25 Natronomonas pharaonis DSM 2160 | NP4502A | 1246A | 0.902A | 1326A | 1330A | 1328A | 4500A | 4498A | 4498A | 1546A | 1542A | + | — | + | 25 |
| Organism                                      | hemA<sup>a</sup> | hemL | hemB | hemC | hemD | SUM<sup>T</sup> | PC2-DH | nirD | nirH | nir1 | nir2 | heme<sup>b</sup> | B<sub>12</sub> | sinoheme<sup>c</sup> | F<sub>430</sub> |
|----------------------------------------------|------------------|------|------|------|------|-----------------|--------|------|------|------|------|-----------------|---------|---------------------|-----------|
| 26 Methanosrevibacter smithii ATCC 35061     | MSML<sub>0967</sub> | 1233 | 1476 | 8881 | 1504 | 1550           | 0968   | —    | —    | —    | —    | —                | +       | +                   | 26        |
| 27 Methanosphaera stadtmannae DSM 3091       | MSP<sub>1408</sub> | 1180 | 0416 | 1332 | 1191 | 1192           | 1407   | —    | —    | —    | —    | —                | +       | +                   | 27        |
| 28 Methanothermoautotrophicus delta H<sup>+</sup> | MTH1012          | 228  | 744  | 874  | 166  | 167            | 1013   | —    | —    | —    | —    | +                | +       | +                   | 28        |
| 29 Methanoactoococcus fervens AG86           | MEFER<sub>0496</sub> | 1387 | 0258 | 0770 | 1494 | 0849           | 0978   | —    | —    | —    | —    | —                |         | +                   | 29        |
| 30 Methanoacccus jannaschii DSM 2661         | MJ0143           | 0603 | 0643 | 0569 | 0994 | 0965           | 0140   | —    | —    | —    | —    | +                | +       | +                   | 30        |
| 31 Methanoactoococcus vulcanii M7            | METV<sub>1458</sub> | 0867 | 0542 | 0749 | 0380 | 0448           | 0492   | —    | —    | —    | —    | —                |         | +                   | 31        |
| 32 Methanoacccus aerolicus Nankai-3          | MAE<sub>Q0052</sub> | 1208 | 0916 | 1324 | 0075 | 0861           | 0053   | —    | —    | —    | —    | +                | +       | +                   | 32        |
| 33 Methanoacccus maripaludis S2              | MMP0088          | 0224 | 1258 | 0872 | 0394 | 0966           | 0089   | —    | —    | —    | —    | +                | +       | +                   | 33        |
| 34 Methanoacccus vanielii SB                 | MEVAN<sub>1100</sub> | 1237 | 0571 | 0048 | 1381 | 0289           | 1101   | —    | —    | —    | —    | —                |         | +                   | 34        |
| 35 Methanoarcpoculum lubeaumum Z             | MLAB<sub>0523</sub> | 0525 | 0524 | 0526 | 0073 | 0526           | 0522   | —    | —    | —    | —    | —                | +       | +                   | 35        |
| 36 Methaneallitus marinigri JR1              | MEM<sub>AR0980</sub> | 0982 | 0981 | 0983 | 0539 | 0984           | 0979   | 0878 | 0877 | 0879 | 0876 | —                | +       | +                   | 36        |
| 37 Methanospirillum hungatell JF-1           | MHUN<sub>2562</sub> | 2560 | 2561 | 2559 | 2267 | 2558           | 2563   | —    | —    | —    | —    | —                | +       | +                   | 37        |
| 38 Methanoregula boonei 6A8                  | MBOO<sub>1238</sub> | 1235 | 1236 | 1234 | 0314 | 1233           | 1239   | 0957 | 0956 | 0958 | 0955 | —                | +       | +                   | 38        |
| 39 Methanospheura palustris E1-9c            | MPAL<sub>1728</sub> | 1726 | 1727 | 1725 | 2749 | 1724           | 1729   | 2627 | 2628 | 2626 | 2629 | —                | +       | +                   | 39        |
| 40 Methanosaeta thermophilium PT             | MTHE<sub>0449</sub> | 1126 | 0050 | 1125 | 1124 | 1124           | 1134   | 0047 | 1135 | 1123 | +    | —                | +       | +                   | 40        |
| 41 Methanococoides burtonii DSM 6242         | MBUR<sub>1229</sub> | 1227 | 1228 | 1226 | 1106 | 1105           | 1230   | 1232 | 1231 | 1233 | 1236 | +                | +       | +                   | 41        |
| 42 Methanoscincina acetivorans C2A           | MA<sub>O577</sub> | 0581 | 0578 | 0582 | 3034 | 3035           | 0576   | 0574 | 0575 | 0573 | 3035 | +                | +       | +                   | 42        |
| 43 Methanoscincina burkeni tusaro            | MBAR<sub>A1462</sub> | 1146 | 1146 | 1145 | 1792 | 11791          | 1146   | 1145 | 1145 | 1145 | 11793 | +               | +       | +                   | 43        |
| 44 Methanoscincina mazei Goei                | ML<sub>1741</sub> | 1741 | 1742 | 1744 | 0308 | 0307           | 1740   | 1738 | 1739 | 1737 | 0309 | +               | +       | +                   | 44        |
| 45 Methanopyrus kandleri AV19                | MK<sub>2000</sub> | 0198 | 0198 | 0746 | 1550 | 1548           | 1495   | —    | —    | 0896 | 0890 | —                | +       | +                   | 45        |
| 46 Pyrococcus abyssi GE5                     | —                | —    | —    | —    | —    | —              | —      | —    | —    | —    | —    | —                | —       | —                   | —         |
| 47 Pyrococcus furiosus DSM 3638              | —                | —    | —    | —    | —    | —              | —      | —    | —    | —    | —    | —                | —       | —                   | —         |
| 48 Pyrococcus horikoshii OT3                 | —                | —    | —    | —    | —    | —              | —      | —    | —    | —    | —    | —                | —       | —                   | —         |
| 49 Thermococcus gammatolerans EJ3            | —                | —    | —    | —    | —    | —              | —      | —    | —    | —    | —    | —                | —       | —                   | —         |
| 50 Thermococcus kodakawaensis KOD1           | —                | —    | —    | —    | —    | —              | —      | —    | —    | —    | —    | —                | —       | —                   | —         |
| Organism                                | hemA<sup>a</sup> | hemL | hemB | hemC | hemD | SUMT | PC2-DH | nirD | nirH | nirJ1 | nirJ2 | heme<sup>b</sup> | heme d<sub>1</sub> | B<sub>12</sub><sup>c</sup> | siroheme<sup>d</sup> | F<sub>430</sub> |
|----------------------------------------|-----------------|------|------|------|------|------|--------|------|------|-------|-------|-----------------|----------------|----------------|------------------|-------------|
| 51 Thermococcus onnurineus NA1         | —               | —    | —    | —    | —    | —    | —      | —    | —    | —     | —     | —               | —             | —             | +                | —           |
| 52 Thermococcus sibiricus MM 739       | PTO0918         | 0248 | 1311 | 0249 | 0250 | 1435 | —      | —    | —    | —     | —     | —               | —             | —             | —                | —           |
| 53 Picrophilus torridus DSM 9790       | TA0536          | 0571 | 0955 | 0572 | 0573 | 0652 | 0924   | —    | —    | —     | —     | +               | —             | —             | +                | —           |
| 54 Thermoplasma acidophilum DSM 1728   | TVN03590        | 0635 | 1100 | 0634 | 0633 | 0924 | —      | —    | —    | —     | —     | —               | —             | —             | +                | —           |
| 55 Thermoplasma volcanium GSS1         | RCIX911         | 913  | 912  | 914  | 916  | 915  | 909    | —    | —    | —     | —     | —               | —             | —             | +                | —           |
| 56 uncultured methanogenic archaeon RC-I| NA               | —    | —    | —    | —    | —    | —      | —    | —    | —     | —     | —               | —             | —             | +                | —           |
| 57 Korarchaeum cryptoFilum OPF8         | —               | —    | —    | —    | —    | —    | —      | —    | —    | —     | —     | —               | —             | —             | +                | —           |
| 58 Nanoarchaeum equitans Kin4-M         | —               | —    | —    | —    | —    | —    | —      | —    | —    | —     | —     | —               | —             | —             | —                | —           |
| 59 Nitrosopumilus maritimus SCM1        | NMAR_0510       | 0509 | 0490 | 0491 | 0492 | 0511 | 0512   | 0512 | —    | —     | —     | —               | —             | —             | —                | 5           |
| 60 Escherichia coli K-12 MG1655         | B1210           | 0154 | 0369 | 3805 | 3804 | 3368 | —      | —    | —    | —     | —     | +               | —             | —             | —                | —           |
| 61 Pseudomonas aeruginosa PAO1          | PA4666          | 3977 | 5243 | 5260 | 5259 | 0510<sup>f</sup> | 2611 | 0515<sup>f</sup> | 0512<sup>f</sup> | — | 0511<sup>f</sup> | + | + | + | + | — | 61 |
| 62 Desulfovibrio vulgaris Hildenborough  | DVU1461         | 3168 | 0856 | 1890 | 0734 | 0734 | 1463   | 0854 | 3167 | 0855 | 0857 | + | + | + | — | 62 |

<sup>a</sup>hemA<sup>a</sup> gene encoding glutamyl-tRNA reductase in the listed organisms. In the case of two or more copies for one gene, these copies are only listed if none of them is located in the gene cluster, otherwise only the copy within the cluster is listed. This is valid for all listed genes. Genes colocated within the same gene cluster in a certain species are highlighted in the same font (bold, italics, bold italics).

<sup>b</sup>Organisms possessing heme-containing proteins according to the literature [14–22] are marked with +. For organisms with empty fields the presence of heme-containing proteins was not clearly obvious from literature and bioinformatics data.

<sup>c</sup>Organisms possessing cobalamin biosynthesis genes according to [35] are marked with +. Organisms that do not synthesize cobalamin are marked with — [35]. For organisms with empty fields the ability to synthesize cobalamin was not obvious from the literature.

<sup>d</sup>Organisms possessing siroheme-containing sulfite or nitrite reductases based on sequence homology are marked with +. For organisms with empty fields the presence of siroheme-containing sulfite and nitrite reductases was not clearly obvious from literature and bioinformatics data.

<sup>e</sup>Numbers represent the database gene number. The full database gene numbers including the strain abbreviation is given for the hemA<sup>a</sup> genes. For all other genes the database gene numbers are given without the strain abbreviation. For genes that are located as direct or near neighbors in an individual strain the numbers are written in the same font.

<sup>f</sup>P. aeruginosa nirE, nirD, nirH, and nirJ genes involved in heme d<sub>1</sub> biosynthesis.
with hemA<sup>6</sup>, hemL, hemB, hemC, hemD, SUMT, and PC2-DH genes (Table 1 and Figure 2). It was previously reported that D. vulgaris and some methanogenic archaea harbor these nir genes on their genomes. It was speculated that the nir genes might be involved in the alternative heme biosynthesis pathway in these organisms [26]. Here, we show not only that the methanogenic archaea contain nir genes, but also that the majority of archaea that synthesize heme also contain these genes (Table 1). These nir genes encode proteins that are homologous to proteins involved in heme d<sub>1</sub> biosynthesis in denitrifying bacteria such as P. aeruginosa. The dioxygenobacteriochlorin heme d<sub>1</sub> serves as an essential prosthetic group in the cytochrome cd<sub>1</sub> nitrite reductase which catalyzes the second step of denitrification [36]. However, based on amino acid sequence homology searches, only Pyrobaculum aerophilum, Pyrobaculum arsenaticum, and Pyrobaculum calidifontis possess a potential cytochrome cd<sub>1</sub> nitrite reductase. All other archaeal genomes analyzed in this study do not. Consequently, the majority of archaeal nir genes are not involved in heme d<sub>1</sub> biosynthesis. Rather, they are likely to be involved in heme biosynthesis. Therefore, we renamed these nir-like genes in the Archaea ahb (archaeal heme biosynthesis)-nir genes.

### 3.4. Structures of Potential Heme Biosynthesis Gene Clusters in Archaea

As mentioned above, the ahb-nir genes are often clustered with the hem, SUMT, and PC2-DH genes on the archaeal genomes. The most complete gene clusters, comprising ten out of the eleven potential heme biosynthesis genes, were found in the genomes of P. aerophilum and P. arsenaticum (Figure 2). In P. aerophilum these genes form one large, uninterrupted gene cluster. Another striking clustering of the potential heme biosynthesis genes was observed in the Methanosarcinales (Figure 2). For example, in the genomes of M. acetivorans and M. barkeri the genes hemA<sup>6</sup>, hemL, hemB, hemC, ahb-nirD, ahb-nirH, ahb-nirJ1, and PC2-DH are organized as one continuous gene cluster, while the genes hemD, ahb-nirJ2, and SUMT are localized together in a second gene cluster. In Halobacterium sp. NRC-1 three heme biosynthesis-related gene clusters were found. The first cluster comprises the genes hemL, hemB, hemC, hemD, and SUMT, the second consists of hemA<sup>6</sup>, PC2-DH, ahb-nirD, and ahb-nirH and the third contains ahb-nirJ1 and ahb-nirJ2 (Figure 2). In the other archaeal species that possess ahb-nir genes the clustering with the hem, SUMT, and PC2-DH genes is less distinct, but there is still often a colocalization of one or two ahb-nir genes with one or several hem genes (Table 1).

#### 3.5. Proposed Function of the ahb-nir Genes during Heme Biosynthesis in the Archaea

As already mentioned above the ahb-nir genes encode proteins that are similar to proteins involved in heme d<sub>1</sub> biosynthesis. We compared the amino acid sequences of the Ahb-Nir proteins from M. barkeri with the Nir proteins involved in heme d<sub>1</sub> biosynthesis in...
P. aeruginosa. We found the following sequence identities: M. barkeri Ahb-NirD and P. aeruginosa NirD: 36.3%; Ahb-NirH and NirH: 40.1%; Ahb-NirJ1 and NirJ: 29.5%; Ahb-NirJ2 and NirJ: 38.8%. Thus, the Ahb-NirJ2 is more similar to the heme \(d_1\) biosynthesis protein NirJ than is the Ahb-NirJ1. Additionally, we found an amino acid sequence identity of 31.8% between the two Ahb-NirJ proteins.

Although the precise functions of the Nir proteins involved in heme \(d_1\) biosynthesis has not yet been established, several reasonable proposals were made [26, 36–40]. First of all, it is known that heme \(d_1\) is biosynthesized from precorrin-2 [27, 35]. In order to obtain heme \(d_1\) from this precursor the following modifications have to take place: (a) decarboxylation of the acetate groups on rings C and D, (b) removal of the propionate side chains on rings A and B and replacement by oxo groups on rings A and B, (c) formation of an acrylate side chain on ring D, (d) oxidation of the tetrapyrrole macrocycle, and (e) iron insertion. The order of these reactions is not known. However, it was proposed that the oxidation reaction (b) might be catalyzed by the NirJ protein during heme \(d_1\) formation [37]. NirJ belongs to the so-called Radical SAM enzyme family whose members are known to catalyze chemically challenging reactions through radical-based mechanisms [41, 42]. It was also speculated that the NirD, NirL, NirG, and NirH proteins might be responsible for the decarboxylation reaction (a) [38].

In order to form heme from precorrin-2 the decarboxylation of the acetate groups on rings C and D as in reaction (a) is required and thus might be catalyzed by Ahb-NirD and Ahb-NirH. Further, the acetate side chains on rings A and B have to be removed, probably in a reaction which resembles the mechanism of reaction (b). Consequently, the Radical SAM enzyme Ahb-NirJ2 which shares 38.8% sequence identity with P. aeruginosa NirJ is a good candidate for catalyzing this reaction. This function was previously proposed for one of the NirJ-like proteins from D. vulgaris [26]. Another reaction which is required for heme formation from precorrin-2 that has, however, no equivalent in heme \(d_1\) biosynthesis is the oxidative decarboxylation of the propionate side chains on rings A and B to the corresponding vinyl groups. This reaction also takes place during the classical heme biosynthesis route in most bacteria and the Eukaryota. In bacteria it is catalyzed by either HemF or HemN (see Figure 1(a)). HemN also belongs to the Radical SAM enzyme family [43]. Thus, Ahb-NirJ1 (Radical SAM family member) might catalyze the formation of the required vinyl groups. In summary, we propose that the Ahb-Nir proteins catalyze some of the late reaction steps during archaeal heme biosynthesis from precorrin-2 (Figure 3).

3.6. Distribution of the ahb-nir Genes over the Archaeal Genomes. In accordance with the proposed function of the ahb-nir gene products during the late steps of archaeal heme biosynthesis we failed to detect any of the ahb-nir genes in those archaeal genomes without any hem genes (Table 1). However, the presence of the hem genes in an archaeological genome does not necessarily mean that the ahb-nir genes are also present. As mentioned above, out of

---

Figure 3: Proposal for the novel, alternative heme biosynthesis pathway in archaea. Archaeal heme biosynthesis starts with the SAM-dependent methylation of UROGEN to precorrin-2 by a SUMT and potentially proceeds via oxidation of precorrin-2 to sirohydrochlorin by PC2-DH. The side chain modifications (highlighted by dashed circles) including acetate group decarboxylation on rings C and D, acetate group removal on rings A and B, and vinyl group formation on rings A and B are potentially catalyzed by the Ahb-Nir proteins.
47 archaeal genomes containing all five early hem genes only 32 also contain the ahb-nir genes. The 15 archaeal species which possess the hem, but no ahb-nir genes probably synthesize their UROGEN solely as precursor for siroheme [44, 45], cobalamin [31] and, in the case of methanogens, for coenzyme F₄₃₀ [46]. However, most likely they do not form heme. Accordingly, almost all of these 15 species also possess both a SUMT and a PC2-DH gene which are required for siroheme, cobalamin, and coenzyme F₄₃₀ biosynthesis. One exception is P. torridus which does not contain a recognizable PC2-DH gene. Moreover, T. acidophilum and T. volcanium do not possess a SUMT gene. Interestingly, these three species are the only representatives of the Archaea for which hemE and hemH genes were found (see above). For P. torridus and T. acidophilum heme-containing proteins were biochemically characterized [18, 20]. However, considering the observation that their genomes lack recognizable ahb-nir genes and some of the late hem genes their route of heme biosynthesis remains currently unclear.

Within the group of the 32 archaeal species that contain ahb-nir genes several subgroups can be recognized. First of all, there are those species for which a complete set of ahb-nir genes (ahb-nirD, ahb-nirH, ahb-nirJ1, ahb-nirJ2) was found. Out of the 32 genomes containing ahb-nir genes 27 contain all four of them. Among the 5 genomes in which not all ahb-nir genes are present, those of Haloarcula vutahensis and Methanopyrus kandleri are missing ahb-nirD and ahb-nirH. The genomes of Sulfolobus acidocaldarius, Haloquadratum walsbyi, and Nicrosopumilus maritimus do not contain ahb-nirJ1 and ahb-nirJ2. Thus, for these five species it is questionable whether they synthesize heme themselves, if at all required. For example, for M. kandleri and N. maritimus no indications were found in the literature or databases that they possess heme-containing proteins. Further, two subgroups of ahb-nir containing archaea can be distinguished depending on whether they possess two distinct ahb-nirD and ahb-nirH genes or whether they contain an ahb-nirD gene fusion. In fact, almost all archaea possess the fused genes except for those methanogenic archaea which contain the ahb-nir genes (Table 1). However, in these methanogens the ahb-nirD and ahb-nirH genes are always located aside to each other on the genome with the only exception being Methanoseta thermophila. Likewise, the two ahb-nir genes are also often (15 out of 29) co-localized on the genomes, either as direct neighbors or in close proximity to each other, indicating gene duplication as the origin of the two copies.

Our bioinformatics investigation of 59 archaeal genomes in combination with the experimental evidence that two methyl groups of archaeal heme are derived from S-adenosyl-L-methionine strongly suggests that heme biosynthesis in the Archaea follows a novel, yet mostly unknown route. It starts with the methylation of UROGEN to precorrin-2 catalyzed by SUMT, followed by the oxidation of precorrin-2 to sirohydrochlorin by PC2-DH and further transformations (decarboxylation of acetate groups, removal of acetate groups, oxidative decarboxylation of propionate to vinyl groups, and insertion of iron) of the macrocycle side chains which are most likely performed by the Ahb-Nir proteins (Figure 3). Clearly, these proposals need to be tested experimentally. Thus, we decided to first verify the predicted functions of the M. barkeri proteins Mba_1791 and Mba_1461 as SUMT and PC2-DH, respectively.

3.7. Production and Purification of Recombinant Mba_1791 and Mba_1461. The M. barkeri proteins Mba_1791 and Mba_1461 were recombinantly produced as N-terminal His-tagged fusion proteins in E. coli. In both cases the recombinant proteins were produced in a soluble form and in a high yield. We purified Mba_1791 and Mba_1461 to apparent homogeneity using a single affinity-chromatographic step on Ni Sepharose 6 Fast Flow (Figure 4(a)). The purified Mba_1791 exhibited a slight red-brown color. UV-visible absorption spectroscopy suggested the presence of a copurified tetrapyrrole, probably the reaction product of Mba_1791 (data not shown). For other SUMTs (e.g., P. aeruginosa NirE) the co-purification of their reaction product has been previously reported [27, 35, 47]. Therefore, the presence of a tetrapyrrole in the purified Mba_1791 was a first hint towards the function of this protein as a SUMT. In contrast, purified Mba_1461 appeared colorless.

The oligomeric state of Mba_1791 and Mba_1461 was determined by gel permeation chromatography. This experiment revealed a native relative molecular mass of 55,300 ± 840 Da for Mba_1791 and 60,800 ± 7,300 Da for Mba_1461, respectively (Figure 4(a)). The calculated molecular masses based on the amino acid sequences of the proteins are 26,350 Da for Mba_1461 and 27,230 Da for Mba_1791. Thus, gel permeation chromatography suggests a dimeric structure for both proteins. Other SUMTs [48, 49] and PC2-DH [33, 50] are also thought to be dimeric proteins.

3.8. Mba_1791 Acts as a SUMT In Vivo. During production of Mba_1791 in E. coli a red compound accumulated and remained in the soluble protein fraction of the cell-free extract after disruption of the cells and ultracentrifugation. This compound was extracted using C₁₈-reversed phase silica gel and analyzed by UV-visible absorption spectroscopy. The UV-visible absorption spectrum of the extracted compound exhibited an absorption maximum at 378 nm which strongly resembled the previously reported spectra of sirohydrochlorin (Figure 4(b)) [50]. Obviously, recombinantly produced M. barkeri Mba_1791 showed significant SUMT activity in the production host E. coli which led to the accumulation of sirohydrochlorin, the oxidized form of the SUMT reaction product precorrin-2. Such tetropyrrole accumulation during recombinant SUMT production has been described before. Depending on the source of the enzyme the formation of either trimethylpyrrocorphin, which is a nonphysiological trimethylated reaction product, or sirohydrochlorin was reported [27, 35, 47, 51–53]. Apparently, Mba_1791 belongs to the class of SUMT enzymes that accumulates sirohydrochlorin and does not catalyze the overmethylation of precorrin-2 to trimethylpyrrocorphin.

3.9. M. barkeri Mba_1791 Is a SUMT. In order to investigate the in vitro activity of Mba_1791 a coupled enzyme assay was performed. The substrate uroporphyrinogen III
Figure 4: Purification and characterization of Mba\textsubscript{1791} and Mba\textsubscript{1461}. (a) SDS-PAGE analysis of purified Mba\textsubscript{1791} (lane 1) and Mba\textsubscript{1461} (lane 2). Gel permeation chromatography revealed native relative molecular masses of 55,300 ± 840 Da for Mba\textsubscript{1791} (dotted line) and 60,800 ± 7,300 Da for Mba\textsubscript{1461} (solid line), respectively. (b) UV-visible absorption spectrum of extracted tetrapyrroles which accumulated during production of recombinant Mba\textsubscript{1791} in E. coli. (c) UV-visible absorption spectra of enzyme assays after overnight incubation at 37°C in the anaerobic chamber. Uroporphyrinogen III was produced from ALA by the enzymes HemB, HemC, and HemD (dashed double dotted line). Addition of purified Mba\textsubscript{1791} and SAM to the assay mixture resulted in precorrin-2 formation (solid line). Addition of purified NirE, Mba\textsubscript{1461}, and NAD\textsuperscript{+} to the assay resulted in formation of sirohydrochlorin (dotted line). In a coupled enzyme assay containing purified Mba\textsubscript{1791} and Mba\textsubscript{1461} the formation of sirohydrochlorin was also observed (dashed line). For exact details see Section 2.

was produced enzymatically and incubated overnight with recombinant purified Mba\textsubscript{1791}. The formation of reaction products was followed using UV-visible absorption spectroscopy (Figure 4(c)). The absorption spectrum of a control assay mixture containing only the uroporphyrinogen III producing enzymes showed no characteristic absorption features under anaerobic conditions. In contrast, the addition of purified Mba\textsubscript{1791} and SAM to the reaction mixture resulted in a yellow colored solution after overnight incubation. The corresponding absorption spectrum exhibited a broad absorption between 350–400 nm and 400–500 nm which is characteristic for precorrin-2 [50, 54]. Consequently, Mba\textsubscript{1791} is indeed a SUMT. Therefore, we will name the enzyme from now on Mba\textsubscript{SUMT}. The specific activity of Mba\textsubscript{SUMT} was determined with uroporphyrinogen III (produced by chemical reduction of uroporphyrin III) at
a concentration of 17 μM, a SAM concentration of 200 μM and a Mba\textsubscript{SUMT} concentration of 1.5 μM. Under these conditions we observed a specific activity of 616 nmol precorrin-2 × h\(^{-1}\)× mg\(^{-1}\) Mba\textsubscript{SUMT} corresponding to a turnover of 17 h\(^{-1}\). This specific activity is in the same range as the activities observed for other SUMTs [27, 55–57].

3.10. M. barkeri Mba\textsubscript{1461} is a PC2-Dehydrogenase. In order to verify the postulated PC2-DH activity for Mba\textsubscript{1461} in an in vitro assay the enzymatically produced uroporphyrinogen III was converted to the PC2-DH substrate precorrin-2 by addition of the SUMT NirE from P. aeruginosa. Additionally, purified M. barkeri Mba\textsubscript{1461} and NAD\(^+\) were added to the reaction mixture. After overnight incubation a UV-visible absorption spectrum of the now purple reaction mixture was measured. The spectrum (Figure 4(c)) corresponds to a typical absorption spectrum of spheroidenone with an absorption maximum at 378 nm [50]. Thus, Mba\textsubscript{1461} exhibited strong PC2-DH activity in vitro and can be safely assigned as Mba\textsubscript{PC2-DH}.

We also tested the activities of the M. barkeri enzymes Mba\textsubscript{SUMT} and Mba\textsubscript{PC2-DH} in a coupled assay. Both were added to the reaction mixture containing all enzymes for uroporphyrinogen III generation. After overnight incubation the UV-visible absorption spectrum of this assay mixture was similar to the spectrum of the reaction mixture containing P. aeruginosa NirE (SUMT) and Mba\textsubscript{PC2-DH} showing again the formation of spheroidenone (Figure 4(c)). When SAM and/or NAD\(^+\) were omitted from this activity assay no formation of spheroidenone was observed (data not shown).

4. Conclusion

In this study we have identified gene clusters in many archaeal genomes that are likely required for the biosynthesis of heme via a novel pathway. These gene clusters consist of (i) the known hem genes (hem\textsubscript{A}, hemL, hemB, hemC, hemD) necessary for the formation of the heme precursor UROGEN, (ii) the SUMT and PC2-DH genes required for the synthesis of the intermediates precorrin-2 and spheroidenone, and (iii) the ahb-nir genes whose protein products are probably responsible for the conversion of spheroidenone into heme (Figure 3). We propose that the detected ahb-nir genes are involved in archaean heme biosynthesis and not in heme \(d_1\) biosynthesis since almost all archaebacteria do not possess a cytochrome \(cd_1\) nitrite reductase. However, we failed to detect any obvious genes encoding potential ferrochelatases within the archaean heme biosynthesis gene clusters. The Archaea do not usually possess a hemH gene encoding the bacterial-type ferrochelatase but do contain several copies of genes encoding putative cobalt- and/or magnesium chelatases, although they do not synthesize chlorophylls. These gene products might be involved in the archaean heme biosynthesis. To confirm our bioinformatics findings and predictions further experimental verification will be required in order to determine the exact function of the Ahb-Nir proteins in the archaean heme biosynthesis process.

Acknowledgments

The authors would like to thank Professor R. Thauer (MPI Marburg) for the gift of genomic DNA from M. barkeri. They also thank Dr. J. Moser (TU Braunschweig) for helpful discussions. This work was financially supported by grants from the Deutsche Forschungsgemeinschaft, the Fonds der Chemischen Industrie to G. Layer and the BBSRC to M. J. Warren.

References

[1] G. Layer, I. Reichelt, D. Jahn, and D. W. Heinz, “Structure and function of enzymes in heme biosynthesis,” Protein Science, vol. 19, no. 6, pp. 1137–1161, 2010.
[2] A. R. Battersby, “Tetrapyrroles: the pigments of life,” Natural Product Reports, vol. 17, no. 6, pp. 507–526, 2000.
[3] G. A. Hunter and G. C. Ferreira, "5-Aminolevulinate synthase: catalysis of the first step of heme biosynthesis," Cellular and Molecular Biology, vol. 55, no. 1, pp. 102–110, 2009.
[4] C. Lüer, S. Schauer, K. Möbius et al., “Complex formation between glutamyl-tRNA reductase and glutamate-1-semialdehyde 2,1-aminomutase in Escherichia coli during the initial reactions of porphyrin biosynthesis,” Journal of Biological Chemistry, vol. 280, no. 19, pp. 18568–18572, 2005.
[5] E. K. Jaffe, “The porphobilinogen synthase catalyzed reaction mechanism,” Bioorganic Chemistry, vol. 32, no. 5, pp. 316–325, 2004.
[6] P. M. Shoollingin-Jordan, “Porphobilinogen deaminase and uroporphyrinogen III synthase: structure, molecular biology, and mechanism,” Journal of Bioenergetics and Biomembranes, vol. 27, no. 2, pp. 181–195, 1995.
[7] H. A. Dailey, “Terminal steps of haem biosynthesis,” Biochemical Society Transactions, vol. 30, no. 4, pp. 590–595, 2002.
[8] H. A. Dailey, T. A. Dailey, C. K. Wu, A. E. Medlock, J. P. Rose, and K. F. Wang, “Ferrochelatase at the millennium: structures, mechanisms and [2Fe-2S] clusters,” Cellular and Molecular Life Sciences, vol. 57, no. 13–14, pp. 1909–1926, 2000.
[9] N. Frankenberg, J. Moser, and D. Jahn, “Bacterial heme biosynthesis and its biotechnological application,” Applied Microbiology and Biotechnology, vol. 63, no. 2, pp. 115–127, 2003.
[10] P. Panek and M. R. O’Brien, “A whole genome view of prokaryotic haem biosynthesis,” Microbiology, vol. 148, no. 8, pp. 2273–2282, 2002.
[11] G. Cavallaro, L. Decaria, and A. Rosato, “Genome-based analysis of heme biosynthesis and uptake in prokaryotic systems,” Journal of Proteome Research, vol. 7, no. 11, pp. 4946–4954, 2008.
[12] S. Shima, M. Sordel-Klippert, A. Brioukhanov, A. Netrusov, D. Linder, and R. K. Thauer, “Characterization of a heme-dependent catalase from Methanobrevibacter arboriphilus,” Applied and Environmental Microbiology, vol. 67, no. 7, pp. 3041–3045, 2001.
[13] P. Duwat, S. Sourice, B. Cesselin et al., “Respiration capacity of the fermenting bacterium Lactococcus lactis and its positive effects on growth and survival,” Journal of Bacteriology, vol. 183, no. 15, pp. 4509–4516, 2001.
[14] J. W. A. Allen, E. M. Harvat, J. M. Stevens, and S. J. Ferguson, "A variant System I for cytochrome c biogenesis in archaea and some bacteria has a novel CcmE and no CcmH," FEBS Letters, vol. 580, no. 20, pp. 4827–4834, 2006.

[15] I. Bertini, G. Cavallaro, and A. Rosato, "Cytochrome c: occurrence and functions," Chemical Reviews, vol. 106, no. 1, pp. 90–115, 2006.

[16] I. Bertini, G. Cavallaro, and A. Rosato, "Evolution of mitochondrial-type cytochrome c domains and of the protein machinery for their assembly," Journal of Inorganic Biochemistry, vol. 101, no. 11-12, pp. 1798–1811, 2007.

[17] T. A. K. Freitas, S. Hou, E. M. Dioum et al., "Ancestral hemoglobins in Archaea," Proceedings of the National Academy of Sciences of the United States of America, vol. 101, no. 17, pp. 6675–6680, 2004.

[18] M. M. Pereira, T. M. Bandeiras, A. S. Fernandes, R. S. Lemos, A. M. P. Melo, and M. Teixeira, "Respiratory chains from aerobic thermophilic prokaryotes," Journal of Bioenergetics and Biomembranes, vol. 36, no. 1, pp. 93–105, 2004.

[19] G. Schäfer, M. Engelhard, and V. Müller, "Bioenergetics of the Archaea," Microbiology and Molecular Biology Reviews, vol. 63, no. 3, pp. 570–620, 1999.

[20] R. K. Thauer, A. K. Kaster, H. Seedorf, W. Buckel, and R. Hedderich, "Methanogenic archaea: ecologically relevant differences in energy conservation," Nature Reviews Microbiology, vol. 6, no. 8, pp. 579–591, 2008.

[21] T. Ishida, L. Yu, H. Akutsu et al., "A primitive pathway of porphyrin biosynthesis and enzymology in Desulfovibrio vulgaris," Proceedings of the National Academy of Sciences of the United States of America, vol. 95, no. 9, pp. 4853–4858, 1998.

[22] S. A. L. Lobo, A. Brindley, M. J. Warren, and L. M. Saraiva, "Functional characterization of the early steps of tetrapyrrole biosynthesis and modification in Desulfovibrio vulgaris Hildenborough," Biochemical Journal, vol. 420, no. 2, pp. 317–325, 2009.

[23] S. Storbeck, J. Walther, J. Müller et al., "The Pseudomonas aeruginosa nirE gene encodes the S-adenosyl-L-methionine-dependent uroporphyrinogen III methyltransferase required for heme d biosynthesis," FEBS Journal, vol. 276, no. 20, pp. 5973–5982, 2009.

[24] I. Uchiyama, "MBGD: microbial genome database for comparative analysis," Nucleic Acids Research, vol. 31, no. 1, pp. 58–62, 2003.

[25] I. Uchiyama, "MBGD: a platform for microbial comparative genomics based on the automated construction of orthologous groups," Nucleic Acids Research, vol. 35, no. 1, pp. D343–D346, 2007.

[26] I. Uchiyama, T. Higuchi, and M. Kawai, "MBGD update 2010: toward a comprehensive resource for exploring microbial genome diversity," Nucleic Acids Research, vol. 38, no. 1, pp. D361–D365, 2010.

[27] Y. Zhang, D. A. Rodionov, M. S. Gelfand, and V. N. Gladyshev, "Comparative genomic analyses of nickel, cobalt and vitamin B12 utilization," BMC Genomics, vol. 10, article 78, 2009.

[28] A. Pfaltz, A. Kobelt, R. Hunster, and R. K. Thauer, "Bio-synthesi- sism of coenzyme F430 in methanogenic bacteria. Identification of 15,17-seco-F430-17-acid as an intermediate," European Journal of Biochemistry, vol. 170, no. 1-2, pp. 459–467, 1988.

[29] H. L. Schubert, R. S. Rose, H. K. Leech et al., "Structure and function of SirC from Bacillus megaterium: a metal-binding precorrin-2 dehydrogenase," Biochemical Journal, vol. 415, no. 2, pp. 257–263, 2008.

[30] M. J. Warren, E. Raux, H. L. Schubert, and J. C. Escalante-Semerena, "The biosynthesis of adenosylcobalamin (vitamin B12)," Natural Product Reports, vol. 19, no. 4, pp. 390–412, 2002.

[31] R. S. Zajicek, S. Bali, S. Arnold, A. A. Brindley, M. J. Warren, and S. J. Ferguson, "D haem biogenesis—assessing the roles of three nir gene products," FEBS Journal, vol. 276, no. 21, pp. 6399–6411, 2009.

[32] W. G. Zumft, "Cell biology and molecular basis of denitrifica- tion?" Microbiology and Molecular Biology Reviews, vol. 61, no. 4, pp. 533–616, 1997.

[33] A. A. Brindley, R. Zajicek, M. J. Warren, S. J. Ferguson, and S. E. J. Rigby, "NirJ, a radical SAM family member of the d1 heme biogenesis cluster," FEBS Letters, vol. 584, no. 11, pp. 2461–2466, 2010.

[34] S. Kawasaki, H. Arai, T. Kodama, and Y. Igarashi, "Gene cluster for dissimilatory nitrite reductase (nir) from Pseudomonas aeruginosa: sequencing and identification of a locus for heme d1 biosynthesis," Journal of Bacteriology, vol. 179, no. 1, pp. 235–242, 1997.

[35] G. Palmredo, P. Seither, H. Korner et al., "Resolution of the nirD locus for heme d1 synthesis of cytochrome cd (respiratory nitrite reductase) from Pseudomonas stutzeri," European Journal of Biochemistry, vol. 232, no. 3, pp. 737–746, 1995.

[36] M. Suzuki, T. Hirai, H. Arai, M. Ishii, and Y. Igarashi, "Purification, characterization, and gene cloning of thermophilic cytochrome cd nitrite reductase from Hydrogenobacter thermophilus TK-6," Journal of Bioscience and Bioengineering, vol. 101, no. 5, pp. 391–397, 2006.

[37] S. J. Booker, "Anaerobic functionalization of unactivated C-H bonds," Current Opinion in Chemical Biology, vol. 13, no. 1, pp. 58–73, 2009.

[38] P. A. Frey, A. D. Hegeman, and F. J. Ruzicka, "The radical SAM superfamily," Critical Reviews in Biochemistry and Molecular Biology, vol. 43, no. 1, pp. 63–88, 2008.

[39] G. Layer, J. Moser, D. W. Heinz, D. Jahn, and W. D. Schubert, "Crystal structure of coproporphyrinogen III oxidase reveals cofactor geometry of Radical SAM enzymes," EMBO Journal, vol. 22, no. 23, pp. 6214–6224, 2003.

[40] E. E. Johnson and B. Mukhopadhyay, "Coenzyme F-dependent sulfite reductase-enabled sulfite detoxification and use of sulfite as a sole sulfur source by Methanococcus maripaludis," Applied and Environmental Microbiology, vol. 74, no. 11, pp. 3591–3595, 2008.

[41] A. Schiffer, K. Parey, E. Warkentin et al., "Structure of the Dissimilatory Sulfite Reductase from the Hyperthermophilic Archaea
Archaea

Archaeon *Archaeoglobus fulgidus*, *Journal of Molecular Biology*, vol. 379, no. 5, pp. 1063–1074, 2008.

[46] R. K. Thauer and L. G. Bonacker, “Biosynthesis of coenzyme F430, a nickel porphinoid involved in methanogenesis,” *Ciba Foundation symposium*, vol. 180, pp. 210–222, 1994.

[47] J. Fan, D. Wang, Z. Liang, M. Guo, M. Teng, and L. Niu, “Maize uroporphyrinogen III methyltransferase: overexpression of the functional gene fragments in *Escherichia coli* and one-step purification,” *Protein Expression and Purification*, vol. 46, no. 1, pp. 40–46, 2006.

[48] M. E. Stroupe, H. K. Leech, D. S. Daniels, M. J. Warren, and E. D. Getzoff, “CysG structure reveals tetrapyrole-binding features and novel regulation of siroheme biosynthesis,” *Nature Structural Biology*, vol. 10, no. 12, pp. 1064–1073, 2003.

[49] J. Večnová, R. M. Graham, E. Raux et al., “Structure/function studies on a S-adenosyl-L-methionine-dependent uroporphyrinogen III C methyltransferase (SUMT), a key regulatory enzyme of tetrapyrole biosynthesis,” *Journal of Molecular Biology*, vol. 344, no. 2, pp. 419–433, 2004.

[50] H. L. Schubert, E. Raux, A. A. Brindley et al., “The structure of *Saccharomyces cerevisiae* Met8p, a bifunctional dehydrogenase and ferrochelatase,” *EMBO Journal*, vol. 21, no. 9, pp. 2068–2075, 2002.

[51] . Chou Pei-Lin, M. Ohtsuka, T. Minowa et al., “Reddish *Escherichia coli* cells caused by overproduction of *Bacillus stearothermophilus* uroporphyrinogen III methylase: cloning, sequencing, and expression of the gene,” *Bioscience, Biotechnology and Biochemistry*, vol. 59, no. 10, pp. 1817–1824, 1995.

[52] T. Leustek, M. Smith, M. Murillo et al., “Siroheme biosynthesis in higher plants. Analysis of an S-adenosyl-L-methionine-dependent uroporphyrinogen III methyltransferase from *Arabidopsis thaliana*,” *Journal of Biological Chemistry*, vol. 272, no. 5, pp. 2744–2752, 1997.

[53] I. Sattler, C. A. Roessner, N. J. Stolowich et al., “Cloning, sequencing, and expression of the uroporphyrinogen III methyltransferase cobA gene of *Propionibacterium freudenreichii* (shermanii),” *Journal of Bacteriology*, vol. 177, no. 6, pp. 1564–1569, 1995.

[54] M. J. Warren, N. J. Stolowich, P. J. Santander, C. A. Roessner, B. A. Sowa, and A. I. Scott, “Enzymatic synthesis of dihydrosirosi-hydrochlorin (precorrin-2) and of a novel pyrrocorphin by uroporphyrinogen III methylase,” *FEBS Letters*, vol. 261, no. 1, pp. 77–80, 1990.

[55] F. Blanche, L. Debuissche, D. Thibaut, J. Crouzet, and B. Cameron, “Purification and characterization of S-adenosyl-L-methionine: uroporphyrinogen III methyltransferase from *Pseudomonas denitrificans*,” *Journal of Bacteriology*, vol. 171, no. 8, pp. 4222–4231, 1989.

[56] F. Blanche, C. Robin, M. Couder et al., “Purification, characterization, and molecular cloning of S-adenosyl-L-methionine: uroporphyrinogen III methyltransferase from *Methanobacterium ivanovii*,” *Journal of Bacteriology*, vol. 173, no. 15, pp. 4637–4645, 1991.

[57] C. Robin, F. Blanche, L. Cauchois, B. Cameron, M. Couder, and J. Crouzet, “Primary structure, expression in *Escherichia coli*, and properties of S-adenosyl-L-methionine: uroporphyrinogen III methyltransferase from *Bacillus megaterium*,” *Journal of Bacteriology*, vol. 173, no. 15, pp. 4893–4896, 1991.