Open issues for protein function assignment in *Haloferax volcanii* and other halophilic archaea.

Friedhelm Pfeiffer[1#] and Mike Dyall-Smith[1,2]

[1]Computational Biology Group, Max-Planck-Institute of Biochemistry, Martinsried, Germany
[2]Veterinary Biosciences, Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, Australia

Corresponding authors addresses:

Dr. Friedhelm Pfeiffer

Computational Biology Group, Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

e-mail: fpf@biochem.mpg.de

Abstract

Background: Annotation ambiguities and annotation errors are a general challenge in genomics. While a reliable protein function assignment can be obtained by experimental characterization, this is expensive and time-consuming, and the number of such Gold Standard Proteins (GSP) with experimental support remains very low compared to proteins annotated by sequence homology, usually through automated pipelines. Even a GSP may give a misleading assignment when used as a reference: the homolog may be close enough to support isofunctionality, but the substrate of the GSP is absent from the species being annotated. In such cases the enzymes cannot be isofunctional. Here, we examine a variety of such issues in halophilic archaea (class Halobacteria), with a strong focus on the model haloarchaeon *Haloferax volcanii*.

Results: Annotated proteins of *Hfx. volcanii* were identified for which public databases tend to assign a function that is probably incorrect. In some cases, an alternative, probably correct, function can be predicted or inferred from the available evidence but this has not been adopted by public databases because experimental validation is lacking. In other cases, a probably invalid specific function is predicted by homology, and while there is evidence that this assigned function is unlikely, the true function remains elusive. We list 50 of those cases, each with detailed background information so that a conclusion about the most likely biological function can be drawn. For reasons of brevity and comprehension, only key aspects are listed in the main text, with detailed information being provided in a corresponding section of the Supplementary Material.

Conclusions: Compiling, describing and summarizing these open annotation issues and functional predictions will benefit the scientific community in the general effort to improve the evaluation of protein function assignments and more thoroughly detail them. By highlighting the gaps and likely annotation errors currently in the databases, we hope this study will provide a framework for experimentalists to sytematically confirm (or disprove) our function predictions or to uncover yet unexpected functions.

Keywords: haloarchaea, genome annotation, Gold Standard Protein, *Haloferax volcanii*, annotation error

1 INTRODUCTION

Haloferax volcanii is a model organism for halophilic archaea (Hartman et al., 2010, Schulze et al., 2020, Leigh et al., 2011, Perez-Arnaiz et al., 2020, Soppa, 2011, Haque et al., 2020), for which an elaborate set of genetic tools has been developed (Allers et al., 2010, Allers and Mevarech, 2005, Kiljunen et al., 2014). Its genome has been sequenced and carefully annotated (Hartman et al., 2010, Pfeiffer et al., 2008a, Pfeiffer and Oesterhelt, 2015). A plethora of biological aspects have been successfully tackled in this species, with examples including DNA replication (Perez-Arnaiz et al., 2020); cell division and cell shape (Turkowyd et al., 2020, Walsh et al., 2019, de Silva et al., 2021, Duggin et al., 2015, Liao et al., 2021); metabolism (Brasen and Schonheit, 2001, Johnsen et al., 2009, Pickl et al., 2012, Sutter et al., 2016, Reinhardt et al., 2019, Kuprat et al., 2021, Kuprat et al., 2020, Sutter et al., 2020, Tästensen et al., 2020); protein secretion (Abdul-Halim et al., 2020, Abdul Halim et al., 2018, Abdul Halim et al., 2013, Storf et al., 2010); motility and biofilms (Schiller et al., 2020, Pohlschroder and Esquivel, 2015, Li et al., 2020, Collins et al., 2020, Quax et al., 2018, Nussbaum et al., 2020); mating (Shalev et al., 2017); signalling (Braun et al., 2019); virus defence (Maier et al., 2019); proteolysis (Reuter and Maupin-Furlow, 2004, Reuter et al., 2010, Prunetti et al., 2014, Cerletti et al., 2018, Cerletti et al., 2014, Costa et al., 2018); posttranslational modification (N-glycosylation; SAMPylation)(Cao et al., 2015, Kaminski and Eichler, 2014, Tripepi et al., 2012, Schulze et al., 2021, Shalev et al., 2018, Kandiba et al., 2016); gene regulation (Qi et al., 2016, Rawls et al., 2010, Hattori et al., 2016, Hwang et al., 2017, Johnsen et al., 2015, Tästensen et al., 2020, Reinhardt et al., 2019); microproteins (Zahn et al., 2021, Nagel et al., 2019, Kubatova et al., 2020) and small noncoding RNAs (sRNAs) (Straub et al., 2009, Heyer et al., 2012, Babski et al., 2014, Wyss et al., 2018).

Genome annotations are frequently compromised by annotation errors (Schnoes et al., 2009, Pfeiffer and Oesterhelt, 2015, Promponas et al., 2015, Danchin et al., 2018). Many of these errors are caused by invalid annotation transfer between presumed homologs, which, once introduced, are further spread by annotation robots. This problem can be partially overcome by using a Gold Standard Protein (GSP) based annotation strategy (Pfeiffer and Oesterhelt, 2015). Because the GSP has itself been subjected to experimental analysis, its annotation cannot be caused by an invalid

annotation transfer process. The GSP strategy had already been applied to a detailed analysis of the metabolism of halophilic archaea (Falb et al., 2008). However, with decreasing level of sequence identity, the assumption of isofunctionality becomes increasingly insecure. Although this may be counterbalanced by additional evidence, e.g. gene clustering, experimental confirmation would be the best option for validation of the annotation.

There are additional and much more subtle genome annotation problems. In some cases, GSPs are true homologs and the annotated function in the database is correct. Nevertheless, the biological context in the query organism makes it unlikely that the homologs are isofunctional, e.g. when the substrate of the GSP is lacking in the query organism. Also, paralogs may have distinct but related functions, which cannot be assigned by sequence analysis but may be assigned based on phylogenetic considerations. Here, again, experimental confirmation is the preferred option for validation of the annotation. Lack of experimental confirmation may keep high-level databases like KEGG or the SwissProt section of UniProt from adopting assignments based on well-supported bioinformatic analyses, so that the database entries continue to provide information which is probably incorrect. We refer to annotation problems in these databases solely to underscore that the biological issues which we raise are far from trivial. We have no intention to question the exceedingly high quality of the SwissProt and KEGG databases (UniProt, 2021, Kanehisa et al., 2019) and their tremendous value for the scientific community.

In this study, we describe a number of annotation issues for haloarchaea with a strong emphasis on *Hfx. volcanii*. We denote such cases as ‘open annotation issues’ with the hope of attracting members of the *Haloferax* community and other groups working with halophilic archaea to apply experimental analyses to elucidate the true function(s) of these proteins. This will increase the number of Gold Standard Proteins which originate from *Hfx. volcanii* or other haloarchaea, reduce genome annotation ambiguities, and perhaps uncover novel metabolic processes.

2 MATERIALS AND METHODS

2.1 Curation of genome annotation and Gold Standard Protein identification

The Gold Standard Protein based curation of haloarchaeal genomes has been described (Pfeiffer and Oesterhelt, 2015). Since then, a systematic comparison to KEGG data was performed for a subset of the curated genomes (Pfeiffer et al., 2020). The *Hfx. volcanii* genome annotation is continuously scrutinized, especially when a closely related genome is annotated (Tittes et al., 2021). The 16 haloarchaeal genomes which are currently under survey are listed in Suppl. Table S3.

2.2 Additional bioinformatics tools

Key databases were UniProtKB/Swiss-Prot (UniProt, 2021), InterPro (Hunter et al., 2009), KEGG (Kanehisa et al., 2019), and OrthoDB (Kriventseva et al., 2019). The SyntTax server was used for inspecting coservation of gene neighbourhood analysis (Oberto, 2013). As general tools the BLAST suite of programs (Johnson et al., 2008, Altschul et al., 1997) was used for genome comparisons.

3 RESULTS

Open issues are organised below under the subsections (3.1), the respiratory chain and oxidative decarboxylation; (3.2), amino acid metabolism; (3.3), heme and cobalamin biosynthesis; (3.4), coenzyme F420; (3.5), tetrahydrofolate as opposed to methanopterin; (3.6), NAD and riboflavin; (3.7), lipid metabolism; (3.8), genetic information processing, and (3.9), stand-alone (miscellaneous) cases.

3.1 The respiratory chain and oxidative decarboxylation

In the respiratory chain, coenzymes that have been reduced during catabolism (e.g. glycolysis) are reoxidized, with the energy being saved as an ion gradient. The textbook example of a respiratory chain are the five mitochondrial complexes (Rich and Marechal, 2010, Guo et al., 2018): complex I (NADH dehydrogenase), complex II (succinate dehydrogenase), complex III (cytochrome bc_1 complex), complex IV (cytochrome-c oxidase as prototype for a terminal oxidase) and complex V (F-type ATP synthase). In mitochondria, a significant part of the NADH which feeds into the respiratory chain originates from oxidative decarboxylation: conversion of pyruvate to acetyl-CoA by the pyruvate dehydrogenase complex and conversion of alpha-ketoglutarate to succinyl-CoA by the homologous 2-oxoglutarate dehydrogenase complex. While complex I and II transfer reducing elements to a lipid-embedded two-electron carrier (ubiquinone), the bc_1 complex transfers the electrons to the one-electron carrier cytochrome-c, a heme (and thus iron) protein, which then transfers electrons to the terminal oxidase.

Bacteria like *Escherichia coli* and *Paracoccus* have related complexes and enzymes: NADH dehydrogenase (encoded by the *nuo* operon), succinate dehydrogenase (encoded by *sdhABCD*) and the related fumarate reductase (encoded by *frdABCD*) (Crofts et al., 2013), several terminal oxidases (e.g. products of *cyoABCDE*, *cydABC*), and an F-type ATP synthase (encoded by *atp* genes). *E. coli* lacks a bc_1 complex, which, however, occurs in *Paracoccus denitrificans* (Kaila and Wikstrom, 2021). *E. coli* contains the canonical complexes of oxidative decarboxylation (pyruvate dehydrogenase complex, encoded by *aceEF+lpdA*, and 2-oxoglutarate dehydrogenase complex, encoded by *sucAB+lpdA*).

The respiratory chain of *Hfx. volcanii* and other haloarchaea deviates considerably from those of mitochondria and bacteria such as *Paracoccus* and *E. coli* (reviewed by (Schafer et al., 1999)), and a number of questions remain unresolved. We focus on the equivalents of complex I, III and IV, because these have unresolved issues. We also cover some aspects relevant for NADH levels (oxidative decarboxylation enzymes and type II NADH dehydrogenase). We do not cover complexes which have already been studied in haloarchaea: complex II (succinate dehydrogenase) (Scharf et al., 1997, Sreeramulu et al., 1998, Gradin et al., 1985) and complex V (ATP synthase) (Steinert et al., 1997, Nanba and Mukohata, 1987).

(a) In haloarchaea, oxidative decarboxylation is not linked to reduction of NAD to NADH but to reduction of a ferredoxin (encoded by *fdx*, e.g. OE_4217R, HVO_2995) which has a redox potential similar to that of the NAD/NADH pair (Kerscher and Oesterhelt, 1977). The enzymes for oxidative decarboxylation are pyruvate--ferredoxin oxidoreductase (*porAB*, e.g. OE_2623R/2622R, HVO_1305/1304) and 2-oxoglutarate--ferredoxin oxidoreductase (*korAB*, e.g. OE_1711R/1710R, HVO_0888/0887), and these have been characterized from *Halobacterium salinarum* (Plaga et al., 1992, Kerscher and Oesterhelt, 1981b, Kerscher and Oesterhelt, 1981a).

(b) In *Hfx. volcanii*, ferredoxin Fdx (HVO_2995) plays an essential role in nitrate assimilation (Zafrilla et al., 2011). It may well be involved in additional metabolic processes and it is yet unresolved how ferredoxin Fdx is reoxidised, but it may well be possible that this is achieved by the Nuo complex.

(c) The *nuo* cluster of haloarchaea resembles that of *E. coli*, with genes and gene order highly conserved, and just a few domain fissions and fusions. However, haloarchaea lack NuoEFG (Falb et al., 2005), which is a subcomplex mediating interaction with NADH (Leif et al., 1995, Braun et al., 1998). Thus, the haloarchaeal *nuo* complex is unlikely to function as NADH dehydrogenase, despite its annotation as such in KEGG (as of April 2021).

(d) Other catabolic enzymes generate NADH, which also must be reoxidized. Based on inhibitor studies in *Hbt. salinarum*, NADH is not reoxidized by a type I but rather by a type II NADH dehydrogenase (Sreeramulu et al., 1998). A gene has been assigned for *Natronomonas pharaonis* (Falb et al., 2008). However, for reasons detailed in Suppl.Text.S1, this assignment is highly questionable, so that this issue calls for experimental analysis.

(e) About one-third of the haloarchaea, especially the *Natrialbales*, do not code for a complex III equivalent (cytochrome bc_1 complex encoded by *petABC*) according to OrthoDB analysis. The bc_1 complex is required to transfer electrons from the lipid-embedded two-electron carrier (menaquinone in haloarchaea) to the one-electron carrier associated with terminal oxidases (probably halocyanin). How electrons flow in the absence of a complex III equivalent is currently unresolved.

The haloarchaeal *petABC* genes resemble those of the chloroplast b6-f complex rather than those of the mitochondrial bc_1 complex (see Suppl.Text S1 Section 1 for more details).

(f) A bc cytochrome has been purified from *Nmn. pharaonis*, but with an atypical 1:1 ratio between the b-type and c-type heme (Scharf et al., 1997). The complex is heterodimeric, with subunits of 18 kDa and 14 kDa. The 18 kDa subunit carries the covalently attached heme group (Scharf et al., 1997). An attempt was made to identify the genes coding for these subunits (Mattar, 1996) (for details see Suppl.Text S1 Section 1). Two approaches were used to obtain protein sequence data, one being N-terminal protein sequencing of the two subunits extracted from a SDS-polyacrylamide gel. In the other attempt, peptides from the purified complex were separated by HPLC, and a peptide absorbing at 280 nm (protein) as well as 400 nm (heme) was isolated. Absorption at 400 nm clearly indicates covalent attachment of the heme group to the peptide. The sequences from the two approaches overlapped and resulted in a contigous sequence of 41 aa, with only the penultimate position remaining undefined (Mattar, 1996). Based on this information, a PCR probe was generated

(designated “cyt-C Sonde”) that allowed the gene to be identified and sequenced, including its genomic neighbourhood. It turned out that the genes coding for the four subunits of succinate dehydrogenase (*sdhCDBA*) had been isolated. The obtained protein sequence corresponds to the N-terminal region of *sdhD* (with the initiator methionine cleaved off) and only 2 sequence discrepancies in addition to the unresolved penultimate residue.

In the PhD thesis (Mattar, 1996), this unambigous result was rated to be a failure (and the data were never formally published). The reason is that SdhD is free of cysteine residues, while textbook knowledge states that a pair of cysteines is required for covalent heme attachment (Kletzin et al., 2015). The lack of the required cysteine pair was taken to indicate that the results were incorrect and that the identified genes did not encode the cytochrome bc that the study had been seeking (Mattar, 1996). In contrast, we speculate that the results were completely correct, despite being in conflict with the cysteine pair paradigm. In our view, a paradigm shift is required. The obtained results call for a yet unanticipated novel mode of covalent heme attachment, exemplified by the 18 kDa subunit of *Natronomonas* succinate dehydrogenase subunit SdhD. It should be noted that the 41 aa protein sequence, which had been obtained, turned out to contain three histidine residues upon translation of the gene, but none of these had been detected upon Edman degradation.

In *Halobacterium*, a small c-type cytochrome was purified (cytochrome c_{552}, 14.1 kDa) (Sreeramulu, 2003). Heme staining after SDS-PAGE indicated a covalent heme attachment, but no sequence or composition data were reported, so that it is not possible to identify the protein based on the available information. We speculate that the *Halobacterium* cytochrome c_{552} also represents SdhD (as detailed in Suppl.Text S1 Section 1). In that case, the proposed novel type of covalent heme attachment would not be restricted to *Nmn. pharaonis* but might be a general property of haloarchaea. This would also solve the “*Halobacterium* paradox” (Kletzin et al., 2015).

(g) The haloarchaeal one-electron carrier is the copper protein halocyanin rather than the iron-containing heme protein cytochrome-c. A halocyanin from *Nmn. pharaonis* (NP_3954A) has been characterized, including its redox potential (Mattar et al., 1994, Scharf and Engelhard, 1993, Hildebrandt et al., 1994). A gene fusion supports the close connection of a halocyanin with a subunit of a terminal oxidase. For further details see Suppl.Text S1 Section 1.

(h) Terminal oxidases are highly diverse in haloarchaea and we restrict our analysis to three species (*Nmn. pharaonis*, *Hfx. volcanii*, and *Hbt. salinarum*) because in each of these at least one terminal oxidase has been experimentally studied. Details are described in Suppl. Text S1 Section 1.

(i) NAD-dependent oxidative decarboxylation is a canonical reaction to convert pyruvate into acetyl-CoA, and alpha-ketoglutarate into succinyl-CoA. In haloarchaea, the conversion of pyruvate to acetyl-CoA and alpha-ketoglutarate to succinyl-CoA is dependent on ferredoxin, not on NAD (see above). Nevertheless, most haloarchaeal genomes also code for homologs of enzymes catalyzing NAD-dependent oxidative decarboxylation, such as the *E. coli* pyruvate dehydrogenase complex. In most cases, the substrates could not be identified, an exception being a paralog involved in isoleucine catabolism (Sisignano et al., 2010). In several cases the enzymes were found not to show catalytic activity with pyruvate or alpha-ketoglutarate (see Suppl.Text S1 Section 1 for details). Also, a conditional lethal *porAB* mutant was unable to grow on glucose or pyruvate, thus excluding that alternative enzymes for conversion of pyruvate to acetyl-CoA exist in *Hfx. volcanii* (Kuprat et al., 2021). Nonetheless, despite experimental results to the contrary, pyruvate is assigned as substrate for some of the homologs of the pyruvate dehydrogenase complex in KEGG (as of April 2021).

3.2 Amino acid metabolism

While most amino acid biosynthesis and degradation pathways can be reliably reconstructed, a few open issues remain, which are discussed below.

(a) The first and last steps of arginine biosynthesis deal with blocking and unblocking of the alpha-amino group of the substrate (glutamate) and a product intermediate (ornithine). As detailed in Suppl. Text S1 Section 2, it is highly likely that glutamate is attached to the gamma-carboxyl group of a carrier protein, and ornithine is released from that carrier protein. This is based on characterized proteins from *Thermus thermophilus* (Horie et al., 2009), *Thermococcus kodakarensis* (Yoshida et al., 2016) and *Sulfolobus acidocaldarius* (Ouchi et al., 2013). The assignment is strongly supported by clustering of the arginine biosynthesis genes. Some of the homologs are bifunctional, being involved in arginine biosynthesis but also in lysine biosynthesis via the prokaryotic variant of the alpha-aminoadipate pathway. This ambiguity is not assumed to occur in haloarchaea, which use the diaminopimelate pathway for Lys biosynthesis (Hochuli et al., 1999) (see Suppl. Text S1 Section 2 for further discussion of this issue).

Expanding the above, we provide full details underlying our reconstruction of arginine and lysine biosynthesis in *Hfx, volcanii* in Table 1.

(b) Archaea use a different precursor for aromatic amino acid biosynthesis than the classical pathway. This has been resolved for *Methanocaldococcus jannaschii* and for *Methanococcus maripaludis* (White, 2004, Porat et al., 2006). However, the initial steps may differ from those reported for *Methanocaldococcus* in that fructose-1,6-disphosphate rather than 6-deoxy-5-ketofructose might be a substrate (Gulko et al., 2014). Up to now, a clean deletion of the corresponding enzymes and confirmation with in vitro assays has not yet been achieved (for details see Suppl. Text S1 Section 2).

(c) The gene for tryptophanase (*tpa*) is stringently regulated in *Haloferax*, which is the basis for using its promoter in the toolbox for regulated gene expression (Large et al., 2007). The shutdown of this gene avoids tryptophan degradation when supplies are scarce. Tryptophanase cleaves Trp into indole, pyruvate and ammonia. The fate of indole is, however, yet unresolved.

(d) A probable histidine utilization cluster exists, based on characterized homologs from *Bacillus subtilis*, but has not yet been experimentally verified.

(e) Among 16 auxotrophic mutants observed in a *Hfx. volcanii* transposon insertion library (Kiljunen et al., 2014), some could grow only in the presence of one (or several) supplied amino acids. In many cases, the affected genes were known to be involved in the corresponding pathway, but the others may lead to novel function assignments. One affected gene resulted in histidine auxotrophy and the product of this gene (HVO_0431) is an interesting candidate. The InterPro domain assignment (HAD family hydrolase) fits to the only remaining pathway gap in histidine biosynthesis (histidinol-phosphatase). In this context it should be noted that the enzyme which catalyzes the preceding reaction (encoded by *hisC*) is part of a highly conserved three-gene operon involved in polar lipid biosynthesis (see below). For details see Suppl. Text S1 Section 2. One affected gene resulted in isoleucine auxotrophy. The product of this gene (HVO_0644) is currently annotated to catalyze two reactions, one being an early step in isoleucine biosynthesis (EC 2.3.1.182), the other being the first step after leucine biosynthesis branches off from valine biosynthesis (EC 2.3.3.13) (see below, f) (for details see Suppl. Text S1 Section 2).

(f) *Hfx. volcanii* codes for two paralogs to which a function as 2-isopropylmalate synthase (EC 2.3.3.13) has been attributed. This is the first reaction specific to leucine biosynthesis, when the pathway branches off valine biosynthesis. One paralog, HVO_0644, is annotated as bifunctional, also catalyzing a chemically similar reaction which is an early step in isoleucine biosynthesis (EC 2.3.1.182). When the gene encoding HVO_0644 is disrupted by transposon integration, cells cannot grow in the absence of isoleucine. It is unclear if the protein is really bifunctional and is really involved in leucine biosynthesis, catalyzing the reaction of EC 2.3.3.13. The other paralog, HVO_1510, belongs to an ortholog set with major problems concerning start codon assignment. The ortholog set from the 16 genomes listed in Suppl. Table S3 were analyzed. When only canonical start codons are considered (ATG, GTG, TTG), then the orthologs from *Haloferax mediterranei*, *Nmn. pharaonis*, *Natronomonas moolapensis* and *Halohasta litchfieldiae* either lack a long highly conserved N-

terminal region, or they are disrupted (pseudogenes), being devoid of a potential start codon. The gene from *Hfx. volcanii* has a start codon (GTG) which is consistent to that of *Haloferax gibbonsii* strain LR2-5 (but a GTA in *Hfx. gibbonsii* strain ARA6). In this region, the gene from *Hfx. mediterranei* is closely related but has in-frame stop codons. HVO_1510 is considerably longer than the orthologs from *Haloquadratum walsbyi*, *Haloarcula hispanica*, and *Natrialba magadii*. The first alternative start codon for HVO_1510 codes for Met-93. This protein was proteomically identified in three ArcPP datasets (Schulze et al., 2020), and peptides upstream of Met-93 were identified. This gene might be translated from an atypical start codon, either an in-frame CTG, or an out-of-frame ATG, which would require ribosomal slippage (for details see Suppl. Text S1 Section 2). It is tempting to speculate that translation occurs only when leucine is not available.

3.3 Coenzymes I: cobalamin and heme

The classical heme biosynthesis pathway branches off cobalamin biosynthesis at the level of uroporphyrinogen III. The alternative heme biosynthesis pathway (Bali et al., 2011), which is used by haloarchaea, has an additional common step, the conversion of uroporphyrinogen III to precorrin-2. For heme biosynthesis, precorrin-2 is converted to siroheme. This pathway has been reconstructed (Siddaramappa et al., 2012), except for the iron insertion step. For de novo cobalamin biosynthesis, haloarchaea use the cobalt-early pathway with a cobalt-dependent key reaction being catalyzed by CbiG (Moore et al., 2013). Several aspects of heme and cobalamin biosynthesis in haloarchaea are yet unresolved.

(a) *Hfx. volcanii* contains two annotated *cbiX* genes. For reasons detailed in Suppl.Text S1 Section 3, we predict that one is a cobaltochelatase, involved in cobalamin biosynthesis, while the other is a ferrochelatase, responsible for conversion of precorrin-2 to siroheme in the alternative heme biosynthesis pathway.

(b) De novo cobalamin biosynthesis has been extensively reconstructed upon curation of the genome annotation (Pfeiffer and Oesterhelt, 2015). All enzymes of the pathway and their associated GSPs are listed in Table 1. Only two pathway gaps remained, and

because these are consecutive, it may be possible that the haloarchaeal pathway is non-canonical and proceeds via a novel biosynthetic intermediate. There are only four genes with yet unassigned function in the *Hfx. volcanii* cobalamin gene cluster, and their synteny is well conserved in the majority of haloarchaeal genomes. Thus, these genes are obvious candidates for filling the pathway gaps (for details see Suppl.Text S1 Section 3).

(c) The cobalamin biosynthesis and salvage reactions (those beyond ligand cobyrinate a,c diamide) involve "adenosylation of the corrin ring, attachment of the aminopropanol arm, and assembly of the nucleotide loop that bridges the lower ligand dimethylbenzimidazole and the corrin ring" (Rodionov et al., 2003). The enzymes of these braches of cobalamin biosynthesis and their associated GSPs are listed in Table 1. Only two pathway gaps remain open. For one of these, a candidate was proposed upon detailed bioinformatic analysis (Rodionov et al., 2003) (for further details see Suppl.Text S1 Section 3).

(d) Haloarchaea may code for a late cobaltochelatase of the heterotrimeric type. Distantly related GSPs are either cobalt or magnesium chelatases. A late cobaltochelatase is not required for de novo cobalamin biosynthesis via the cobalt-early pathway. We speculate that it may be involved in cobalamin salvage. The chelatase has a mosaic subunit structure as also reported previously (Rodionov et al., 2003) (see Suppl.Text S1 Section 3 for details).

(e) In the alternative heme biosynthesis pathway, siroheme is decarboxylated to 12,18-didecarboxysiroheme, which is attributed to the proteins encoded by *ahbA* and *ahbB*. These are homologous to each other and are organized as two two-domain proteins. It is unclear if AhbA and AhbB function independently or if they form a complex.

3.4 Coenzymes II: coenzyme F420

Even though coenzyme F420 is predominantly associated with methanogenic archaea (Eirich et al., 1979, Jaenchen et al., 1984), it occurs also in bacteria and a small amount of this coenzyme has been detected in non-methanogenic archaea, including halophiles (Lin and White, 1986). The genes required for the biosynthesis of this coenzyme are encoded in haloarchaeal genomes, but the origin and attachment of the phospholactate moiety are not completely resolved (see below). To the best of our knowledge, only a single coenzyme F420 dependent enzymatic reaction has yet been reported for halophilic archaea (de Wit and Eker, 1987). Thus, the importance of this coenzyme in haloarchaeal biology is currently enigmatic and awaits experimental analysis.

(a) The pathway which creates the carbon backbone of this coenzyme has been reconstructed. We list the enzymes with their associated GSPs in Table 1. Coenzyme F420 contains a phospholactate moiety which was reported to orginate from 2-phospho-lactate (Grochowski et al., 2008), but this compound is not well connected to the remainder of metabolism. As summarized in Suppl.Text S1 Section 4, there are various new insights regarding this pathway from recent studies in other prokaryotes (Bashiri et al., 2019, Braga et al., 2019). To the best of our knowledge, the haloarchaeal coenzyme F420 biosynthesis pathway has never been experimentally analyzed.

(b) The prediction of coenzyme F420-specific oxidoreductases in *Mycobacterium* and actinobacteria has been reported (Selengut and Haft, 2010), leading to patterns and domains which are also found in haloarchaea. Several such enzymes are described in Suppl.Text S1 Section 4.

(c) HVO_1937 might be a coenzyme F420-dependent 5,10-methylenetetrahydrofolate reductase (see also below, C1 metabolism, and Suppl.Text S1 Section 4).

(d) The precursor for coenzyme F420 may be used by a photo-lyase involved in DNA repair.

3.5 Coenzymes III: coenzymes of C1 metabolism: tetrahydrofolate in haloarchaea, methanopterin in methanogens

Halophilic and methanogenic archaea use distinct coenzymes as one-carbon carrier (C1 metabolism): tetrahydrofolate in haloarchaea and methanopterin in methanogens (White, 1988, Maden, 2000). Several characterized methanogenic proteins that act on or with methanopterin have comparably close homologs in haloarchaea, which results in misannotation of haloarchaeal proteins (e.g. in SwissProt) as being involved in methanopterin biology. We assume that the haloarchaeal proteins function with the haloarchaeal one-carbon carrier tetrahydrofolate and that this shift in coenzyme specificity is possible due to the structural similarity between methanopterin and tetrahydrofolate (a near-identical core structure consists of a pterin heterocyclic ring linked via a methylene bridge to a phenyl ring; both also have a polyglutamate tail). A detailed review on the many variants of the tetrahydrofolate biosynthetic pathway is available (de Crecy-Lagard, 2014).

(a) Folate biosynthesis requires aminobenzoate. We had proposed candidates for a pathway from chorismate to para-aminobenzoate (Falb et al., 2008, Pfeiffer et al., 2008b) (for details see Suppl.Text S1 Section 5). However, these predictions have not been adopted by KEGG (accessed April 2021) and without experimental confirmation this is unlikely to ever happen.

(b) GTP cyclohydrolase MptA (HVO_2348) catalyzes a reaction in the common part of tetrahydrofolate and methanopterin biosynthesis. The enzymes specific for methanopterin biosynthesis are absent from haloarchaea and thus the assignment of HVO_2348 to the methanopterin biosynthesis pathway in UniProt is invalid (accessed March 2021).

The next common pathway step (EC 3.1.4.56) has been resolved in *M. jannaschii* (MJ0837) but still is a pathway gap in halophilic archaea. MJ0837 is very distantly

related to HVO_A0533, which thus is a promising candidate for experimental analysis.

HVO_2628 shows 30% protein sequence identity to the enzyme catalyzing the first committed step to methanopterin biosynthesis. As detailed in Suppl.Text S1 Section 5, we consider it likely that it does not catalyze that reaction.

(c) Two enzymes which alter the oxidation level of the coenzyme-attached one-carbon compound probably function with tetrahydrofolate, even though their methanogenic homologs function with methanopterin. In contrast to their assignments in KEGG and UniProt (as of March 2021), their probable functions are thus methenyltetrahydrofolate cyclohydrolase (HVO_2573) and 5,10-methylenetetrahydrofolate reductase (HVO_1937) (see Suppl.Text S1 Section 5)

3.6 Coenzymes IV: NAD and FAD (riboflavin)

(a) The energy source for NAD kinase may be ATP or polyphosphate. This is unresolved for the two paralogs of probable NAD kinase (HVO_2363, *nadK1*, HVO_0837, *nadK2*). These show only 25% protein sequence identity to each other (see Suppl. Text S1 Section 6). Polyphosphate was not found in exponentially growing *Hfx. volcanii* cells (Zerulla et al., 2014), so that ATP is the more likely energy source.

(b) HVO_0782 is an enzyme involved in NAD biosynthesis, which is encoded in most haloarchaeal and archaeal genomes. The adjacent gene, HVO_0781, is encoded in nearly all haloarchaeal genomes according to OrthoDB, and with very strong syntenic coupling revealed by SyntTax analysis. Thus, HVO_0781 is a candidate to also be involved in NAD biosynthesis. Characterized homologs to HVO_0781 decompose S-adenosyl-methionine into methionine and adenosine, a reaction which seems wasteful and might not be expected to be highly conserved with respect to existence and gene clustering (see Suppl. Text S1 Section 6).

(c) We describe the reconstruction of riboflavin biosynthesis based on a detailed bioinformatic reconstruction (Rodionova et al., 2017). Three pathway gaps remain, with candidate genes predicted for two of these (Rodionova et al., 2017) (for details see Suppl. Text S1 Section 6).

3.7 Biosynthesis of membrane lipids, bacterioruberin and menaquinone

Archaeal membrane lipids contain ether-linked isoprenoid side chains (see (Caforio and Driessen, 2017) and references cited therein). The isoprenoid precursor isopentenyl diphosphate is synthesized in haloarchaea by a modified version of the mevalonate pathway (Vannice et al., 2014). Isoprenoid units are then linearly condensed to the C20 compound geranylgeranyl diphosphate. The haloarchaeal core lipid, archaeol, consists of 2,3-sn-glycerol with two C20 isoprenoid side chains attached by ether linkages. In some archaea, especially alkaliphiles, C25 isoprenoids are also found (see e.g. (De and Gambacorta, 1988, Dawson et al., 2012)). Also, a number of distinct headgroups are found in polar lipids (phospholipids) (reviewed in (Caforio and Driessen, 2017)). Even though polar lipids are used as important taxonomic markers (Oren et al., 1997) their biosynthetic pathways are not completely resolved.

Haloarchaea typically have a red color, which is due to carotenoids, mainly the C50 carotenoid bacterioruberin (Oren, 2002, Kushwaha et al., 1975, Yang et al., 2015). For carotenoid biosynthesis, two molecules of geranylgeranyl diphosphate, a C20 compound, are linked head to head to generate phytoene, which is desaturated to lycopene (Falb et al., 2008, Giani et al., 2020). The pathway from lycopene to the C50 compound bacterioruberin has been experimentally characterized (Dummer et al., 2011, Yang et al., 2015).

(a) We assigned HVO_2725 (*idsA1*, paralog of NP_3696A) and HVO_0303 (*idsA2*, paralog of NP_0604A) for the linear isoprenoid condensation reactions resulting in a C20 isoprenoid (EC 2.5.1.10, 2.5.29, short chain isoprenyl diphosphate synthase) (see also Suppl.Text S1 Section 7). Some archaea, mainly haloalkaliphiles, also contain C25 isoprenoid side chains. Geranylfarnesyl diphosphate synthase, the enzyme which

generates the C25 isoprenoids, has been purified and enzymatically characterized from *Nmn. pharaonis* (Tachibana, 1994), but data which allow the assignment to a specific gene have not been collected. Three paralogous genes from *Nmn. pharaonis* are candidates for this function (NP_0604A, NP_3696A, and NP_4556A). Because NP_0604A and NP_3696A have orthologs in *Hfx. volcanii*, a species devoid of C25 lipids, we assign the synthesis of C25 isoprenoids (geranylfarnesyl diphosphate synthase activity) to the third paralog, NP_4556A. UniProt assigns C25 biosynthesis activity to NP_3696A for undescribed reasons (as of April 2021) and KEGG does not make this assignment for any of the three paralogs (as of April 2021). Our assignments are supported by analysis of key residues which determine the length of the isoprenoid chain (Bale et al., 2019). These authors label the cluster containing NP_3696A (WP011323557.1) as “C15/C20” and the cluster containing NP_4556A (WP011323984.1) as “C20->C25->C30?”.

(b) Typical polar lipids in haloarchaea are phosphatidylglycerophosphate methyl ester (PGP-Me) and phosphatidylglycerol (PG), but also phosphatidylglycerosulfate (PGS) (Kates, 1993, Kates et al., 1993, Bale et al., 2019). Other polar lipids are archaetidylserine and its decarboxylation product archaetidylethanolamine, both of which are found in rather low quantities in *Haloferax* (Kellermann et al., 2016). A third group of polar lipids has a headgroup derived from myo-inositol. The biosynthetic pathway of the head groups is only partially resolved. One CDP-archaeol 1-archaetidyltransferase which is part of a highly conserved three-gene operon may attach either glycerol phosphate or myo-inositol phosphate. In Suppl. Text S1 Section 7 we summarize arguments in favor of each of these candidates, but the true function can only be decided by experimental analysis.

(c) Carotenoid biosynthesis involves the head-to-head condensation of the C20 isoprenoid geranylgeranyl diphosphate to phytoene, which is desaturated to lycopene (Falb et al., 2008, Giani et al., 2020). The *crtB* gene product (e.g. HVO_2524) catalyzes the head-to-head condensation. It is yet uncertain which gene product is responsible for the desaturation of phytoene to lycopene. The further pathway from lycopene to bacterioruberin has been experimentally characterized in *Haloarcula japonica* (Yang et al., 2015). A three gene cluster (*crtD-lyeJ-cruF*) codes for the three

enzymes of this pathway. The synteny of this three gene cluster is strongly conserved according to SyntTax analysis. Several genes which are certainly or possibly involved in carotenoid biosynthesis are encoded in the vicinity of this cluster (for details see Suppl. Text S1 Section 7).

(d) Halophilic archaea contain menaquinone as a lipid based two electron carrier of the respiratory chain (Elling et al., 2016, Kellermann et al., 2016). We describe the reconstruction of the menaquinone biosynthesis pathway, with two pathway gaps remaining open (see Suppl.Text S1 Section 7 for details).

3.8 Issues concerning RNA polymerase, protein translation components and signal peptide degradation

(a) Haloarchaeal RNA polymerase consists of a set of canonical subunits (encoded by *rpoA1A2B1B2DEFHKLNP*). *Hbt. salinarum* and a subset of other haloarchaea contain an additional subunit called epsilon (Leffers et al., 1989, Madon and Zillig, 1983). Purified RNA polymerase containing the epsilon subunit transcribes native templates efficiently, in contrast to RNA polymerase devoid of this subunit (Madon and Zillig, 1983). The biological relevance of this subunit is enigmatic (see Suppl.Text S1 Section 8).

(b) Two distant paralogs are found for haloarchaeal ribosomal protein S10 in nearly all haloarchaeal genomes. It is uncertain if both occur in the ribosome, whether they occur together or are mutually exclusive. The latter distribution would result in heterogeneity of the ribosomes. Alternatively, one of the paralogs may exclusively have a non-ribosomal function.

In a subset of archaea, two distant paralogs are found for haloarchaeal ribosomal protein S14 (ca 20% of the genomes, e.g. in *Nmn. pharaonis*). For more details see Suppl.Text S1 Section 8.

(c) The ribosomal protein L43e shows heterogeneity with respect to the present of C2-C2 type zinc finger motif. This zinc finger is found in L34e from all *Halobacteriales* and all euryarchaeal proteins outside the order *Halobacteria*, but is not found in *Haloferacales* and very rare in Natrialbales, . Eukaryotic orthologs (e.g. from rat and yeast) contain this zinc finger and its biological importance has been experimentally shown for the yeast protein (Rivlin et al., 1999) (for details see Suppl.Text S1 Section 8).

(d) Diphthamide is a complex covalent modification of a histidine residue of translation elongation factor a-EF2. The pathway has been reconstructed, based on distant homologs (enzymes encoded by *dph2*, *dph5*) and by detailed bioinformatic analysis (enzyme encoded by *dph6*) (de Crecy-Lagard et al., 2012) (for details see Suppl.Text S1 Section 8). These uncertain function assignments await experimental confirmation.

(e) N-terminal signal sequences target proteins to the secretion machinery. Subsequent to membrane insertion or transmembrane transfer, the signal sequence is cleaved off by signal peptidase. After cleavage, the signal peptide must be degraded to avoid clogging of the membrane. Degradation is catalyzed by signal peptide peptidase. Candidates for this activity have been predicted from two protein families (Ng et al., 2007, Raut et al., 2021) (for details see Suppl.Text S1 Section 8).

3.9 Miscellaneous metabolic enzymes and proteins with other functions

Here, we list a few other enzymatic or non-enzymatic functions for which candidate genes have been assigned, but without experimental validation.

(a) Ketohexokinase from *Haloarcula vallismortis* has been experimentally characterized (Rangaswamy and Altekar, 1994). However, the activity was not assigned to a gene. Detailed bioinformatic analyses have been made (Anderson et al., 2011, Williams et al., 2019) and point to a small set of orthologs represented by

Hmuk_2662, the ortholog of HVO_1812 (for further details see Suppl. Text S1 Section 9).

(b) The assignment of fructokinase activity to the *Hht. litchfieldiae* candidate gene halTADL_1913 (UniProt:A0A1H6QYL4) is based on differential proteomic analysis (Williams et al., 2019) (see Suppl. Text S1 Section 9 for details). Very close homologs are rare in haloarchaea. For this protein family (carbohydrate kinase) it is unclear if more distant homologs (with about 50% protein sequence identity) are isofunctional.

(c) A candidate gene for glucoamylase is HVO_1711 for reasons described in Suppl. Text S1 Section 9. The enyzme from *Halorubrum sodomense* has been characterized (Chaga et al., 1993) but the activity was not yet assigned to a gene.

(d) A strong candidate for having glucose-6-phosphate isomerase activity is *Hfx. volcanii* (HVO_1967, pgi), based on 36% protein sequence identity to the characterized enzyme from *M. jannaschii* (MJ1605) (Rudolph et al., 2004).

(e) A candidate gene for specifying an enzyme with 2-dehydro-3-deoxy-(phospho)gluconate aldolase activity is *Hbt. salinarum kdgA* (OE_1665R). It is rather closely related (36% protein sequence) to *Hfx, volcanii* HVO_1101 (encoded by *dapA*), which is involved in lysine biosynthesis, a biosynthetic pathway which is absent from *Hbt. salinarum*. The function assignment is based on distant homologs from *Saccharolobus (Sulfolobus) solfataricus* and *Thermoproteus tenax* which have been characterized (Ahmed et al., 2005) (for details see Suppl. Text S1 Section 9).

(f) Haloarchaea may contain an NAD-independent L-lactate dehydrogenase, LudBC (HVO_1692 and HVO_1693). Deletion of this gene pair impairs growth on rhamnose, which is catabolized to pyruvate and lactate (Reinhardt et al., 2019). There is a very distant relationship (for details see Suppl.Text S1 Section 9) to the LldABC subunits of the characterized L-lactate dehydrogenase from *Pseudomonas stutzeri* A1501 (Gao

et al., 2015) and to the LutABC proteins from *B. subtilis*, which have been shown to be involved in lactose utilization (Chai et al., 2009).

(g) *Hfx. volcanii* may be able to convert urate to allantoin, using the gene cluster HVO_B0299-HVO_B0302. This could be part of a complete degradation pathway for purines, which, however, has to be considered highly speculative (see Suppl.Text S1 Section 9).

(h) *Hfx. volcanii* may contain an enzyme having a "nickel-pincer cofactor". The biogenesis of this cofactor may be catalyzed by *larBCE* (as detailed in Suppl. Text S1 Section 9).

(i) cyclic diguanylate hydrolysis

Cyclic di-AMP (c-di-AMP) is an important nucleotide signalling molecule in bacteria and archaea. It is generated from two molecules of ATP by diadenylate cyclase (encoded by *dacZ*) and is degraded to pApA by phosphodiesterases (Corrigan and Grundling, 2013). The level of this signalling molecule is strictly controlled (Gundlach et al., 2015, Commichau et al., 2019), thus requiring a sophisticated interplay of cyclase and phosphodiesterase. DacZ from *Hfx. volcanii* has been characterized and it was shown that c-di-AMP levels must be tightly regulated (Braun et al., 2019). The degrading enzyme, however, has not yet been identified in *Haloferax* but candidates have been proposed (Corrigan and Grundling, 2013, He et al., 2020, Yin et al., 2020) (see Suppl. Text S1 Section 9).

(j) homolog to RNaseZ

HVO_2763 is distantly related to RNase Z (HVO_0144, *rnz*). The experimental characterization of HVO_2763 (Fischer et al., 2012) excluded activity as exonuclease but did not reveal its physiological function. Upon transcriptome analysis, the downregulation of several genes was detected. Several of these were uncharacterized at the time of the experiment but have since been shown to be involved in the minor

N-glycosylation pathway which was initially detected under low salt conditions (see Suppl. Text S1 Section 9 for further details) .

(k) A pair of genes (*dabAB*) is predicted to function as CO_2 transporter.

HVO_2410 and HVO_2411 probably function as carbon dioxide transporter, based on the identification of such transporters in *Halothiobacillus neapolitanus* (Desmarais et al., 2019). Being a member of the proton-conducting membrane transporter family, this protein may be misannotated as a subunit of the *nuo* or *mrp* complex (see Suppl. Text S1 Section 9 for further details).

4. Conclusion

We describe a large number of cases where protein function cannot be correctly predicted when restricting considerations to computational analyses without taking the biological context into account. An example is the switch from methanopterin to tetrahydrofolate as C1 carrier in haloarchaea. Homologous enzymes, inherited from the common ancestor, have adapted to the new C1 carrier, rather being replaced by non-homologous proteins. Function prediction tools may misannotate haloarchaeal proteins to work with methanopterin. Another example is the *nuo* complex and its misannotation as a type I NADH dehydrogenase. In other cases, even distant sequence similarity may allow a valid function prediction if additional evidence (e.g. from gene neighbourhood analysis or from detailed evaluation of metabolic pathway gaps) is taken into account. Examples include cobalamin cluster proteins, which probably close the two residual pathway gaps, and the predicted degradation pathway for purines. In all these cases, we have presented reasonable hypotheses based on current knowledge, and in many cases these are so well supported as to be compelling, but to be certain, experimental data are required. With this overview, we attempt to arouse the curiosity of colleagues, hoping that they will confirm or disprove our speculations and thus advance the knowledge about haloarchaeal biology. *Hfx. volcanii* is a model species for halophilic archaea, and the more complete and correctly its genome is annotated, the higher will be its value for systems biology analyses (modelling) and for synthetic biology (metabolic engineering) and biotechnology.

Acknowledgments

We thank all members of the *Haloferax* community for more than a decade of fruitful cooperation and for sharing their deep knowledge about the biology of this model species. We thank our colleagues for kindly reading and providing thoughtful comments on the manuscript before submission: Sonja-Verena Albers, Thorsten Allers, Maria Jose Bonete, Rosana de Castro, Anita Marchfelder, Mechthild Pohlschroder, Joerg Soppa. We thank Birgit Scharf for insightful discussions on the *Natronomonas* respiratory chain.

Data Availability Statement

None.

Funding

This research received no specific grant from any funding agency in the public, commercial, or non-for-profit sectors.

Authors contribution

Conceptualization, F.P.
Data curation, F.P.
Project administration, F.P.
Formal analysis, F.P., M.D-S.
Investigation F.P., M.D-S.
Validation, F.P., M.D-S.
Visualization, F.P., M.D-S.
Writing – original draft, F.P., M.D-S.;
Writing – review & editing, F.P., M.D-S.

Conflict of interest

None declared.

Ethics statement

Not applicable

REFERENCES

ABDUL HALIM, M. F., PFEIFFER, F., ZOU, J., FRISCH, A., HAFT, D., WU, S., TOLIC, N., BREWER, H., PAYNE, S. H., PASA-TOLIC, L. & POHLSCHRODER, M. 2013. *Haloferax volcanii* archaeosortase is required for motility, mating, and C-terminal processing of the S-layer glycoprotein. *Mol Microbiol,* 88**,** 1164-75.

ABDUL HALIM, M. F., RODRIGUEZ, R., STOLTZFUS, J. D., DUGGIN, I. G. & POHLSCHRODER, M. 2018. Conserved residues are critical for *Haloferax volcanii* archaeosortase catalytic activity: Implications for convergent evolution of the catalytic mechanisms of non-homologous sortases from archaea and bacteria. *Mol Microbiol,* 108**,** 276-87.

ABDUL-HALIM, M. F., SCHULZE, S., DILUCIDO, A., PFEIFFER, F., BISSON FILHO, A. W. & POHLSCHRODER, M. 2020. Lipid anchoring of archaeosortase substrates and midcell growth in haloarchaea. *mBio,* 11**,** e00349-20.

AHMED, H., ETTEMA, T. J., TJADEN, B., GEERLING, A. C., VAN DER OOST, J. & SIEBERS, B. 2005. The semi-phosphorylative Entner-Doudoroff pathway in hyperthermophilic archaea: a re-evaluation. *Biochem J,* 390**,** 529-40.

ALLERS, T., BARAK, S., LIDDELL, S., WARDELL, K. & MEVARECH, M. 2010. Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in *Haloferax volcanii*. *Appl Environ Microbiol,* 76**,** 1759-69.

ALLERS, T. & MEVARECH, M. 2005. Archaeal genetics - the third way. *Nat. Rev. Genet.,* 6**,** 58-73.

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res,* 25**,** 3389-402.

ANDERSON, I., SCHEUNER, C., GOKER, M., MAVROMATIS, K., HOOPER, S. D., PORAT, I., KLENK, H. P., IVANOVA, N. & KYRPIDES, N. 2011. Novel insights into the diversity of catabolic metabolism from ten haloarchaeal genomes. *PLoS One,* 6**,** e20237.

BABSKI, J., MAIER, L. K., HEYER, R., JASCHINSKI, K., PRASSE, D., JAGER, D., RANDAU, L., SCHMITZ, R. A., MARCHFELDER, A. & SOPPA, J. 2014. Small regulatory RNAs in Archaea. *RNA Biol,* 11**,** 484-93.

BALE, N. J., SOROKIN, D. Y., HOPMANS, E. C., KOENEN, M., RIJPSTRA, W. I. C., VILLANUEVA, L., WIENK, H. & SINNINGHE DAMSTE, J. S. 2019. New

insights into the polar lipid composition of extremely halo(alkali)philic euryarchaea from hypersaline lakes. *Front Microbiol,* 10**,** 377.

BALI, S., LAWRENCE, A. D., LOBO, S. A., SARAIVA, L. M., GOLDING, B. T., PALMER, D. J., HOWARD, M. J., FERGUSON, S. J. & WARREN, M. J. 2011. Molecular hijacking of siroheme for the synthesis of heme and d1 heme. *Proc Natl Acad Sci U S A,* 108**,** 18260-5.

BASHIRI, G., ANTONEY, J., JIRGIS, E. N. M., SHAH, M. V., NEY, B., COPP, J., STUTELEY, S. M., SREEBHAVAN, S., PALMER, B., MIDDLEDITCH, M., TOKURIKI, N., GREENING, C., SCOTT, C., BAKER, E. N. & JACKSON, C. J. 2019. A revised biosynthetic pathway for the cofactor F420 in prokaryotes. *Nat Commun,* 10**,** 1558.

BRAGA, D., LAST, D., HASAN, M., GUO, H., LEICHNITZ, D., UZUM, Z., RICHTER, I., SCHALK, F., BEEMELMANNS, C., HERTWECK, C. & LACKNER, G. 2019. Metabolic pathway rerouting in *Paraburkholderia rhizoxinica* evolved long-overlooked derivatives of coenzyme F420. *ACS Chem Biol,* 14**,** 2088-94.

BRASEN, C. & SCHONHEIT, P. 2001. Mechanisms of acetate formation and acetate activation in halophilic archaea. *Arch Microbiol,* 175**,** 360-8.

BRAUN, F., THOMALLA, L., VAN DER DOES, C., QUAX, T. E. F., ALLERS, T., KAEVER, V. & ALBERS, S. V. 2019. Cyclic nucleotides in archaea: Cyclic di-AMP in the archaeon *Haloferax volcanii* and its putative role. *Microbiologyopen,* 8**,** e00829.

BRAUN, M., BUNGERT, S. & FRIEDRICH, T. 1998. Characterization of the overproduced NADH dehydrogenase fragment of the NADH:ubiquinone oxidoreductase (complex I) from *Escherichia coli*. *Biochemistry,* 37**,** 1861-7.

CAFORIO, A. & DRIESSEN, A. J. M. 2017. Archaeal phospholipids: Structural properties and biosynthesis. *Biochim Biophys Acta Mol Cell Biol Lipids,* 1862**,** 1325-39.

CAO, S., HEPOWIT, N. & MAUPIN-FURLOW, J. A. 2015. Ubiquitin-like protein SAMP1 and JAMM/MPN+ metalloprotease HvJAMM1 constitute a system for reversible regulation of metabolic enzyme activity in Archaea. *PLoS One,* 10**,** e0128399.

CERLETTI, M., MARTINEZ, M. J., GIMENEZ, M. I., SASTRE, D. E., PAGGI, R. A. & DE CASTRO, R. E. 2014. The LonB protease controls membrane lipids composition and is essential for viability in the extremophilic haloarchaeon *Haloferax volcanii*. *Environ Microbiol,* 16**,** 1779-92.

CERLETTI, M., PAGGI, R., TROETSCHEL, C., FERRARI, M. C., GUEVARA, C. R., ALBAUM, S., POETSCH, A. & DE CASTRO, R. 2018. LonB protease is a

novel regulator of carotenogenesis controlling degradation of phytoene synthase in *Haloferax volcanii*. *J Proteome Res,* 17**,** 1158-71.

CHAGA, G., PORATH, J. & ILLENI, T. 1993. Isolation and purification of amyloglucosidase from *Halobacterium sodomense*. *Biomed Chromatogr,* 7**,** 256-61.

CHAI, Y., KOLTER, R. & LOSICK, R. 2009. A widely conserved gene cluster required for lactate utilization in *Bacillus subtilis* and its involvement in biofilm formation. *J Bacteriol,* 191**,** 2423-30.

COLLINS, M., AFOLAYAN, S., IGIRANEZA, A. B., SCHILLER, H., KRESPAN, E., BEITING, D. P., DYALL-SMITH, M., PFEIFFER, F. & POHLSCHRODER, M. 2020. Mutations affecting HVO_1357 or HVO_2248 cause hypermotility in *Haloferax volcanii*, suggesting roles in motility regulation. *Genes (Basel),* 12**,** 58.

COMMICHAU, F. M., HEIDEMANN, J. L., FICNER, R. & STULKE, J. 2019. Making and breaking of an essential poison: the cyclases and phosphodiesterases that produce and degrade the essential second messenger cyclic di-AMP in bacteria. *J Bacteriol,* 201**,** e00462-18.

CORRIGAN, R. M. & GRUNDLING, A. 2013. Cyclic di-AMP: another second messenger enters the fray. *Nat Rev Microbiol,* 11**,** 513-24.

COSTA, M. I., CERLETTI, M., PAGGI, R. A., TROTSCHEL, C., DE CASTRO, R. E., POETSCH, A. & GIMENEZ, M. I. 2018. *Haloferax volcanii* proteome response to deletion of a rhomboid protease gene. *J Proteome Res,* 17**,** 961-77.

CROFTS, A. R., HONG, S., WILSON, C., BURTON, R., VICTORIA, D., HARRISON, C. & SCHULTEN, K. 2013. The mechanism of ubihydroquinone oxidation at the Qo-site of the cytochrome bc1 complex. *Biochim Biophys Acta,* 1827**,** 1362-77.

DANCHIN, A., OUZOUNIS, C., TOKUYASU, T. & ZUCKER, J. D. 2018. No wisdom in the crowd: genome annotation in the era of big data - current status and future prospects. *Microb Biotechnol,* 11**,** 588-605.

DAWSON, K. S., FREEMAN, K. H. & MACALADY, J. L. 2012. Molecular characterization of core lipids from halophilic archaea grown under different salinity conditions. *Organic Geochemistry,* 48**,** 1-8.

DE CRECY-LAGARD, V. 2014. Variations in metabolic pathways create challenges for automated metabolic reconstructions: Examples from the tetrahydrofolate synthesis pathway. *Comput Struct Biotechnol J,* 10**,** 41-50.

DE CRECY-LAGARD, V., FOROUHAR, F., BROCHIER-ARMANET, C., TONG, L. & HUNT, J. F. 2012. Comparative genomic analysis of the DUF71/COG2102

family predicts roles in diphthamide biosynthesis and B12 salvage. *Biol Direct,* 7**,** 32.

DE, R. M. & GAMBACORTA, A. 1988. The lipids of archaebacteria. *Prog Lipid Res,* 27**,** 153-75.

DE SILVA, R. T., ABDUL-HALIM, M. F., PITTRICH, D. A., BROWN, H. J., POHLSCHRODER, M. & DUGGIN, I. G. 2021. Improved growth and morphological plasticity of *Haloferax volcanii*. *Microbiology (Reading),* 167.

DE WIT, L. E. A. & EKER, A. P. M. 1987. 8-Hydroxy-5-deazaflavin-dependent electron transfer in the extreme halophile *Halobacterium cutirubrum*. *FEMS Microbiology Letters,* 48**,** 121-25.

DESMARAIS, J. J., FLAMHOLZ, A. I., BLIKSTAD, C., DUGAN, E. J., LAUGHLIN, T. G., OLTROGGE, L. M., CHEN, A. W., WETMORE, K., DIAMOND, S., WANG, J. Y. & SAVAGE, D. F. 2019. DABs are inorganic carbon pumps found throughout prokaryotic phyla. *Nat Microbiol,* 4**,** 2204-15.

DUGGIN, I. G., AYLETT, C. H., WALSH, J. C., MICHIE, K. A., WANG, Q., TURNBULL, L., DAWSON, E. M., HARRY, E. J., WHITCHURCH, C. B., AMOS, L. A. & LOWE, J. 2015. CetZ tubulin-like proteins control archaeal cell shape. *Nature,* 519**,** 362-5.

DUMMER, A. M., BONSALL, J. C., CIHLA, J. B., LAWRY, S. M., JOHNSON, G. C. & PECK, R. F. 2011. Bacterioopsin-mediated regulation of bacterioruberin biosynthesis in *Halobacterium salinarum*. *J Bacteriol,* 193**,** 5658-67.

ELLING, F. J., BECKER, K. W., KONNEKE, M., SCHRODER, J. M., KELLERMANN, M. Y., THOMM, M. & HINRICHS, K. U. 2016. Respiratory quinones in Archaea: phylogenetic distribution and application as biomarkers in the marine environment. *Environ Microbiol,* 18**,** 692-707.

FALB, M., MULLER, K., KONIGSMAIER, L., OBERWINKLER, T., HORN, P., VON GRONAU, S., GONZALEZ, O., PFEIFFER, F., BORNBERG-BAUER, E. & OESTERHELT, D. 2008. Metabolism of halophilic archaea. *Extremophiles,* 12**,** 177-96.

FALB, M., PFEIFFER, F., PALM, P., RODEWALD, K., HICKMANN, V., TITTOR, J. & OESTERHELT, D. 2005. Living with two extremes: conclusions from the genome sequence of *Natronomonas pharaonis. Genome Res,* 15**,** 1336-43.

FISCHER, S., JOHN VON FREYEND, S., SABAG-DAIGLE, A., DANIELS, C. J., ALLERS, T. & MARCHFELDER, A. 2012. Assigning a function to a conserved archaeal metallo-beta-lactamase from *Haloferax volcanii*. *Extremophiles,* 16**,** 333-43.

GAO, C., WANG, Y., ZHANG, Y., LV, M., DOU, P., XU, P. & MA, C. 2015. NAD-independent L-lactate dehydrogenase required for L-lactate utilization in *Pseudomonas stutzeri* A1501. *J Bacteriol,* 197**,** 2239-47.

GIANI, M., MIRALLES-ROBLEDILLO, J. M., PEIRO, G., PIRE, C. & MARTINEZ-ESPINOSA, R. M. 2020. Deciphering pathways for carotenogenesis in haloarchaea. *Molecules,* 25**,** 1197.

GRADIN, C. H., HEDERSTEDT, L. & BALTSCHEFFSKY, H. 1985. Soluble succinate dehydrogenase from the halophilic archaebacterium, *Halobacterium halobium*. *Arch Biochem Biophys,* 239**,** 200-5.

GROCHOWSKI, L. L., XU, H. & WHITE, R. H. 2008. Identification and characterization of the 2-phospho-L-lactate guanylyltransferase involved in coenzyme F420 biosynthesis. *Biochemistry,* 47**,** 3033-7.

GULKO, M. K., DYALL-SMITH, M., GONZALEZ, O. & OESTERHELT, D. 2014. How do haloarchaea synthesize aromatic amino acids? *PLoS One,* 9**,** e107475.

GUNDLACH, J., MEHNE, F. M., HERZBERG, C., KAMPF, J., VALERIUS, O., KAEVER, V. & STULKE, J. 2015. An essential poison: synthesis and degradation of cyclic di-AMP in *Bacillus subtilis*. *J Bacteriol,* 197**,** 3265-74.

GUO, R., GU, J., ZONG, S., WU, M. & YANG, M. 2018. Structure and mechanism of mitochondrial electron transport chain. *Biomed J,* 41**,** 9-20.

HAQUE, R. U., PARADISI, F. & ALLERS, T. 2020. *Haloferax volcanii* for biotechnology applications: challenges, current state and perspectives. *Appl Microbiol Biotechnol,* 104**,** 1371-82.

HARTMAN, A. L., NORAIS, C., BADGER, J. H., DELMAS, S., HALDENBY, S., MADUPU, R., ROBINSON, J., KHOURI, H., REN, Q., LOWE, T. M., MAUPIN-FURLOW, J., POHLSCHRODER, M., DANIELS, C., PFEIFFER, F., ALLERS, T. & EISEN, J. A. 2010. The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS One,* 5**,** e9605.

HATTORI, T., SHIBA, H., ASHIKI, K., ARAKI, T., NAGASHIMA, Y. K., YOSHIMATSU, K. & FUJIWARA, T. 2016. Anaerobic growth of haloarchaeon *Haloferax volcanii* by denitrification is controlled by the transcription regulator NarO. *J Bacteriol,* 198**,** 1077-86.

HE, J., YIN, W., GALPERIN, M. Y. & CHOU, S. H. 2020. Cyclic di-AMP, a second messenger of primary importance: tertiary structures and binding mechanisms. *Nucleic Acids Res,* 48**,** 2807-29.

HEYER, R., DORR, M., JELLEN-RITTER, A., SPATH, B., BABSKI, J., JASCHINSKI, K., SOPPA, J. & MARCHFELDER, A. 2012. High throughput sequencing

reveals a plethora of small RNAs including tRNA derived fragments in *Haloferax volcanii*. *RNA biology,* 9.

HILDEBRANDT, P., MATYSIK, J., SCHRADER, B., SCHARF, B. & ENGELHARD, M. 1994. Raman spectroscopic study of the blue copper protein halocyanin from *Natronobacterium pharaonis*. *Biochemistry,* 33**,** 11426-31.

HOCHULI, M., PATZELT, H., OESTERHELT, D., WUTHRICH, K. & SZYPERSKI, T. 1999. Amino acid biosynthesis in the halophilic archaeon *Haloarcula hispanica*. *J Bacteriol,* 181**,** 3226-37.

HORIE, A., TOMITA, T., SAIKI, A., KONO, H., TAKA, H., MINEKI, R., FUJIMURA, T., NISHIYAMA, C., KUZUYAMA, T. & NISHIYAMA, M. 2009. Discovery of proteinaceous N-modification in lysine biosynthesis of *Thermus thermophilus*. *Nat Chem Biol,* 5**,** 673-9.

HUNTER, S., APWEILER, R., ATTWOOD, T. K., BAIROCH, A., BATEMAN, A., BINNS, D., BORK, P., DAS, U., DAUGHERTY, L., DUQUENNE, L., FINN, R. D., GOUGH, J., HAFT, D., HULO, N., KAHN, D., KELLY, E., LAUGRAUD, A., LETUNIC, I., LONSDALE, D., LOPEZ, R., MADERA, M., MASLEN, J., MCANULLA, C., MCDOWALL, J., MISTRY, J., MITCHELL, A., MULDER, N., NATALE, D., ORENGO, C., QUINN, A. F., SELENGUT, J. D., SIGRIST, C. J., THIMMA, M., THOMAS, P. D., VALENTIN, F., WILSON, D., WU, C. H. & YEATS, C. 2009. InterPro: the integrative protein signature database. *Nucleic Acids Res,* 37**,** D211-5.

HWANG, S., CORDOVA, B., ABDO, M., PFEIFFER, F. & MAUPIN-FURLOW, J. A. 2017. ThiN as a versatile domain of transcriptional repressors and catalytic enzymes of thiamine biosynthesis. *J Bacteriol,* 199.

JOHNSEN, U., DAMBECK, M., ZAISS, H., FUHRER, T., SOPPA, J., SAUER, U. & SCHONHEIT, P. 2009. D-xylose degradation pathway in the halophilic archaeon *Haloferax volcanii*. *J Biol Chem,* 284**,** 27290-303.

JOHNSEN, U., SUTTER, J. M., SCHULZ, A. C., TASTENSEN, J. B. & SCHONHEIT, P. 2015. XacR - a novel transcriptional regulator of D-xylose and L-arabinose catabolism in the haloarchaeon *Haloferax volcanii*. *Environ Microbiol,* 17**,** 1663-76.

JOHNSON, M., ZARETSKAYA, I., RAYTSELIS, Y., MEREZHUK, Y., MCGINNIS, S. & MADDEN, T. L. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res,* 36**,** W5-9.

KAILA, V. R. I. & WIKSTROM, M. 2021. Architecture of bacterial respiratory chains. *Nat Rev Microbiol,* 19**,** 319-30.

KAMINSKI, L. & EICHLER, J. 2014. *Haloferax volcanii* N-glycosylation: delineating the pathway of dTDP-rhamnose biosynthesis. *PLoS One,* 9**,** e97441.

KANDIBA, L., LIN, C. W., AEBI, M., EICHLER, J. & GUERARDEL, Y. 2016. Structural characterization of the N-linked pentasaccharide decorating glycoproteins of the halophilic archaeon Haloferax volcanii. *Glycobiology,* 26**,** 745-56.

KANEHISA, M., SATO, Y., FURUMICHI, M., MORISHIMA, K. & TANABE, M. 2019. New approach for understanding genome variations in KEGG. *Nucleic Acids Res,* 47**,** D590-D95.

KATES, M. 1993. Biology of halophilic bacteria, Part II. Membrane lipids of extreme halophiles: biosynthesis, function and evolutionary significance. *Experientia,* 49**,** 1027-36.

KATES, M., MOLDOVEANU, N. & STEWART, L. C. 1993. On the revised structure of the major phospholipid of *Halobacterium salinarium*. *Biochim Biophys Acta,* 1169**,** 46-53.

KELLERMANN, M. Y., YOSHINAGA, M. Y., VALENTINE, R. C., WORMER, L. & VALENTINE, D. L. 2016. Important roles for membrane lipids in haloarchaeal bioenergetics. *Biochim Biophys Acta,* 1858**,** 2940-56.

KERSCHER, L. & OESTERHELT, D. 1977. Ferredoxin is the coenzyme of α-ketoacid oxidoreductases in *Halobacterium halobium*. *FEBS Letters,* 83**,** 197-201.

KERSCHER, L. & OESTERHELT, D. 1981a. The catalytic mechanism of 2-oxoacid:ferredoxin oxidoreductases from *Halobacterium halobium*. One-electron transfer at two distinct steps of the catalytic cycle. *Eur J Biochem,* 116**,** 595-600.

KERSCHER, L. & OESTERHELT, D. 1981b. Purification and properties of two 2-oxoacid:ferredoxin oxidoreductases from *Halobacterium halobium*. *Eur J Biochem,* 116**,** 587-94.

KILJUNEN, S., PAJUNEN, M. I., DILKS, K., STORF, S., POHLSCHRODER, M. & SAVILAHTI, H. 2014. Generation of comprehensive transposon insertion mutant library for the model archaeon, *Haloferax volcanii*, and its use for gene discovery. *BMC Biol,* 12**,** 103.

KLETZIN, A., HEIMERL, T., FLECHSLER, J., VAN NIFTRIK, L., RACHEL, R. & KLINGL, A. 2015. Cytochromes c in Archaea: distribution, maturation, cell architecture, and the special case of *Ignicoccus hospitalis*. *Front Microbiol,* 6**,** 439.

KRIVENTSEVA, E. V., KUZNETSOV, D., TEGENFELDT, F., MANNI, M., DIAS, R., SIMAO, F. A. & ZDOBNOV, E. M. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res,* 47**,** D807-D11.

KUBATOVA, N., JONKER, H. R. A., SAXENA, K., RICHTER, C., VOGEL, V., SCHREIBER, S., MARCHFELDER, A. & SCHWALBE, H. 2020. Solution Structure and Dynamics of the Small Protein HVO_2922 from *Haloferax volcanii*. *Chembiochem,* 21**,** 149-56.

KUPRAT, T., JOHNSEN, U., ORTJOHANN, M. & SCHONHEIT, P. 2020. Acetate metabolism in Archaea: characterization of an acetate transporter and of enzymes involved in acetate activation and gluconeogenesis in *Haloferax volcanii*. *Front Microbiol,* 11**,** 604926.

KUPRAT, T., ORTJOHANN, M., JOHNSEN, U. & SCHONHEIT, P. 2021. Glucose metabolism and acetate switch in Archaea: the enzymes in *Haloferax volcanii*. *J Bacteriol,* 203**,** e00690-20.

KUSHWAHA, S. C., KRAMER, J. K. G. & KATES, M. 1975. Isolation and characterization of C50-carotenoid pigments and other polar isoprenoids from *Halobacterium cutirubrum*. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism,* 398**,** 303-14.

LARGE, A., STAMME, C., LANGE, C., DUAN, Z., ALLERS, T., SOPPA, J. & LUND, P. A. 2007. Characterization of a tightly controlled promoter of the halophilic archaeon *Haloferax volcanii* and its use in the analysis of the essential *cct1* gene. *Mol Microbiol,* 66**,** 1092-106.

LEFFERS, H., GROPP, F., LOTTSPEICH, F., ZILLIG, W. & GARRETT, R. A. 1989. Sequence, organization, transcription and evolution of RNA polymerase subunit genes from the archaebacterial extreme halophiles *Halobacterium halobium* and *Halococcus morrhuae*. *J Mol Biol,* 206**,** 1-17.

LEIF, H., SLED, V. D., OHNISHI, T., WEISS, H. & FRIEDRICH, T. 1995. Isolation and characterization of the proton-translocating NADH: ubiquinone oxidoreductase from *Escherichia coli*. *Eur J Biochem,* 230**,** 538-48.

LEIGH, J. A., ALBERS, S. V., ATOMI, H. & ALLERS, T. 2011. Model organisms for genetics in the domain Archaea: methanogens, halophiles, *Thermococcales* and *Sulfolobales*. *FEMS Microbiol Rev,* 35**,** 577-608.

LI, Z., RODRIGUEZ-FRANCO, M., ALBERS, S. V. & QUAX, T. E. F. 2020. The switch complex ArlCDE connects the chemotaxis system and the archaellum. *Mol Microbiol,* 114**,** 468-79.

LIAO, Y., ITHURBIDE, S., EVENHUIS, C., LOWE, J. & DUGGIN, I. G. 2021. Cell division in the archaeon *Haloferax volcanii* relies on two FtsZ proteins with distinct functions in division ring assembly and constriction. *Nat Microbiol,* 6**,** 594-605.

LIN, X. L. & WHITE, R. H. 1986. Occurrence of coenzyme F420 and its gamma-monoglutamyl derivative in nonmethanogenic archaebacteria. *J Bacteriol,* 168**,** 444-8.

MADEN, B. E. 2000. Tetrahydrofolate and tetrahydromethanopterin compared: functionally distinct carriers in C1 metabolism. *Biochem J,* 350 Pt 3**,** 609-29.

MADON, J. & ZILLIG, W. 1983. A form of the DNA-dependent RNA polymerase of *Halobacterium halobium*, containing an additional component, is able to transcribe native DNA. *Eur J Biochem,* 133**,** 471-4.

MAIER, L. K., STACHLER, A. E., BRENDEL, J., STOLL, B., FISCHER, S., HAAS, K. A., SCHWARZ, T. S., ALKHNBASHI, O. S., SHARMA, K., URLAUB, H., BACKOFEN, R., GOPHNA, U. & MARCHFELDER, A. 2019. The nuts and bolts of the Haloferax CRISPR-Cas system I-B. *RNA Biol,* 16**,** 469-80.

MATTAR, S. 1996. *Molekularbiologische und Biochemische Charakterisierung zweier Komplexe der Atmungskette von Natronobacterium pharaonis.* Ph.D., Ruhr-Universität Bochum.

MATTAR, S., SCHARF, B., KENT, S. B., RODEWALD, K., OESTERHELT, D. & ENGELHARD, M. 1994. The primary structure of halocyanin, an archaeal blue copper protein, predicts a lipid anchor for membrane fixation. *Journal of Biological Chemistry,* 269**,** 14939-45.

MOORE, S. J., LAWRENCE, A. D., BIEDENDIECK, R., DEERY, E., FRANK, S., HOWARD, M. J., RIGBY, S. E. & WARREN, M. J. 2013. Elucidation of the anaerobic pathway for the corrin component of cobalamin (vitamin B12). *Proc Natl Acad Sci U S A,* 110**,** 14906-11.

NAGEL, C., MACHULLA, A., ZAHN, S. & SOPPA, J. 2019. Several one-domain zinc finger micro-proteins of *Haloferax volcanii* are important for stress adaptation, biofilm formation, and swarming. *Genes (Basel),* 10**,** 361.

NANBA, T. & MUKOHATA, Y. 1987. A membrane-bound ATPase from *Halobacterium halobium*: purification and characterization. *J Biochem (Tokyo),* 102**,** 591-8.

NUSSBAUM, P., ITHURBIDE, S., WALSH, J. C., PATRO, M., DELPECH, F., RODRIGUEZ-FRANCO, M., CURMI, P. M. G., DUGGIN, I. G., QUAX, T. E. F. & ALBERS, S. V. 2020. An oscillating MinD protein determines the cellular positioning of the motility machinery in Archaea. *Curr Biol,* 30**,** 4956-72 e4.

OBERTO, J. 2013. SyntTax: a web server linking synteny to prokaryotic taxonomy. *BMC Bioinformatics,* 14**,** 4.

OREN, A. 2002. Molecular ecology of extremely halophilic Archaea and Bacteria. *FEMS Microbiol. Ecol.,* 39**,** 1-7.

OREN, A., VENTOSA, A. & GRANT, W. D. 1997. Proposed minimal standards for description of new taxa in the order *Halobacteriales*. *International Journal of Systematic Bacteriology,* 47**,** 233-38.

OUCHI, T., TOMITA, T., HORIE, A., YOSHIDA, A., TAKAHASHI, K., NISHIDA, H., LASSAK, K., TAKA, H., MINEKI, R., FUJIMURA, T., KOSONO, S., NISHIYAMA, C., MASUI, R., KURAMITSU, S., ALBERS, S. V., KUZUYAMA, T. & NISHIYAMA, M. 2013. Lysine and arginine biosyntheses mediated by a common carrier protein in *Sulfolobus*. *Nat Chem Biol,* 9**,** 277-83.

PEREZ-ARNAIZ, P., DATTANI, A., SMITH, V. & ALLERS, T. 2020. *Haloferax volcanii*-a model archaeon for studying DNA replication and repair. *Open Biol,* 10**,** 200293.

PFEIFFER, F., BROICHER, A., GILLICH, T., KLEE, K., MEJIA, J., RAMPP, M. & OESTERHELT, D. 2008a. Genome information management and integrated data analysis with HaloLex. *Arch Microbiol,* 190**,** 281-99.

PFEIFFER, F., LOSENSKY, G., MARCHFELDER, A., HABERMANN, B. & DYALL-SMITH, M. 2020. Whole-genome comparison between the type strain of *Halobacterium salinarum* (DSM 3754(T)) and the laboratory strains R1 and NRC-1. *Microbiologyopen,* 9**,** e974.

PFEIFFER, F. & OESTERHELT, D. 2015. A manual curation strategy to improve genome annotation: application to a set of haloarchael genomes. *Life (Basel),* 5**,** 1427-44.

PFEIFFER, F., SCHUSTER, S. C., BROICHER, A., FALB, M., PALM, P., RODEWALD, K., RUEPP, A., SOPPA, J., TITTOR, J. & OESTERHELT, D. 2008b. Evolution in the laboratory: the genome of *Halobacterium salinarum* strain R1 compared to that of strain NRC-1. *Genomics,* 91**,** 335-46.

PICKL, A., JOHNSEN, U. & SCHONHEIT, P. 2012. Fructose degradation in the haloarchaeon *Haloferax volcanii* involves a bacterial type phosphoenolpyruvate-dependent phosphotransferase system, fructose-1-phosphate kinase, and class II fructose-1,6-bisphosphate aldolase. *Journal of bacteriology,* 194**,** 3088-97.

PLAGA, W., LOTTSPEICH, F. & OESTERHELT, D. 1992. Improved purification, crystallization and primary structure of pyruvate:ferredoxin oxidoreductase from *Halobacterium halobium*. *Eur J Biochem,* 205**,** 391-7.

POHLSCHRODER, M. & ESQUIVEL, R. N. 2015. Archaeal type IV pili and their involvement in biofilm formation. *Front Microbiol,* 6**,** 190.

PORAT, I., SIEPRAWSKA-LUPA, M., TENG, Q., BOHANON, F. J., WHITE, R. H. & WHITMAN, W. B. 2006. Biochemical and genetic characterization of an early step in a novel pathway for the biosynthesis of aromatic amino acids and p-aminobenzoic acid in the archaeon *Methanococcus maripaludis*. *Mol Microbiol,* 62**,** 1117-31.

PROMPONAS, V. J., ILIOPOULOS, I. & OUZOUNIS, C. A. 2015. Annotation inconsistencies beyond sequence similarity-based function prediction - phylogeny and genome structure. *Stand Genomic Sci,* 10**,** 108.

PRUNETTI, L., REUTER, C. J., HEPOWIT, N. L., WU, Y., BARRUETO, L., MIRANDA, H. V., KELLY, K. & MAUPIN-FURLOW, J. A. 2014. Structural and biochemical properties of an extreme 'salt-loving' proteasome activating nucleotidase from the archaeon *Haloferax volcanii*. *Extremophiles,* 18**,** 283-93.

QI, Q., ITO, Y., YOSHIMATSU, K. & FUJIWARA, T. 2016. Transcriptional regulation of dimethyl sulfoxide respiration in a haloarchaeon, *Haloferax volcanii*. *Extremophiles,* 20**,** 27-36.

QUAX, T. E. F., ALTEGOER, F., ROSSI, F., LI, Z., RODRIGUEZ-FRANCO, M., KRAUS, F., BANGE, G. & ALBERS, S. V. 2018. Structure and function of the archaeal response regulator CheY. *Proc Natl Acad Sci U S A,* 115**,** E1259-E68.

RANGASWAMY, V. & ALTEKAR, W. 1994. Ketohexokinase (ATP:D-fructose 1-phosphotransferase) from a halophilic archaebacterium, *Haloarcula vallismortis*: purification and properties. *J Bacteriol,* 176**,** 5505-12.

RAWLS, K. S., YACOVONE, S. K. & MAUPIN-FURLOW, J. A. 2010. GlpR represses fructose and glucose metabolic enzymes at the level of transcription in the haloarchaeon *Haloferax volcanii*. *J Bacteriol,* 192**,** 6251-60.

REINHARDT, A., JOHNSEN, U. & SCHONHEIT, P. 2019. L-Rhamnose catabolism in archaea. *Mol Microbiol,* 111**,** 1093-108.

REUTER, C. J. & MAUPIN-FURLOW, J. A. 2004. Analysis of proteasome-dependent proteolysis in *Haloferax volcanii* cells, using short-lived green fluorescent proteins. *Appl. Environ. Microbiol.,* 70**,** 7530-38.

REUTER, C. J., UTHANDI, S., PUENTES, J. A. & MAUPIN-FURLOW, J. A. 2010. Hydrophobic carboxy-terminal residues dramatically reduce protein levels in the haloarchaeon *Haloferax volcanii*. *Microbiology (Reading),* 156**,** 248-55.

RICH, P. R. & MARECHAL, A. 2010. The mitochondrial respiratory chain. *Essays Biochem,* 47**,** 1-23.

RODIONOV, D. A., VITRESCHAK, A. G., MIRONOV, A. A. & GELFAND, M. S. 2003. Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. *J Biol Chem,* 278**,** 41148-59.

RODIONOVA, I. A., VETTING, M. W., LI, X., ALMO, S. C., OSTERMAN, A. L. & RODIONOV, D. A. 2017. A novel bifunctional transcriptional regulator of riboflavin metabolism in Archaea. *Nucleic Acids Res,* 45**,** 3785-99.

RUDOLPH, B., HANSEN, T. & SCHONHEIT, P. 2004. Glucose-6-phosphate isomerase from the hyperthermophilic archaeon *Methanococcus jannaschii*: characterization of the first archaeal member of the phosphoglucose isomerase superfamily. *Arch Microbiol,* 181**,** 82-7.

SCHAFER, G., ENGELHARD, M. & MULLER, V. 1999. Bioenergetics of the Archaea. *Microbiol Mol Biol Rev,* 63**,** 570-620.

SCHARF, B. & ENGELHARD, M. 1993. Halocyanin, an archaebacterial blue copper protein (type I) from *Natronobacterium pharaonis*. *Biochemistry,* 32**,** 12894-900.

SCHARF, B., WITTENBERG, R. & ENGELHARD, M. 1997. Electron transfer proteins from the haloalkaliphilic archaeon *Natronobacterium pharaonis*: possible components of the respiratory chain include cytochrome bc and a terminal oxidase cytochrome ba3. *Biochemistry,* 36**,** 4471-9.

SCHILLER, H., SCHULZE, S., MUTAN, Z., DE VAULX, C., RUNCIE, C., SCHWARTZ, J., RADOS, T., BISSON FILHO, A. W. & POHLSCHRODER, M. 2020. *Haloferax volcanii* immersed liquid biofilms develop independently of known biofilm machineries and exhibit rapid honeycomb pattern formation. *mSphere,* 5.

SCHNOES, A. M., BROWN, S. D., DODEVSKI, I. & BABBITT, P. C. 2009. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput Biol,* 5**,** e1000605.

SCHULZE, S., ADAMS, Z., CERLETTI, M., DE CASTRO, R., FERREIRA-CERCA, S., FUFEZAN, C., GIMENEZ, M. I., HIPPLER, M., JEVTIC, Z., KNUPPEL, R., LEGERME, G., LENZ, C., MARCHFELDER, A., MAUPIN-FURLOW, J., PAGGI, R. A., PFEIFFER, F., POETSCH, A., URLAUB, H. & POHLSCHRODER, M. 2020. The Archaeal Proteome Project advances knowledge about archaeal cell biology through comprehensive proteomics. *Nat Commun,* 11**,** 3145.

SCHULZE, S., PFEIFFER, F., GARCIA, B. A. & POHLSCHRODER, M. 2021. Glycoproteomics of *Haloferax volcanii* reveals an extensive glycoproteome and concurrence of different N-glycosylation pathways. biorxiv.org: biorxiv.org.

SELENGUT, J. D. & HAFT, D. H. 2010. Unexpected abundance of coenzyme F(420)-dependent enzymes in *Mycobacterium tuberculosis* and other actinobacteria. *J Bacteriol,* 192**,** 5788-98.

SHALEV, Y., SOUCY, S. M., PAPKE, R. T., GOGARTEN, J. P., EICHLER, J. & GOPHNA, U. 2018. Comparative analysis of surface layer glycoproteins and genes involved in protein glycosylation in the genus *Haloferax*. *Genes (Basel),* 9.

SHALEV, Y., TURGEMAN-GROTT, I., TAMIR, A., EICHLER, J. & GOPHNA, U. 2017. Cell surface glycosylation is required for efficient mating of *Haloferax volcanii*. *Front Microbiol,* 8**,** 1253.

SIDDARAMAPPA, S., CHALLACOMBE, J. F., DECASTRO, R. E., PFEIFFER, F., SASTRE, D. E., GIMENEZ, M. I., PAGGI, R. A., DETTER, J. C., DAVENPORT, K. W., GOODWIN, L. A., KYRPIDES, N., TAPIA, R., PITLUCK, S., LUCAS, S., WOYKE, T. & MAUPIN-FURLOW, J. A. 2012. A comparative genomics perspective on the genetic content of the alkaliphilic haloarchaeon *Natrialba magadii* ATCC 43099^{T}. *BMC Genomics,* 13**,** 165.

SOPPA, J. 2011. Functional genomic and advanced genetic studies reveal novel insights into the metabolism, regulation, and biology of *Haloferax volcanii*. *Archaea,* 2011**,** 602408.

SREERAMULU, K. 2003. Purification and partial characterization of cytochrome c552 from *Halobacterium salinarium*. *Indian Journal of Biochemistry and Biophysics,* 40**,** 274-77.

SREERAMULU, K., SCHMIDT, C. L., SCHAFER, G. & ANEMULLER, S. 1998. Studies of the electron transport chain of the euryarcheon *Halobacterium salinarum*: indications for a type II NADH dehydrogenase and a complex III analog. *J Bioenerg Biomembr,* 30**,** 443-53.

STEINERT, K., WAGNER, V., KROTH PANCIC, P. G. & BICKEL SANDKOETTER, S. 1997. Characterization and subunit structure of the ATP synthase of the halophilic archaeon *Haloferax volcanii* and organization of the ATP synthase genes. *Journal of Biological Chemistry,* 272**,** 6261-69.

STORF, S., PFEIFFER, F., DILKS, K., CHEN, Z. Q., IMAM, S. & POHLSCHRODER, M. 2010. Mutational and bioinformatic analysis of haloarchaeal lipobox-containing proteins. *Archaea,* 2010**,** Article ID 410975.

STRAUB, J., BRENNEIS, M., JELLEN-RITTER, A., HEYER, R., SOPPA, J. & MARCHFELDER, A. 2009. Small RNAs in haloarchaea: Identification, differential expression and biological function. *RNA Biol,* 6.

SUTTER, J. M., JOHNSEN, U., REINHARDT, A. & SCHONHEIT, P. 2020. Pentose degradation in archaea: *Halorhabdus* species degrade D-xylose, L-arabinose and D-ribose via bacterial-type pathways. *Extremophiles,* 24**,** 759-72.

SUTTER, J. M., TASTENSEN, J. B., JOHNSEN, U., SOPPA, J. & SCHONHEIT, P. 2016. Key enzymes of the semiphosphorylative Entner-Doudoroff pathway in the haloarchaeon *Haloferax volcanii*: Characterization of glucose dehydrogenase, gluconate dehydratase and 2-keto-3-deoxy-6-phosphogluconate aldolase. *J Bacteriol,* 198**,** 2251-62.

TACHIBANA, A. 1994. A novel prenyltransferase, farnesylgeranyl diphosphate synthase, from the haloalkaliphilic archaeon, *Natronobacterium pharaonis*. *FEBS Letters,* 341**,** 291-94.

TÄSTENSEN, J. B., JOHNSEN, U., REINHARDT, A., ORTJOHANN, M. & SCHONHEIT, P. 2020. D-galactose catabolism in archaea: operation of the DeLey-Doudoroff pathway in *Haloferax volcanii*. *FEMS Microbiol Lett,* 367**,** fnaa029.

TITTES, C., SCHWARZER, S., PFEIFFER, F., DYALL-SMITH, M., RODRIGUEZ-FRANCO, M., OKSANEN, H. M. & QUAX, T. E. F. 2021. Cellular and genomic properties of *Haloferax gibbonsii* LR2-5, the host of euryarchaeal virus HFTV1. *Front Microbiol,* 12**,** 625599.

TRIPEPI, M., YOU, J., TEMEL, S., ONDER, O., BRISSON, D. & POHLSCHRODER, M. 2012. N-Glycosylation of *Haloferax volcanii* flagellins requires known Agl proteins and Is essential for biosynthesis of stable flagella. *Journal of bacteriology,* 194**,** 4876-87.

TURKOWYD, B., SCHREIBER, S., WORTZ, J., SEGAL, E. S., MEVARECH, M., DUGGIN, I. G., MARCHFELDER, A. & ENDESFELDER, U. 2020. Establishing live-cell single-molecule localization microscopy imaging and single-particle tracking in the archaeon *Haloferax volcanii*. *Front Microbiol,* 11**,** 583010.

UNIPROT, C. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res,* 49**,** D480-D89.

VANNICE, J. C., SKAFF, D. A., KEIGHTLEY, A., ADDO, J. K., WYCKOFF, G. J. & MIZIORKO, H. M. 2014. Identification in *Haloferax volcanii* of phosphomevalonate decarboxylase and isopentenyl phosphate kinase as

catalysts of the terminal enzyme reactions in an archaeal alternate mevalonate pathway. *J Bacteriol,* 196**,** 1055-63.

WALSH, J. C., ANGSTMANN, C. N., BISSON-FILHO, A. W., GARNER, E. C., DUGGIN, I. G. & CURMI, P. M. G. 2019. Division plane placement in pleomorphic archaea is dynamically coupled to cell shape. *Mol Microbiol*.

WHITE, R. H. 1988. Analysis and characterization of the folates in the nonmethanogenic archaebacteria. *J Bacteriol,* 170**,** 4608-12.

WHITE, R. H. 2004. L-Aspartate semialdehyde and a 6-deoxy-5-ketohexose 1-phosphate are the precursors to the aromatic amino acids in *Methanocaldococcus jannaschii*. *Biochemistry,* 43**,** 7618-27.

WILLIAMS, T. J., ALLEN, M. A., LIAO, Y., RAFTERY, M. J. & CAVICCHIOLI, R. 2019. Sucrose metabolism in haloarchaea: reassessment using genomics, proteomics, and metagenomics. *Appl Environ Microbiol,* 85**,** e02935-18.

WYSS, L., WASER, M., GEBETSBERGER, J., ZYWICKI, M. & POLACEK, N. 2018. mRNA-specific translation regulation by a ribosome-associated ncRNA in *Haloferax volcanii*. *Sci Rep,* 8**,** 12502.

YANG, Y., YATSUNAMI, R., ANDO, A., MIYOKO, N., FUKUI, T., TAKAICHI, S. & NAKAMURA, S. 2015. Complete biosynthetic pathway of the C50 carotenoid bacterioruberin from lycopene in the extremely halophilic archaeon *Haloarcula japonica*. *J Bacteriol,* 197**,** 1614-23.

YIN, W., CAI, X., MA, H., ZHU, L., ZHANG, Y., CHOU, S. H., GALPERIN, M. Y. & HE, J. 2020. A decade of research on the second messenger c-di-AMP. *FEMS Microbiol Rev,* 44**,** 701-24.

YOSHIDA, A., TOMITA, T., ATOMI, H., KUZUYAMA, T. & NISHIYAMA, M. 2016. Lysine biosynthesis of *Thermococcus kodakarensis* with the capacity to function as an ornithine biosynthetic system. *J Biol Chem,* 291**,** 21630-43.

ZAFRILLA, B., MARTINEZ-ESPINOSA, R. M., BONETE, M. J., BUTT, J. N., RICHARDSON, D. J. & GATES, A. J. 2011. A haloarchaeal ferredoxin electron donor that plays an essential role in nitrate assimilation. *Biochem Soc Trans,* 39**,** 1844-8.

ZAHN, S., KUBATOVA, N., PYPER, D. J., CASSIDY, L., SAXENA, K., THOLEY, A., SCHWALBE, H. & SOPPA, J. 2021. Biological functions, genetic and biochemical characterization, and NMR structure determination of the small zinc finger protein HVO_2753 from *Haloferax volcanii*. *FEBS J,* 288**,** 2042-62.

ZERULLA, K., CHIMILESKI, S., NATHER, D., GOPHNA, U., PAPKE, R. T. & SOPPA, J. 2014. DNA as a phosphate storage polymer and the alternative advantages of polyploidy for growth or survival. *PLoS One,* 9**,** e94819.

1285