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 systematically 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 that is probably incorrect. We refer to annotation problems in these databases solely to underscore that the biological issues raised by us are far from trivial. There is 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. We have actively supported them by providing feedback and encourage others to do the same, e.g. with the recently implemented "Add a publication" functionality in UniProt entries, that allows users to connect a protein to a publication that describes its experimental characterization (https://community.uniprot.org/bbsub/bbsubinfo.html).

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 that are currently under survey are listed in Suppl. Table S11.

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 conservation 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 that 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 this might be 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 contiguous 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 unambiguous 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 (Table 1). 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).

			Gold Standard Protein						
Section	**Code**	**Gene**	**isofunc**	**%seq_id**	**Locus tag**	**UniProt**	**Reference**	**PMID**	**Comment**
1a	HVO_1305 HVO_1304	*porAB*	yes	67% 80%	OE2623R OE2622R	B0R4X6 B0R4X5	(Plaga et al., 1992) (Kerscher and Oesterhelt, 1981b) (Kerscher and Oesterhelt, 1981a)	1555599 6266826 6266827	
1a	HVO_0888 HVO_0887	*korAB*	yes	77% 77%	OE1711R OE1710R	B0R3G0 B0R3F9	(Kerscher and Oesterhelt, 1981b) (Kerscher and Oesterhelt, 1981a)	6266826 6266827	
1a/1b	HVO_2995	*fdx*	yes	88%	OE4217R	B0R7I9	(Kerscher and Oesterhelt, 1976) (Kerscher et al., 1976) (Kerscher and Oesterhelt, 1977)	964365 188650 201489	role in oxidative decarboxylation
1a/1b	HVO_2995 (cont.)				self	D4GY89	(Zafrilla et al., 2011)	22103537	role in nitrate assimilation
1c	HVO_0979 (complex)	*nuoB*	possibly	50%	tlr0705	Q8DKZ4	(Zhang et al., 2005) (Schuller et al., 2019) (Pan et al., 2020)	15910282 30573545 32001694	reoxidizes ferredoxin
1c	HVO_0979 (cont.)		no	48%	b2287	P0AFC7	(Leif et al., 1995) (Braun et al., 1998)	7607227 9485311	reoxidizes NADH in *E.coli*
1d	NP_3508A	*ndh1*	special	26% (N-term 140 aa)	-	Q7ZAG8			function of Q7ZAG8 was reassigned (from ndh1 to sqr) after annotation transfer
1d	NP_3508A (cont.)		possibly	30%	BpOF4_04810	A7LKG4	(Liu et al., 2008)	18359284	type II NADH dehydrogenase
1e	HVO_2620 HVO_0842 HVO_0841	*petABD*	yes	39%	SYNPCC7002_A0842	P28056	(Lee et al., 2001)	11245788	HVO_0842 (*petB*) related to cytochrome b6
1f	HVO_2810	*sdhD*	yes	66%	NP_4268A	Q3INS7	(Scharf et al., 1997) (Mattar, 1996)	9109654 PhD_Mattar	
1g	HVO_0943	*cbaD*	yes	57%	NP_2966A	A0A1U7EWW4	(Mattar and Engelhard, 1997)	9428682	
	HVO_0943 (cont.)		-	63%	OE_4073R (C-term)	B0R7A9		-	halocyanin/cbaD fusion protein, uncharacterized

1g	HVO_2150	*hcpG*	-	44%	OE_4073R (N-term)	B0R7A9		-	halocyanin/cbaD fusion protein, uncharacterized
1h	HVO_0945 (complex)	*cbaA*	yes	64%	NP_2966A	A0A1U7EWW4	(Mattar and Engelhard, 1997)	9428682	
1h	HVO_0907 (complex)	*coxA1*			self		(Tanaka et al., 2002)	11790755	
1h	HVO_0907 (cont.)		yes	70%	VNG_0657G (OE_1979R)	P33588	(Fujiwara et al., 1989) (Denda et al., 1991)	2542239 1659810	
1h	HVO_1645 (complex)	*coxAC2*	yes	43%	APE_0793.1	Q9YdX6	(Ishikawa et al., 2002)	12471503	
1h	HVO_0462 HVO_0461	*cydAB*	yes	32% 24%	- -	Q09049 Q05780	(Moshiri et al., 1991)	1655703	
1h	HVO_0462 HVO_0461 (cont.)		yes	30% 27%	b0733 b0734	P0ABJ9 P0ABK2	(Miller and Gennis, 1983)	6307994	
1h	NP_4296A NP_4294A	*coxA3* *coxB3*	yes	28% 33%	TTHA1135 TTHA1134	Q5SJ79 Q5SJ80	(Zimmermann et al., 1988) (Keightley et al., 1995)	2842747 7657607	
1i	HVO_2958 HVO_2959	*oadhAB1*			self	D4GY15 D4GY17	(Sisignano et al., 2010)	19910413	Ile indirectly assigned as substrate
1i	HVO_2958 HVO_2959 (cont.)				self		(Jolley et al., 2000) (Al-Mailem et al., 2008) (van Ooyen and Soppa, 2007)	10832633 17571210 17906130	no substrate was identified;pyruvate and alphaKG excluded
1i	HVO_2595 HVO_2596	*oadhAB2*			self		(Wanner and Soppa, 2002) (van Ooyen and Soppa, 2007) (Sisignano et al., 2010)	12003954 17906130 19910413	no substrate was identified;pyruvate and alphaKG excluded
1i	HVO_0669 HVO_0668	*oadhAB3*			self		(van Ooyen and Soppa, 2007) (Sisignano et al., 2010)	17906130 19910413	no substrate was identified;pyruvate and alphaKG excluded
1i	HVO_2209	*oadhA4*			self				not yet analyzed experimentally
1i	HVO_2958 HVO_2959 (cont.)		yes/no	38% 52%	TA1438 TA1437	Q9HIA3 Q9HIA4	(Heath et al., 2007)	17894823	substrates are Ile, Leu, Val

1i	HVO_2595 HVO_2596 (cont.)		no	41% 41%	- -	Q57102 Q57041	(Oppermann et al., 1991)	1898934	substrate is acetoin
1i	HVO_2595 HVO_2596 (cont.)		unknown	40% 43%	BSU08060 BSU08070	O31404 O34591	(Huang et al., 1999)	10368162	substrate is acetoin
1i	HVO_0669 HVO_0668 (cont.)		unknown	54% 47%	BSU08060 BSU08070	O31404 O34591	(Huang et al., 1999)	10368162	substrate is acetoin
1i	HVO_0669 HVO_0668 (cont.)		unknown	49% 43%	- -	Q57102 Q57041	(Oppermann et al., 1991)	1898934	substrate is acetoin
1i	HVO_2209 (cont.)		unknown	38%	TA1438	Q9HIA3	(Heath et al., 2007)	17894823	substrates are Ile, Leu, Val

Table 1: **Proteins with open annotation issues and their Gold Standard Protein homologs (Section 1).** The column Section refers to the Table listing the protein and to the section in the Results and in Suppl. Text S1. As an example, 2c covers topic (c) from the decimal-numbered Results subsection 3.2 Amino Acid Biosynthesis. In Suppl. Text S1, this is covered under Section S2 subsection S2.c. The corresponding proteins are listed in Table 2. For a few proteins, two sections are indicated (e.g. 1a/1b). The column Code refers to a haloarchaeal protein by its locus tag, which is mainly from *Haloferax volcanii* (HVO), but also from *Halobacterium salinarum* (OE), *Natronomonas pharaonis* (NP) and *Halohasta litchfieldiae* (halTADL). When the reconstruction of a complete pathway is presented, the unassigned genes are indicated as a "pathway gap". In one case we indicate the absence of a haloarchaeal ortholog by a dash. In the case of a complex, we either list more than one code, or we list only one subunit together with the term (complex). All subunits of these complexes are listed groupwise in Table S10. A protein may be shown in more than one row. From the 2nd row onwards, this is indicated by the term (cont.). The column Gene lists the assigned gene or a dash if no gene has been assigned. The assigned gene is only indicated in the first row of a protein. A set of four columns is used to relate a query protein to an experimentally characterized homolog, a GSP (Gold Standard Protein) (isofunc, %seq_id, Locus tag, UniProt). The column

isofunc indicates if the query protein and its Gold Standard Protein homolog are isofunctional. The meaning of the terms used in this column in Tables 1-9 (yes, no, yes/no, probably, possibly, unclear, unknown, prediction, special, "-") is described at the end of this legend. The column %seq_id indicates the protein sequence identity between the query protein and the homologous GSP. The column Locus tag contains the locus tag, if assigned. The column UniProt contains the UniProt accession of the GSP. GSPs have been experimentally characterized as described in a publication. The column Reference links to the reference list of the manuscript. The column PMID lists the PubMed ID of the publication, if available. Otherwise, this is indicated as "not in PubMed". Also, one PhD thesis is indicated (PhD_Matter). The column Comment provides various types of additional information. The terms used in the column isofunc in Tables 1-9 have the following meaning: The term "yes" indicates that we consider the two proteins as isofunctional and annotate the query protein accordingly. The term "no" is used when we conclude that the proteins differ in function. Additional terms are used for more difficult cases. The term "yes/no" is used for GSPs which are multifunctional, and we assign only one a subset of these functions to the query protein. The term "probably" is used when we consider isofunctionality likely and annotated the query protein accordingly (with the term probable added to the protein name). The term "possibly" is used when we see a good chance that the proteins are isofunctional, but consider it too speculative to annotate the protein accordingly. The term "unclear" is used when we consider it likely that the same overall reaction is catalyzed, but when reaction details, e.g. the energy-providing compound, is unresolved. The term "unknown" is used when it is not possible to predict the substrate of the query protein. The term "prediction" is used if a function assignment is based on bioinformatic analyses but not yet on an experimentally characterized homologous protein. The term "special" is used when multiple arguments have to be considered with full details provided in the corresponding section of Suppl. Text S1. Finally, a dash ("-") is used when isofunctionality does not apply, e.g. when a homologous Gold Standard Protein could not be identified.

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 2.

			Gold Standard Protein						
Section	**Code**	**Gene**	**isofunc**	**%seq_id**	**Locus tag**	**UniProt**	**Reference**	**PMID**	**Comment**
2a	HVO_0047	*argW*	no	54%	TT_C1544	Q72HE5	(Yoshida et al., 2015)	25392000	for Arg, not for Lys biosynthesis
2a	HVO_0047 (cont.)		yes/no	39%	Saci_0753	Q4JAQ0			only for Arg, not for Lys biosynthesis
2a	HVO_0047 (cont.)		yes/no	61%	TK0279	Q5JFV9	(Yoshida et al., 2016)	27566549	only for Arg, not for Lys biosynthesis
2a	HVO_0046	*argX*	no	44%	TT_C1543	Q72HE6	(Horie et al., 2009)	19620981	for Arg, not for Lys biosynthesis
2a	HVO_0046 (cont.)		yes	30%	Saci_1621	Q4J8E7			only for Arg, not for Lys biosynthesis
2a	HVO_0046 (cont.)		yes/no	37%	TK0278	Q5JFW0	(Yoshida et al., 2016)	27566549	only for Arg, not for Lys biosynthesis
2a	HVO_0044	*argB*	no	41%	TT_C1541	O50147	(Horie et al., 2009) (Yoshida et al., 2015)	19620981 25392000	for Arg, not for Lys biosynthesis
2a	HVO_0044 (cont.)		yes/no	33%	Saci_0751	Q4JAQ2	(Ouchi et al., 2013)	23434852	only for Arg, not for Lys biosynthesis
2a	HVO_0044 (cont.)		yes/no	32%	TK0276	Q5JFW2	(Yoshida et al., 2016)	27566549	only for Arg, not for Lys biosynthesis
2a	HVO_0045	*argC*	no	48%	TT_C1542	O50146	(Horie et al., 2009) (Shimizu et al., 2016)	19620981 26966182	for Arg, not for Lys biosynthesis
2a	HVO_0045 (cont.)		yes/no	42%	Saci_0750	Q4JAQ3	(Ouchi et al., 2013)	23434852	only for Arg, not for Lys biosynthesis
2a	HVO_0045 (cont.)		yes/no	46%	TK0277	Q5JFW1	(Yoshida et al., 2016)	27566549	only for Arg, not for Lys biosynthesis
2a	HVO_0043	*argD*	no	45%	TT_C1393	Q93R93	(Miyazaki et al., 2001)	11489859	for Arg, not for Lys biosynthesis
2a	HVO_0043 (cont.)		yes/no	40%	Saci_0755	Q4JAP8	(Ouchi et al., 2013)	23434852	only for Arg, not for Lys biosynthesis
2a	HVO_0043 (cont.)		yes/no	42%	TK0275	Q5JFW3	(Yoshida et al., 2016)	27566549	only for Arg, not for Lys biosynthesis

2a	HVO_0042	*argE*	no	36%	TT_C1396	Q8VUS5	(Horie et al., 2009) (Fujita et al., 2017)	19620981 28720495	for Arg, not for Lys biosynthesis
2a	HVO_0042 (cont.)		yes/no	29%	Saci_0756	Q4JAP7	(Ouchi et al., 2013)	23434852	only for Arg, not for Lys biosynthesis
2a	HVO_0042 (cont.)		yes/no	37%	TK0274	Q5JFW4	(Yoshida et al., 2016)	27566549	only for Arg, not for Lys biosynthesis
2a	HVO_0041	*argF*	yes	50%	P18186	BSU11250	(Issaly and Issaly, 1974)	4216455	
2a	HVO_0041 (cont.)		yes	47%	OE_5205R	B0R9X3	(Ruepp et al., 1995)	7868583	
2a	HVO_0049	*argG*	yes	35%	-	P00966	(Shaheen et al., 1994)	8792870	human
2a	HVO_0049 (cont.)		yes	23%	b3172	P0A6E4	(Lemke et al., 1999)	10666579	*E. coli*
2a	HVO_0048	*argH*	yes	38%	MMP0013	O74026	(Cohen-Kupiec et al., 1999)	10220900	
2a	HVO_0008	*lysC*	yes	32%	BSU28470	P08495	(Kato et al., 2004)	15033471	
2a	HVO_2487	*asd*	yes	51%	MJ0205	Q57658	(Faehnle et al., 2005)	16225889	
2a/9e	HVO_1101	*dapA*	yes	45%	PA1010	Q9I4W3	(Kaur et al., 2011)	21396954	
2a	HVO_1100	*dapB*	yes	33%	b0031	P04036	(Reddy et al., 1995)	7893644	
2a	HVO_1099	*dapD*	yes	32%	b0166	P0A9D8	(Simms et al., 1984)	6365916	
2a	HVO_1096	*dapE*	yes	29%	b2472	P0AED7	(Lin et al., 1988)	3276674	function supported by gene clustering
2a	HVO_1097	*dapF*	yes	35%	b3809	P0A6K1	(Wiseman and Nichols, 1984)	6378903	
2a	HVO_1098	*lysA*	yes	38%	b2838	P00861	(White and Kelly, 1965)	14343156	
2a	HVO_A0634	-	unknown	25%	b2472	P0AED7	(Lin et al., 1988)	3276674	function assigned to HVO_1096 in *dap* cluster
2b	HVO_0790	*fba2*	special	67%	OE_1472F	B0R334	(Gulko et al., 2014)	25216252	EC 2.2.1.10 activity of OE_1472F not yet confirmed in vitro
2b	HVO_0790 (cont.)		special	45%	MJ0400	Q57843	(White, 2004)	15182204	substrate uncertain
2b	HVO_0792	*aroB*	yes	69%	OE_1475F	B0R336	(Gulko et al., 2014)	25216252	OE_1475F only partially characterized
2b	HVO_0792 (cont.)		yes	44%	MJ1249	Q58646	(White, 2004)	15182204	
2b	HVO_0602	*aroD1*	yes	44%	OE_1477R	B0R338	(Gulko et al., 2014)	25216252	

2b	HVO_0602 (cont.)		yes	31%	MMP1394	Q6LXF7	(Porat et al., 2004)	15262931	
2c	HVO_0009	*tnaA*	yes	41%	b3708	P0A853	(Phillips and Gollnick, 1989) (Newton et al., 1965)	2659590 14284727	
2d	HVO_A0559	*hutH*	yes	42%	BSU39350	P10944	(Oda et al., 1988) (Hartwell and Magasanik, 1963)	2454913 14066617	
2d	HVO_A0562	*hutU*	yes	62%	BSU39360	P25503	(Kaminskas et al., 1970)	4990470	
2d	HVO_A0560	*hutI*	yes	42%	BSU39370	P42084	(Yu et al., 2006)	16990261	
2d	HVO_A0561	*hutG*	yes	33%	BSU39380	P42068	(Kaminskas et al., 1970)	4990470	
2e	HVO_0431	-	-						no GSP available
2e	HVO_0644	*leuA1*	yes/no	47%	MJ1392	Q58787	(Howell et al., 1999)	9864346	HVO_0644 monofunc (CimA) or bifunc (CimA+LeuA); MJ1392 CimA
2e	HVO_0644 (cont.)		unclear	44%	MJ1195	Q58595	(Howell et al., 1998)	9665716	HVO_0644 monofunc (CimA) or bifunc (CimA+LeuA); MJ1195 LeuA
2e/2f	HVO_1510	*leuA2*	yes	47%	MJ1195	Q58595	(Howell et al., 1998)	9665716	HVO_1510 LeuA; MJ1195 LeuA
2e/2f	HVO_1510 (cont.)		no	41%	MJ1392	Q58787	(Howell et al., 1999)	9864346	HVO_1510 LeuA MJ1392 CimA
2e	HVO_A0489	-	no	31%	MJ1392	Q58787	(Howell et al., 1999)	9864346	HVO_A0489 general function only; MJ1392 CimA
2e	HVO_A0489 (cont.)		no	30%	MJ1195	Q58595	(Howell et al., 1998)	9665716	HVO_A0489 general function only; MJ1195 LeuA
2e	HVO_1153	-	-						function unassigned; no GSP

Table 2: **Proteins with open annotation issues and their Gold Standard Protein homologs (Section 2).** For a description of this table see the legend to 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 tryptophan 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 with an attributed function as 2-isopropylmalate synthase (EC 2.3.3.13). 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 S11 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. This is illustrated in Figure 1.

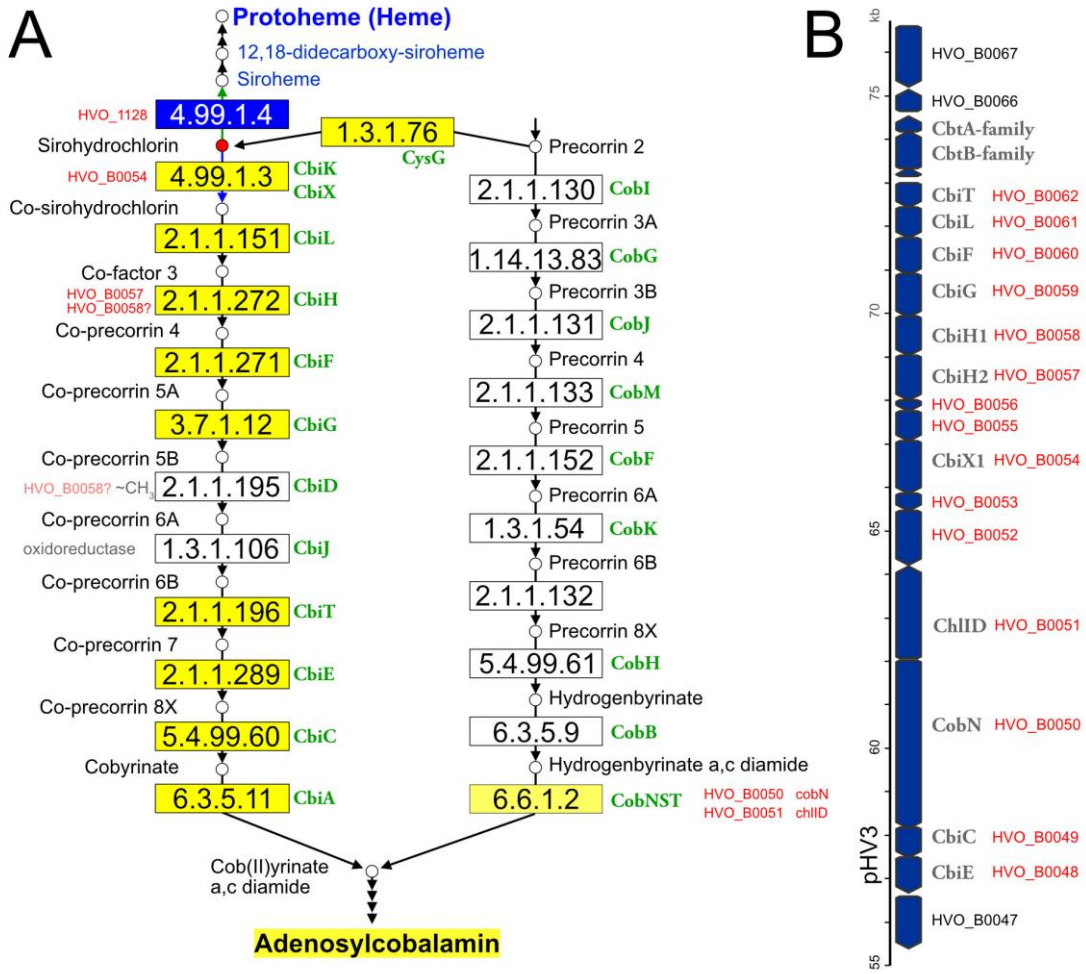

Figure 1. **Illustration of the haloarchaeal cobalamin and heme biosynthesis pathways and of the major cobalamin biosynthesis gene cluster.** (A) Biosynthesis pathways. This illustration is based on the corresponding KEGG map 00860. Small circles represent pathway intermediates and have their names assigned. Pathway intermediates upstream of Precorrin 2 are not displayed. The circle for sirohydrochlorin is highlighted in red as this is the branchpoint for heme and cobalamin biosynthesis in haloarchaea. Enzymatic reactions are shown by arrows, EC numbers being provided in rectangular boxes. Rectangles are colored when the enzyme has been reconstructed for haloarchaea (blue: heme biosynthesis; dark

yellow: de novo cobalamin biosynthesis; light yellow: late cobaltochelatase which may be a salvage reaction). Gene names in green are adopted from KEGG and represent those from bacterial model pathways. Consecutive arrowheads indicate reaction series which are not shown in detail for space reasons. For enzymatic reactions which are considered to be open issues, the *Hfx. volcanii* locus tags are provided. For two pathway gaps (white boxes in the cobalt-early pathway), the type of reaction is indicated (oxidoreductase and ~CH3, indicating a methylation reaction). The question mark after HVO_B0058 indicates that this protein, currently co-attributed to EC 2.1.1.272, is a candidate for the yet unassigned EC 2.1.1.195 reaction. We note that haloarchaea might use a deviating biosynthesis pathway, e.g. by swapping the methylation and oxidoreductase reactions (not illustrated). (B) The major cobalamin cluster, encoded on megaplasmid pHV3. Arrows are used to indicate the coding strand and are roughly drawn to scale. If assigned, the gene name is provided in addition to the *Hfx. volcanii* locus tag. Locus tags in red indicate genes that are part of the cobalamin cluster.

(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 3. 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 branches of cobalamin biosynthesis and their associated GSPs are listed in Table 3. 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.

455

			Gold Standard Protein						
Section	**Code**	**Gene**	**isofunc**	**%seq_id**	**Locus tag**	**UniProt**	**Reference**	**PMID**	**Comment**
3a	HVO_B0054	*cbiX1*	yes	30%	-	O87690	(Raux et al., 2003)	12408752	cobaltochelatase
3a	HVO_B0054 (cont.)		yes	27%	MTH_1397	O27448	(Brindley et al., 2003)	12686546	cobaltochelatase
3a	HVO_1128	*cbiX2*	no	29%	AF0721	O29537	(Yin et al., 2006)	16835730	cobaltochelatase
3a	HVO_1128 (cont.)		no	28%	MTH_1397	O27448	(Brindley et al., 2003)	12686546	cobaltochelatase
3a	HVO_1128 (cont.)		no	29%	AF0721	O29537	(Yin et al., 2006)	16835730	cobaltochelatase
3a	NP_0734A	*cbiX3*	-						function unassigned; no GSP; distantly related to paralogs
3a	HVO_2312	*sirC*	yes/no	31%	Mbar_A1461	Q46CH4	(Storbeck et al., 2010)	21197080	precorrin-2 DH; no analysis for Fe-chelatase
3a	HVO_2312 (cont.)		yes/no	29%	STM3477	P25924	(Stroupe et al., 2003) (Pennington et al., 2020)	14595395 32054833	matches to the N-term domain which is bifunctional as precorrin-2 DH and Fe-chelatase
3a	HVO_2312 (cont.)		yes/no	29%	-	P61818	(Raux et al., 2003) (Schubert et al., 2008)	12408752 18588505	precorrin-2 DH; devoid of Fe-chelatase activity
3b	HVO_B0061	*cbiL*	no	32%	STM2024	Q05593	(Roessner et al., 1992)	1451790	equivalent reaction on cobalt-free substrate
3b	HVO_B0057	*cbiH2*	yes	45%	-	O87689	(Moore et al., 2013)	23922391	corresponds to N-term of O87689 which has a C-term extension
3b	HVO_B0057 (cont.)		no	40%	STM2027	Q05590	(Santander et al., 1997) (Santander et al., 2006)	9331403 16198574	equivalent reaction on cobalt-free substrate
3b	HVO_B0058	*cbiH1*	special	32%	-	O87689	(Moore et al., 2013)	23922391	corresponds to N-term of O87689 which has a C-term extension;

									more distant to O87689 than CbiH2
3b	HVO_B0058 (cont.)		no	30%	STM2027	Q05590	(Santander et al., 1997) (Santander et al., 2006)	9331403 16198574	equivalent reaction on cobalt-free substrate
3b	HVO_B0060	*cbiF*	no	40%	STM2029	P0A2G9	(Roessner et al., 1992) (Kajiwara et al., 2006)	1451790 16866557	equivalent reaction on cobalt-free substrate
3b	HVO_B0060 (cont.)		yes	38%	-	O87686	(Moore et al., 2013)	23922391	
3b	HVO_B0059	*cbiG*	yes	24%	-	O87687	(Moore et al., 2013)	23922391	
3b	pathway gap								EC 2.1.1.195
3b	pathway gap								EC 1.3.1.106
3b	HVO_B0062	*cbiT*	yes	36%	-	O87694	(Moore et al., 2013)	23922391	corresponds to the C-term of bifunctional O87694
3b	HVO_B0048	*cbiE*	yes	28%	-	O87694	(Moore et al., 2013)	23922391	corresponds to the N-term of bifunctional O87694
3b	HVO_B0049	*cbiC*	yes	33%	-	O87692	(Moore et al., 2013)	23922391	
3b	HVO_A0487	*cbiA*	no	37%	STM2035	P29946	(Fresquet et al., 2004)	15311923	equivalent reaction on cobalt-free substrate
3b	HVO_B0052	-	-						function unassigned; no GSP
3b	HVO_B0053	-	-						function unassigned; no GSP
3b	HVO_B0055	-	-						function unassigned; no GSP
3b	HVO_B0056	-	-						function unassigned; no GSP
3c	HVO_A0488	*cobA*	yes	31%	MM_3138	Q8PSE1	(Buan et al., 2006)	16672609	
3c	HVO_A0488 (cont.)		yes	30%	STM1718	P31570	(Fonseca et al., 2002)	12080060	
3c	HVO_2395	*pduO*	yes	37%	-	Q9XDN2	(Johnson et al., 2001)	11160088	PduO and CobA are isofunctional; In Q9XDN2, the PduO domain (N-term) is

									fused to a DUF336 domain
3c	HVO_A0553	*cbiP*	yes	63%	VNG_1576G OE_3246F	Q9HPL5 B0R5X2	(Woodson et al., 2003b)	14645280	
3c	HVO_0587	*cbiB*	yes	58%	VNG_1578H OE_3253F	Q9HPL3 B0R5X4	(Woodson et al., 2003b)	14645280	
3c	HVO_0592	*cbiZ*	yes	57%	VNG_1583C OE_3261F	Q9HPL3 B0R5X8	(Woodson and Escalante-Semerena, 2004)	14990804	
3c	HVO_0589	*cobY*	yes	47%	VNG_1581C OE_3257F	Q9HPL1 B0R5X6	(Woodson et al., 2003a)	12486068	
3c	HVO_0588	*cobS*	yes	30%	STM2017	Q05602	(Zayas and Escalante-Semerena, 2007)	17209023	
3c	-				STM0643	P39701	(O'Toole et al., 1994)	7929373	EC 3.1.3.73; CobC; no homolog in haloarchaea
3c	HVO_0586	-	prediction	-	-	-	(Rodionov et al., 2003)	12869542	EC 3.1.3.73; prediction for HSL01294 (VNG_1577C)
3c	pathway gap								EC 2.7.1.177
3c	HVO_0591	*cobD1*	yes	31%	STM0644	P97084	(Brushaber et al., 1998)	9446573	
3c	HVO_0593	*cobD2*	yes						no GSP; 51% seq_id to HVO_0591 (*cobD1*)
3c	HVO_0590	*cobT*	prediction				(Rodionov et al., 2003)	12869542	prediction for VNG_1572C
3c	halTADL_3045	*cobT*	yes	39%	STM0644	Q05603	(Trzebiatowski et al., 1994)	8206834	
3d	HVO_B0051	*cobN*	yes	34%	-	P29929	(Debussche et al., 1992)	1429466	
3d	HVO_B0051 (cont.)		no	29%	-	Q55284	(Jensen et al., 1996) (Jensen et al., 1998)	8663186 9716491	Mg chelatase
3d	HVO_B0050	*chlID*	no	46%	slr1030	P51634	(Jensen et al., 1996) (Jensen et al., 1998)	8663186 9716491	match to N-term; Mg chelatase
3d	HVO_B0050 (cont.)		no	33%	slr1777	P52772	(Jensen et al., 1996) (Jensen et al., 1998)	8663186 9716491	match to complete sequence, incl distant match to N-term; Mg chelatase
3e	HVO_1121	*ahbC*	yes	47%	Mbar_A1793	Q46BK8	(Bali et al., 2011)	21969545	

							(Kuhner et al., 2014)	24669201	
3e	HVO_2144	*ahbD*	yes	42%	Mbar_A1458	Q46CH7	(Kuhner et al., 2014)	24669201	
3e	HVO_2227	*ahbA*	yes	35%	-	I6UH61	(Bali et al., 2011)	21969545	
3e	HVO_2313	*ahbB*	yes	32%	-	I6UH61	(Bali et al., 2011)	21969545	

Table 3: **Proteins with open annotation issues and their Gold Standard Protein homologs (Section 3).** For a description of this table see the legend to Table 1.

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 that creates the carbon backbone of this coenzyme has been reconstructed. We list the enzymes with their associated GSPs in Table 4. Coenzyme F420 contains a phospholactate moiety, which was reported to originate 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 that 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.

491

			Gold Standard Protein						
Section	**Code**	**Gene**	**isofunc**	**%seq_id**	**Locus tag**	**UniProt**	**Reference**	**PMID**	**Comment**
4a	HVO_2198	*cofH*	yes	35%	MJ1431	Q58826	(Graham et al., 2003) (Philmus et al., 2015)	14593448 25781338	
4a	HVO_2201	*cofG*	yes	43%	MJ0446	Q57888	(Graham et al., 2003) (Decamps et al., 2012) (Philmus et al., 2015)	14593448 23072415 25781338	
4a	HVO_2202	*cofC*	yes	25%	MJ0887	Q58297	(Grochowski et al., 2008) (Bashiri et al., 2019) (Braga et al., 2019)	18260642 30952857 31469543	
4a	HVO_2479	*cofD*	yes	39%	MM_1874	Q8PVT6	(Forouhar et al., 2008) (Braga et al., 2019)	18252724 31469543	
4a	HVO_2479 (cont.)		yes	32%	MJ1256	Q58653	(Graupner et al., 2002a)	11888293	
4a	HVO_1936	*cofE*	yes	47%	AF_2256	O28028	(Nocek et al., 2007)	17669425	
4a	HVO_1936 (cont.)		yes	38%	MJ0768	Q58178	(Li et al., 2003)	12911320	
4b	HVO_0433	*npdG*	yes	38%	AF_0892	O29370	(Kunow et al., 1993)	not in PubMed	
4b	HVO_B0113	-	no	27%	Rv0132c	P96809	(Purwantini and Mukhopadhyay, 2013)	24349169	too distant to assume isofunctionality
4b	HVO_B0342	-	unknown	29%	-	O93734	(Klein et al., 1996) (Aufhammer et al., 2004)	8706724 15016352	too distant to assume isofunctionality
4b	NP_1902A	-	no	28%	-	Q9UXP0	(Haase et al., 1992) (Westenberg et al., 1999)	1735436 9933933	too distant to assume isofunctionality
4b	NP_4006A	-	no	27%	MJ0870	Q58280	(Johnson and Mukhopadhyay, 2005)	16048999	too distant to assume isofunctionality
4c/5c	HVO_1937	*mer*	no	38%	MTH_1752	O27784	(te Brömmelstroet et al., 1990) (Vaupel and Thauer, 1995) (Shima et al., 2000)	2298726 7649177 10891279	
4d	HVO_2911	*phr2*	yes	62%	VNG_1335G OE_2907R	Q9HQ46 B0R5D6	(Takao et al., 1989) (McCready and Marcello, 2003)	2681164 12773185	

4d	HVO_2843	*phr1*	no	45%	sll1629	P77967	(Brudler et al., 2003)	12535521	sll1629 implicated in transcription regulation
4d	HVO_2843 (cont.)		possibly	45%	At5g24850	Q84KJ5	(Kleine et al., 2003) (Selby and Sancar, 2006)	12834405 17062752	mediates photorepair of ssDNA
4d	HVO_1234	*phr3*	possibly	40%	Atu4765	A9CH39	(Zhang et al., 2013)	23589886	

Table 4: **Proteins with open annotation issues and their Gold Standard Protein homologs (Section 4).** For a description of this table see the legend to Table 1.

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 (Table 5), 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 that 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)

537

			Gold Standard Protein						
Section	**Code**	**Gene**	**isofunc**	**%seq_id**	**Locus tag**	**UniProt**	**Reference**	**PMID**	**Comment**
5a	HVO_0709	*pabA*	no	47%	TTHA1843	P05379	(Sato et al., 1988)	2844259	Trp biosynthesis
5a	HVO_0709 (cont.)		yes/no	39%	BSU00750	P28819	(Slock et al., 1990)	2123867	TrpG works with TrpE and with PabB
5a	HVO_0710	*pabB*	no	46%	TTHA1844	P05378	(Sato et al., 1988)	2844259	Trp biosynthesis
5a	HVO_0710 (cont.)		yes	44%	BSU00740	P28820	(Schadt et al., 2009)	19275258	PabB; para-aminobenzoate biosynthesis
5a	HVO_0708	*pabC*	no	36%	AF_0933	O29329	(Isupov et al., 2019)	30733943	branched-chain amino acids
5b	HVO_2348	*mptA*			self		(El Yacoubi et al., 2009)	19478918	gene deletion phenotypes
5b	HVO_2348 (cont.)		yes	41%	MJ0775	Q58185	(Grochowski et al., 2007)	17497938	common part of methanopterin and tetrahydrofolate biosynthesis
5b	HVO_A0533	-	unknown	27%	MJ0837	Q58247	(Mashhadi et al., 2009)	19746965	if isofunctional would resolve a pathway gap
5b	HVO_2628	-	no	31%	AF_2089	O28190	(Scott and Rasche, 2002)	12142414	first committed step to methanopterin biosynthesis
5b	HVO_2628 (cont.)		no	26%	MJ1427	Q58822	(Dumitru and Ragsdale, 2004)	15262968	first committed step to methanopterin biosynthesis
5c	HVO_2573	*mch*	no	45%	MK0625	P94954	(Vaupel et al., 1998)	9676239	acts on a one-carbon attached to methanopterin
4c/5c	HVO_1937	*mer*	no	38%	MTH_1752	O27784	(te Brömmelstroet et al., 1990) (Vaupel and Thauer, 1995) (Shima et al., 2000)	2298726 7649177 10891279	acts on a one-carbon compound attached to methanopterin

538

Table 5: **Proteins with open annotation issues and their Gold Standard Protein homologs (Section 5).** For a description of this table see the legend to Table 1.

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 that 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). The enzymes and their associated GSPs are listed in Table 6.Three pathway gaps remain, with candidate genes predicted for two of these (Rodionova et al., 2017) (for details see Suppl. Text S1 Section 6).

			Gold Standard Protein						
Section	**Code**	**Gene**	**isofunc**	**%seq_id**	**Locus tag**	**UniProt**	**Reference**	**PMID**	**Comment**
6a	HVO_2363	*nadK1*	unclear	37%	Rv1695	P9WHV7	(Kawai et al., 2000)	11006082	can use ATP and PP
6a	HVO_2363 (cont.)		unclear	31%	AF_2373	O30297			ATP or PP usage unresolved
6a	HVO_0837	*nadK2*	unclear	28%	Rv1695	P9WHV7			can use ATP and PP
6a	HVO_0837 (cont.)		unclear	partial	AF_2373	O30297			ATP or PP usage unresolved
6b	HVO_0782	*nadM*	yes	53%	MJ0541	Q57961	(Raffaelli et al., 1997) (Raffaelli et al., 1999)	9401030 10331644	
6b	HVO_0781	-	unknown	42%	Sare_1364	A8M783	(Eustaquio et al., 2008)	18720493	
6b	HVO_0781 (cont.)		unknown	35%	PH0463	O58212	(Deng et al., 2008)	18551689	
6b	HVO_0327	*ribB*	yes	43%	MJ0055	Q60364	(Fischer et al., 2002)	12200440	
6b	HVO_0974	*ribH*	yes	45%	MJ0303	Q57751	(Haase et al., 2003)	12603336	
6b	HVO_1284	*arfA*		self			(Phillips et al., 2012)	21999246	gene deletion leads to riboflavin auxotrophy
6b	HVO_1284 (cont.)		yes	44%	MJ0145	Q57609	(Graham et al., 2002)	12475257	
6b	HVO_1235	-	prediction				(Rodionova et al., 2017)	28073944	*arfB* candidate
6b	HVO_1341	*arfC*	yes	36%	MJ0671	Q58085	(Graupner et al., 2002b) (Romisch-Margl et al., 2008)	11889103 18671734	
6b	HVO_2483	-	prediction	34%	MJ0699	Q58110	(Rodionova et al., 2017)	28073944	predicted also for MJ0699
6b	pathway gap								EC 3.1.3.104
6b	HVO_0326	*rbkR*	yes	37%	TA1064	Q9HJA6	(Rodionova et al., 2017)	28073944	bifunctional as gene regulator and enzyme
6b	HVO_0326 (cont.)		yes/no	32%	MJ0056	Q60365	(Ammelburg et al., 2007)	18073108	enzyme only; lacks an N-terminal HTH domain
6b	HVO_1015	*ribL*	yes	50%	MJ1179	Q58579	(Mashhadi et al., 2010)	20822113	

Table 6: **Proteins with open annotation issues and their Gold Standard Protein homologs (Section 6).** For a description of this table see the legend to Table 1.

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 that 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 that belongs to 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 (Table 7), with two pathway gaps remaining open (see Suppl.Text S1 Section 7 for details).

			Gold Standard Protein						
Section	**Code**	**Gene**	**isofunc**	**%seq_id**	**Locus tag**	**UniProt**	**Reference**	**PMID**	**Comment**
7a	NP_0604A	*idsA2*	yes	32%	GACE_1337	A0A0A7GEY4	(Petrova et al., 2018)	30062607	ortholog of HVO_0303 (66%); produces a C20 isoprenoid (same assignment for NP_0604A)
7a	NP_0604A (cont.)	*idsA2*	no	30%	APE_1764	Q9YB31 Q9UWR6	(Tachibana et al., 2000)	10632701	produces a C25 isoprenoid (C20 assigned to NP_0604A)
7a	NP_3996A	*idsA3*	yes	44%	GACE_1337	A0A0A7GEY4	(Petrova et al., 2018)	30062607	ortholog of HVO_2725 (67%); produces a C20 isoprenoid (same assignment for NP_3996A)
7a	NP_3996A (cont.)	*idsA2*	no	36%	APE_1764	Q9YB31 Q9UWR6	(Tachibana et al., 2000)	10632701	produces a C25 isoprenoid (C20 assigned to NP_3996A)
7a	NP_4556A	*idsA1*	no	34%	GACE_1337	A0A0A7GEY4	(Petrova et al., 2018)	30062607	no ortholog in *Hfx. volcanii*; produces a C20 isoprenoid (C25 assigned to NP_4556A)
7a	NP_4556A (cont.)	*idsA1*	yes	29%	APE_1764	Q9YB31 Q9UWR6	(Tachibana et al., 2000)	10632701	produces a C25 isoprenoid (same assignment for NP_4556A)
7b	HVO_0332	*carS*	yes	45%	AF_1740	O28537	(Jain et al., 2014)	25219966	
7b	HVO_1143	*assA*	yes	32%	MTH_1027	O27106	(Morii and Koga, 2003)	12562787	gene synonym: *pgsA3*
7b	HVO_1297	*aisA*	yes	25%	MTH_1691	O27726	(Morii et al., 2009)	19740749	gene synonym: *pgsA2*

7b	HVO_1136	*pgsA1*	-						only distant partial matches to GSPs
7b	HVO_1971	*pgsA4*	unclear	26%	MTH_1027	O27106	(Morii and Koga, 2003)	12562787	MTH_1027 is less distant to HVO_1143
7b	HVO_0146	*asd*	no	39%	SMc00551	Q9FDI9	(Vences-Guzman et al., 2008)	18708506	equivalent function for the bacterial lipid
7b	HVO_1295	*hisC*		self			(Conover and Doolittle, 1990)	2345144	complements a His auxotrophy mutant
7b	HVO_1295 (cont.)		yes	31%	b2021	P06986	(Grisolia et al., 1985)	2999081	weak support, see text
7b	HVO_1296	*adk2*	unclear	34%	PAB0757	Q9UZK4	(Loc'h et al., 2014)	24823650	*Pyrococcus*: involved in ribosome biogenesis
7b	HVO_1296 (cont.)		unclear	32%	-	Q9Y3D8	(Ren et al., 2005)	15630091	human: adenylate kinase; HVO_1296 may be inositol kinase
7b	HVO_2496	*adk1*	yes	45%	BSU01370	P16304	(Moon et al., 2019)	31111079	*Bacillus*: adenylate kinase
7b	HVO_B0213	-	yes	43%	AF_1794	O28480	(Chen et al., 2000) (Neelon et al., 2011)	11015222 22261071	*Archaeoglobus*: adenylate kinase
7b	HVO_1135	-	-						a SAM-dependent methyltransferase
7c	HVO_2524	*crtB*		self			(Kiljunen et al., 2014) (Maurer et al., 2018)	25488358 29038254	*crtB* mutants are colorless
7c	HVO_2524 (cont.)		yes	32%	Synpcc7942 _1984	P37269	(Chamovitz et al., 1992)	1537409	
7c	HVO_2527	*lyeJ*		self			(Dummer et al., 2011)	21840984	
7c	HVO_2527 (cont.)		yes	65%	VNG_1682C OE_3380R	Q9HPD9 B0R651	(Dummer et al., 2011)	21840984	
7c	HVO_2527 (cont.)		yes	61%	C444_12922	M0L7V9	(Yang et al., 2015)	25712483	
7c	HVO_2528	*crtD*		self			(Maurer et al., 2018)	29038254	a HVO_2528 mutant was white
7c	HVO_2528 (cont.)		yes	71%	C444_12917	A0A0A1GKA2	(Yang et al., 2015)	25712483	
7c	HVO_2526	*cruF*	yes	59%	C444_12927	A0A0A1GNF2	(Yang et al., 2015)	25712483	

7d	HVO_1470	*menF*	yes	38%	PA4231	Q51508	(Serino et al., 1995)	7500944	
7d	HVO_1469	*menD*	yes	37%	BSU30820	P23970	(Dawson et al., 2010)	20600129	
7d	pathway gap								EC 4.2.99.20
7d	HVO_1461	*menC*	no	29%	BSU12980	O34508	(Schmidt et al., 2001)	11747447	Ala/Glu epimerase
7d	HVO_1461 (cont.)		yes	24%	BSU30780	O34514	(Palmer et al., 1999)	10194342	o-succinylbenzoate synthase
7d	HVO_1375	*menE*	yes	36%	BSU30790	P23971	(Chen et al., 2016)	27933791	
7d	HVO_1465	*menB*	yes	66%	Rv0548c	P9WNP5	(Jiang et al., 2010)	20643650	
7d	pathway gap								EC 3.1.2.28
7d	HVO_1462	*menA*	yes	37%	b3930	P32166	(Suvarna et al., 1998)	9573170	
7d	HVO_0309	*menG*	yes/no	44%	At3g63410	Q9LY74	(Cheng et al., 2003)	14508009	*A.thaliana* enzyme also involved in tocopherol biosynthesis
7d	HVO_0309 (cont.)		yes	27%	-	O86169	(Koike-Takeshita et al., 1997)	9139683	

Table 7: **Proteins with open annotation issues and their Gold Standard Protein homologs (Section 7).** For a description of this table see the legend to Table 1.

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 (uS10) 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 (uS14) (ca 20% of the genomes, e.g. in *Nmn. pharaonis*). For more details see Suppl.Text S1 Section 8.

(c) The ribosomal protein L43e (eL43) shows heterogeneity with respect to the present of C2-C2 type zinc finger motif. This zinc finger is found in L43e 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 (Table 8), 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).

681

			Gold Standard Protein						
Section	**Code**	**Gene**	**isofunc**	**%seq_id**	**Locus tag**	**UniProt**	**Reference**	**PMID**	**Comment**
8a	OE_1279R	*rpoeps*		self			(Leffers et al., 1989) (Madon and Zillig, 1983)	2495365 6852054	
8b	HVO_0360	*rps10a*	yes	94%	rrnAC2405	P23357	(Arndt et al., 1991)	1764513	
8b	HVO_1392	*rps10b*	-						no GSP; 24% seq_id to HVO_0360 (*rps10a*)
8b	NP_4882A	*rps14a*	yes	72%	rrnAC1597.1	P26816	(Scholzen and Arndt, 1991)	1832208	full-length similarity; *Haloarcula* protein was not isolated or characterized
8b	NP_4882A (cont.)		yes	57%	YDL061C	P41058	(Otaka et al., 1984)	18782943	yeast YS29B; N-term 20 aa divergent
8b	NP_1768A	*rps14b*	unclear	80%	rrnAC1597.1	P26816	(Scholzen and Arndt, 1991)	1832208	N-term 20 aa divergent
8c	OE_1373R	*rpl43e*	yes	69%	rrnAC1669	P60619	(Ban et al., 2000)	10937989	
8c	OE_1373R (cont.)		yes	39%	YPR043W	P0CX25	(Rivlin et al., 1999) (Dresios et al., 2002)	10588896 11866512	
8c	HVO_0654	*rpl43e*	yes	54%	rrnAC1669	P60619	(Ban et al., 2000)	10937989	*Haloarcula*: has zinc finger; *Haloferax*; lacks zinc finger
8d	HVO_1631	*dph2*	yes	35%	PH1105	O58832	(Zhu et al., 2011)	20931132	
8d	HVO_0916	*dph5*	yes	39%	PH0725	O58456	(Zhu et al., 2010)	20873788	
8d	HVO_1077	*dph6*	yes	31%	YLR143W	Q12429	(Su et al., 2012) (Uthman et al., 2013)	23169644 23468660	
8e	HVO_0881	*sppA1*	yes	33%	BSU19530	O34525	(Bolhuis et al., 1999) (Nam et al., 2012)	10455123 22472423	
8e	HVO_1987	*sppA2*	probably	23%	BSU19530	O34525	(Bolhuis et al., 1999) (Nam et al., 2012)	10455123 22472423	
8e	HVO_1107	-	prediction						no GSP

682

Table 8: **Proteins with open annotation issues and their Gold Standard Protein homologs (Section 8).** For a description of this table see the legend to Table 1.

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) (Table 9).

			Gold Standard Protein						
Section	**Code**	**Gene**	**isofunc**	**%seq_id**	**Locus tag**	**UniProt**	**Reference**	**PMID**	**Comment**
9a	HVO_1812	-	prediction						no GSP
9b	halTADL_1913	-	yes	37%	-	P26984	(Aulkemeyer et al., 1991)	1809835	
9b	halTADL_1913 (cont.)	-	yes	31%	OCC_03567	Q7LYW8 H3ZP68	(Qu et al., 2004)	15138858	
9c	HVO_1711	-	probably	33%	-	P29761	(Ohnishi et al., 1992)	1633799	P29761 matches to C-term half of HVO_1711
9c	HVO_1711 (cont.)	-	probably	51%	SAMN 04487937_ 2677	A0A1I6HD35	(Chaga et al., 1993)	8305855	correlation between PMID:8305855 and A0A1I6HD35 likely (see text)
9d	HVO_1967	*pgi*	yes	36%	MJ1605	Q59000	(Rudolph et al., 2004)	14655001	
9e	OE_1665R	*kdgA*	no	31%	PA1010	Q9I4W3	(Kaur et al., 2011)	21396954	GSP for *dapA* (see under 2a)
9e	OE_1665R (cont.)		probably	30%	TTX_1156.1 TTX_1156a	G4RJQ2	(Ahmed et al., 2005)	15869466	
9e	OE_1665R (cont.)		probably	25%	SSO3197	Q97U28	(Ahmed et al., 2005)	15869466	
9f	HVO_1692	*ludB*		self			(Reinhardt et al., 2019)	30707467	
9f	HVO_1692 (cont.)		probably	35%	BSU34040	O07021	(Chai et al., 2009)	19201793	matches up to HVO_1692 pos 490 of 733
9f	HVO_1692 (cont.)		probably	35%	PST_3338	O4VPR6	(Gao et al., 2015)	25917905	matches up to HVO_1692 pos 400 of 733
9f	HVO_1693	*ludC*		self			(Reinhardt et al., 2019)	30707467	
9f	HVO_1693 (cont.)		probably	30%	BSU34030	O32259	(Chai et al., 2009)	19201793	
9f	HVO_1693 (cont.)		probably	33%	PST_3339	O4VPR7	(Gao et al., 2015)	25917905	partial match
9f	HVO_1697	-	unclear	24%	PST_3340	O4VPR8	(Gao et al., 2015)	25917905	

9f	HVO_1696	*lctP*	probably	44%	PST_3336	O4VPR4	(Gao et al., 2015)	25917905	
9g	HVO_B0300	*pucL1*	yes	49%	BSU32450	O32141	(Pfrimer et al., 2010)	20168977	*Bacillus*: bifunctional, matches to C-term
9g	HVO_B0299	*pucM*	yes	43%	BSU32460	O32142	(Lee et al., 2005)	16098976	
9g	HVO_B0301	*pucL2*	yes	43%	BSU32450	O32141	(Kim et al., 2007)	17567580	*Bacillus*: bifunctional, matches to N-term
9g	HVO_B0302	*pucH1*	no	33%	-	Q8VTT5	(Xu et al., 2002)	12148274	paper in Chinese, abstract in English; pyrimidine degradation
9g	HVO_B0302 (cont.)		yes	30%	STM0523	Q7CR08	(Ho et al., 2013)	23287969	purine degradation
9g	HVO_B0302 (cont.)		yes	29%	BSU32410	O32137	(Schultz et al., 2001)	11344136	purine degradation
9g	HVO_B0306	*amaB4*	no	39%	-	Q53389	(Martinez-Rodriguez et al., 2012)	22904279	carbamoyl-AA hydrolysis
9g	HVO_B0306 (cont.)		yes	34%	At5g43600	Q8VXY9	(Werner et al., 2010) (Werner et al., 2013)	19935661 23940254	purine degradation
9g	HVO_B0308	*coxS*	no	46%	Saci_2270	Q4J6M5	(Kardinahl et al., 1999)	10095793	GAPDH
9g	HVO_B0308 (cont.)		no	41%	-	P19915	(Kang and Kim, 1999)	10482497	CO-DH
9g	HVO_B0308 (cont.)		yes	39%	b2868	Q46801	(Xi et al., 2000)	10986234	xanthine DH
9g	HVO_B0309	*coxL*	yes	33%	b2866	Q46799	(Xi et al., 2000)	10986234	xanthine DH
9g	HVO_B0309 (cont.)		no	28%	-	P19913	(Kang and Kim, 1999)	10482497	CO-DH
9g	HVO_B0309 (cont.)		no	26%	Saci_2271	Q4J6M3	(Kardinahl et al., 1999)	10095793	GAPDH
9g	HVO_B0310	*coxM*	no	31%	Saci_2269	Q4J6M6	(Kardinahl et al., 1999)	10095793	GAPDH
9g	HVO_B0310 (cont.)		no	31%	-	P19914	(Kang and Kim, 1999)	10482497	CO-DH
9g	HVO_B0310 (cont.)		yes	25%	b2867	Q46800	(Xi et al., 2000)	10986234	xanthine DH
9g	HVO_B0303	*uraA4*	yes	38%	b3654	P0AGM9	(Karatza and Frillingos, 2005)	16096267	
9h	HVO_0197	-	possibly	39%	lp_0105	F9UST0	(Desguin et al., 2016)	27114550	LarB family protein
9h	HVO_2381	-	possibly	31%	lp_0106 /	F9UST1	(Desguin et al., 2016)	27114550	LarC family protein

					lp_0107				
9h	HVO_0190	-	possibly	34%	lp_0109	F9UST4	(Desguin et al., 2016)	27114550	LarE family protein
9i	HVO_1660	*dacZ*		self			(Braun et al., 2019)	30884174	
9i	HVO_0756	-	prediction				(He et al., 2020)	32095817	
9i	HVO_0990	-	prediction				(He et al., 2020)	32095817	
9i	HVO_1690	-	prediction				(He et al., 2020)	32095817	
9j	HVO_2763	-		self				22350204	no function could be assigned
9j	HVO_2763 (cont.)		no	27%	HVO_0144	D4GZ88	(Spath et al., 2008)	18437358	Rnase Z
9k	HVO_2410	*dabA*	yes	33%	Hneap_0211	D0KWS7	(Desmarais et al., 2019)	31406332	
9k	HVO_2411	*dabB*	yes	31%	Hneap_0212	D0KWS8	(Desmarais et al., 2019)	31406332	

Table 9: **Proteins with open annotation issues and their Gold Standard Protein homologs (Section 9).** For a description of this table see the legend to Table 1.

(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 that 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 that 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, Sebastien Ferreira-Cerca, 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.

AL-MAILEM, D. M., HOUGH, D. W. & DANSON, M. J. 2008. The 2-oxoacid dehydrogenase multienzyme complex of *Haloferax volcanii*. *Extremophiles,* 12**,** 89-96.

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.

AMMELBURG, M., HARTMANN, M. D., DJURANOVIC, S., ALVA, V., KORETKE, K. K., MARTIN, J., SAUER, G., TRUFFAULT, V., ZETH, K., LUPAS, A. N. & COLES, M. 2007. A CTP-dependent archaeal riboflavin kinase forms a bridge in the evolution of cradle-loop barrels. *Structure,* 15**,** 1577-90.

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.

ARNDT, E., SCHOLZEN, T., KROMER, W., HATAKEYAMA, T. & KIMURA, M. 1991. Primary structures of ribosomal proteins from the archaebacterium *Halobacterium marismortui* and the eubacterium *Bacillus stearothermophilus*. *Biochimie,* 73**,** 657-68.

AUFHAMMER, S. W., WARKENTIN, E., BERK, H., SHIMA, S., THAUER, R. K. & ERMLER, U. 2004. Coenzyme binding in F420-dependent secondary alcohol dehydrogenase, a member of the bacterial luciferase family. *Structure,* 12**,** 361-70.

AULKEMEYER, P., EBNER, R., HEILENMANN, G., JAHREIS, K., SCHMID, K., WRIEDEN, S. & LENGELER, J. W. 1991. Molecular analysis of two fructokinases involved in sucrose metabolism of enteric bacteria. *Mol Microbiol,* 5**,** 2913-22.

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.

BAN, N., NISSEN, P., HANSEN, J., MOORE, P. B. & STEITZ, T. A. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science,* 289**,** 905-20.

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.

BOLHUIS, A., MATZEN, A., HYYRYLAINEN, H. L., KONTINEN, V. P., MEIMA, R., CHAPUIS, J., VENEMA, G., BRON, S., FREUDL, R. & VAN DIJL, J. M. 1999. Signal peptide peptidase- and ClpP-like proteins of *Bacillus subtilis* required

for efficient translocation and processing of secretory proteins. *J Biol Chem,* 274**,** 24585-92.

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.

BRINDLEY, A. A., RAUX, E., LEECH, H. K., SCHUBERT, H. L. & WARREN, M. J. 2003. A story of chelatase evolution: identification and characterization of a small 13-15-kDa "ancestral" cobaltochelatase (CbiXS) in the archaea. *J Biol Chem,* 278**,** 22388-95.

BRUDLER, R., HITOMI, K., DAIYASU, H., TOH, H., KUCHO, K.-I., ISHIURA, M., KANEHISA, M., ROBERTS, V. A., TODO, T., TAINER, J. A. & GETZOFF, E. D. 2003. Identification of a new cryptochrome class. *Molecular Cell,* 11**,** 59-67.

BRUSHABER, K. R., O'TOOLE, G. A. & ESCALANTE-SEMERENA, J. C. 1998. CobD, a novel enzyme with L-threonine-O-3-phosphate decarboxylase activity, is responsible for the synthesis of (R)-1-amino-2-propanol O-2-phosphate, a proposed new intermediate in cobalamin biosynthesis in *Salmonella typhimurium* LT2. *J Biol Chem,* 273**,** 2684-91.

BUAN, N. R., REHFELD, K. & ESCALANTE-SEMERENA, J. C. 2006. Studies of the CobA-type ATP:Co(I)rrinoid adenosyltransferase enzyme of *Methanosarcina mazei* strain Go1. *J Bacteriol,* 188**,** 3543-50.

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.

CHAMOVITZ, D., MISAWA, N., SANDMANN, G. & HIRSCHBERG, J. 1992. Molecular cloning and expression in *Escherichia coli* of a cyanobacterial gene coding for phytoene synthase, a carotenoid biosynthesis enzyme. *FEBS Letters,* 296**,** 305-10.

CHEN, L., ZHOU, C., YANG, H. & ROBERTS, M. F. 2000. Inositol-1-phosphate synthase from *Archaeoglobus fulgidus* is a class II aldolase. *Biochemistry,* 39**,** 12415-23.

CHEN, Y., JIANG, Y. & GUO, Z. 2016. Mechanistic insights from the crystal structure of *Bacillus subtilis* o-succinylbenzoyl-CoA synthetase complexed with the adenylate intermediate. *Biochemistry,* 55**,** 6685-95.

CHENG, Z., SATTLER, S., MAEDA, H., SAKURAGI, Y., BRYANT, D. A. & DELLAPENNA, D. 2003. Highly divergent methyltransferases catalyze a conserved reaction in tocopherol and plastoquinone synthesis in cyanobacteria and photosynthetic eukaryotes. *Plant Cell,* 15**,** 2343-56.

COHEN-KUPIEC, R., KUPIEC, M., SANDBECK, K. & LEIGH, J. A. 1999. Functional conservation between the argininosuccinate lyase of the archaeon *Methanococcus maripaludis* and the corresponding bacterial and eukaryal genes. *FEMS Microbiol Lett,* 173**,** 231-8.

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.

CONOVER, R. K. & DOOLITTLE, W. F. 1990. Characterization of a gene involved in histidine biosynthesis in *Halobacterium* (*Haloferax*) *volcanii*: isolation and rapid mapping by transformation of an auxotroph with cosmid DNA. *J Bacteriol,* 172**,** 3244-9.

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, A., CHEN, M., FYFE, P. K., GUO, Z. & HUNTER, W. N. 2010. Structure and reactivity of *Bacillus subtilis* MenD catalyzing the first committed step in menaquinone biosynthesis. *J Mol Biol,* 401**,** 253-64.

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.

DEBUSSCHE, L., COUDER, M., THIBAUT, D., CAMERON, B., CROUZET, J. & BLANCHE, F. 1992. Assay, purification, and characterization of cobaltochelatase, a unique complex enzyme catalyzing cobalt insertion in hydrogenobyrinic acid a,c-diamide during coenzyme B12 biosynthesis in *Pseudomonas denitrificans*. *J Bacteriol,* 174**,** 7445-51.

DECAMPS, L., PHILMUS, B., BENJDIA, A., WHITE, R., BEGLEY, T. P. & BERTEAU, O. 2012. Biosynthesis of F0, precursor of the F420 cofactor, requires a unique two radical-SAM domain enzyme and tyrosine as substrate. *J Am Chem Soc,* 134**,** 18173-6.

DENDA, K., FUJIWARA, T., SEKI, M., YOSHIDA, M., FUKUMORI, Y. & YAMANAKA, T. 1991. Molecular cloning of the cytochrome aa3 gene from the archaeon (Archaebacterium) *Halobacterium halobium*. *Biochem Biophys Res Commun,* 181**,** 316-22.

DENG, H., BOTTING, C. H., HAMILTON, J. T., RUSSELL, R. J. & O'HAGAN, D. 2008. S-adenosyl-L-methionine:hydroxide adenosyltransferase: a SAM enzyme. *Angew Chem Int Ed Engl,* 47**,** 5357-61.

DESGUIN, B., SOUMILLION, P., HOLS, P. & HAUSINGER, R. P. 2016. Nickel-pincer cofactor biosynthesis involves LarB-catalyzed pyridinium carboxylation and LarE-dependent sacrificial sulfur insertion. *Proc Natl Acad Sci U S A,* 113**,** 5598-603.

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.

DRESIOS, J., CHAN, Y. L. & WOOL, I. G. 2002. The role of the zinc finger motif and of the residues at the amino terminus in the function of yeast ribosomal protein YL37a. *J Mol Biol,* 316**,** 475-88.

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.

DUMITRU, R. V. & RAGSDALE, S. W. 2004. Mechanism of 4-(beta-D-ribofuranosyl)aminobenzene 5'-phosphate synthase, a key enzyme in the methanopterin biosynthetic pathway. *J Biol Chem,* 279**,** 39389-95.

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.

EL YACOUBI, B., PHILLIPS, G., BLABY, I. K., HAAS, C. E., CRUZ, Y., GREENBERG, J. & DE CRECY-LAGARD, V. 2009. A Gateway platform for functional genomics in *Haloferax volcanii*: deletion of three tRNA modification genes. *Archaea,* 2**,** 211-9.

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.

EUSTAQUIO, A. S., HARLE, J., NOEL, J. P. & MOORE, B. S. 2008. S-Adenosyl-L-methionine hydrolase (adenosine-forming), a conserved bacterial and archaeal protein related to SAM-dependent halogenases. *Chembiochem,* 9**,** 2215-9.

FAEHNLE, C. R., OHREN, J. F. & VIOLA, R. E. 2005. A new branch in the family: structure of aspartate-beta-semialdehyde dehydrogenase from *Methanococcus jannaschii*. *J Mol Biol,* 353**,** 1055-68.

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, M., ROMISCH, W., SCHIFFMANN, S., KELLY, M., OSCHKINAT, H., STEINBACHER, S., HUBER, R., EISENREICH, W., RICHTER, G. & BACHER, A. 2002. Biosynthesis of riboflavin in archaea studies on the mechanism of 3,4-dihydroxy-2-butanone-4-phosphate synthase of *Methanococcus jannaschii*. *J Biol Chem,* 277**,** 41410-6.

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.

FONSECA, M. V., BUAN, N. R., HORSWILL, A. R., RAYMENT, I. & ESCALANTE-SEMERENA, J. C. 2002. The ATP:Co(I)rrinoid adenosyltransferase (CobA) enzyme of *Salmonella enterica* requires the 2'-OH group of ATP for function and yields inorganic triphosphate as its reaction byproduct. *J Biol Chem,* 277**,** 33127-31.

FOROUHAR, F., ABASHIDZE, M., XU, H., GROCHOWSKI, L. L., SEETHARAMAN, J., HUSSAIN, M., KUZIN, A., CHEN, Y., ZHOU, W., XIAO, R., ACTON, T. B., MONTELIONE, G. T., GALINIER, A., WHITE, R. H. & TONG, L. 2008. Molecular insights into the biosynthesis of the F420 coenzyme. *J Biol Chem,* 283**,** 11832-40.

FRESQUET, V., WILLIAMS, L. & RAUSHEL, F. M. 2004. Mechanism of cobyrinic acid a,c-diamide synthetase from *Salmonella typhimurium* LT2. *Biochemistry,* 43**,** 10619-27.

FUJITA, S., CHO, S. H., YOSHIDA, A., HASEBE, F., TOMITA, T., KUZUYAMA, T. & NISHIYAMA, M. 2017. Crystal structure of LysK, an enzyme catalyzing the last step of lysine biosynthesis in *Thermus thermophilus*, in complex with lysine: Insight into the mechanism for recognition of the amino-group carrier protein, LysW. *Biochem Biophys Res Commun,* 491**,** 409-15.

FUJIWARA, T., FUKUMORI, Y. & YAMANAKA, T. 1989. Purification and properties of *Halobacterium halobium* “cytochrome aa3” which lacks CuA and CuB. *J Biochem (Tokyo),* 105**,** 287-92.

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.

GRAHAM, D. E., XU, H. & WHITE, R. H. 2002. A member of a new class of GTP cyclohydrolases produces formylaminopyrimidine nucleotide monophosphates. *Biochemistry,* 41**,** 15074-84.

GRAHAM, D. E., XU, H. & WHITE, R. H. 2003. Identification of the 7,8-didemethyl-8-hydroxy-5-deazariboflavin synthase required for coenzyme F(420) biosynthesis. *Arch Microbiol,* 180**,** 455-64.

GRAUPNER, M., XU, H. & WHITE, R. H. 2002a. Characterization of the 2-phospho-L-lactate transferase enzyme involved in coenzyme F(420) biosynthesis in *Methanococcus jannaschii*. *Biochemistry,* 41**,** 3754-61.

GRAUPNER, M., XU, H. & WHITE, R. H. 2002b. The pyrimidine nucleotide reductase step in riboflavin and F(420) biosynthesis in archaea proceeds by the eukaryotic route to riboflavin. *J Bacteriol,* 184**,** 1952-7.

GRISOLIA, V., CARLOMAGNO, M. S., NAPPO, A. G. & BRUNI, C. B. 1985. Cloning, structure, and expression of the *Escherichia coli* K-12 *hisC* gene. *J Bacteriol,* 164**,** 1317-23.

GROCHOWSKI, L. L., XU, H., LEUNG, K. & WHITE, R. H. 2007. Characterization of an Fe(2+)-dependent archaeal-specific GTP cyclohydrolase, MptA, from *Methanocaldococcus jannaschii*. *Biochemistry,* 46**,** 6658-67.

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.

HAASE, I., MORTL, S., KOHLER, P., BACHER, A. & FISCHER, M. 2003. Biosynthesis of riboflavin in archaea. 6,7-dimethyl-8-ribityllumazine synthase of *Methanococcus jannaschii*. *Eur J Biochem,* 270**,** 1025-32.

HAASE, P., DEPPENMEIER, U., BLAUT, M. & GOTTSCHALK, G. 1992. Purification and characterization of F420H2-dehydrogenase from *Methanolobus tindarius*. *Eur J Biochem,* 203**,** 527-31.

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.

HARTWELL, L. H. & MAGASANIK, B. 1963. The molecular basis of histidase induction in *Bacillus subtilis*. *J Mol Biol,* 7**,** 401-20.

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.

HEATH, C., POSNER, M. G., AASS, H. C., UPADHYAY, A., SCOTT, D. J., HOUGH, D. W. & DANSON, M. J. 2007. The 2-oxoacid dehydrogenase multi-enzyme complex of the archaeon *Thermoplasma acidophilum* - recombinant expression, assembly and characterization. *FEBS J,* 274**,** 5406-15.

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**,** 1011-8.

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.

HO, Y. Y., HUANG, Y. H. & HUANG, C. Y. 2013. Chemical rescue of the post-translationally carboxylated lysine mutant of allantoinase and dihydroorotase by metal ions and short-chain carboxylic acids. *Amino Acids,* 44**,** 1181-91.

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.

HOWELL, D. M., HARICH, K., XU, H. M. & WHITE, R. H. 1998. Alpha-keto acid chain elongation reactions involved in the biosynthesis of coenzyme b (7-mercaptoheptanoyl threonine phosphate) in methanogenic archaea. *Biochemistry,* 37**,** 10108-17.

HOWELL, D. M., XU, H. & WHITE, R. H. 1999. (R)-citramalate synthase in methanogenic archaea. *J Bacteriol,* 181**,** 331-3.

HUANG, M., OPPERMANN-SANIO, F. B. & STEINBUCHEL, A. 1999. Biochemical and molecular characterization of the *Bacillus subtilis* acetoin catabolic pathway. *J Bacteriol,* 181**,** 3837-41.

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**,** e00810-16.

ISHIKAWA, R., ISHIDO, Y., TACHIKAWA, A., KAWASAKI, H., MATSUZAWA, H. & WAKAGI, T. 2002. *Aeropyrum pernix* K1, a strictly aerobic and hyperthermophilic archaeon, has two terminal oxidases, cytochrome ba3 and cytochrome aa3. *Arch Microbiol,* 179**,** 42-9.

ISSALY, I. M. & ISSALY, A. S. 1974. Control of ornithine carbamoyltransferase activityby arginase in *Bacillus subtilis*. *Eur J Biochem,* 49**,** 485-95.

ISUPOV, M. N., BOYKO, K. M., SUTTER, J. M., JAMES, P., SAYER, C., SCHMIDT, M., SCHONHEIT, P., NIKOLAEVA, A. Y., STEKHANOVA, T. N., MARDANOV, A. V., RAVIN, N. V., BEZSUDNOVA, E. Y., POPOV, V. O. & LITTLECHILD, J. A. 2019. Thermostable branched-chain amino acid transaminases from the archaea *Geoglobus acetivorans* and *Archaeoglobus fulgidus*: biochemical and structural characterization. *Front Bioeng Biotechnol,* 7**,** 7.

JAIN, S., CAFORIO, A., FODRAN, P., LOLKEMA, J. S., MINNAARD, A. J. & DRIESSEN, A. J. M. 2014. Identification of CDP-archaeol synthase, a missing link of ether lipid biosynthesis in Archaea. *Chem Biol,* 21**,** 1392-401.

JENSEN, P. E., GIBSON, L. C., HENNINGSEN, K. W. & HUNTER, C. N. 1996. Expression of the *chlI*, *chlD*, and *chlH* genes from the Cyanobacterium *Synechocystis* PCC6803 in *Escherichia coli* and demonstration that the three cognate proteins are required for magnesium-protoporphyrin chelatase activity. *J Biol Chem,* 271**,** 16662-7.

JENSEN, P. E., GIBSON, L. C. & HUNTER, C. N. 1998. Determinants of catalytic activity with the use of purified I, D and H subunits of the magnesium protoporphyrin IX chelatase from *Synechocystis* PCC6803. *Biochem J,* 334 (Pt 2)**,** 335-44.

JIANG, M., CHEN, M., GUO, Z. F. & GUO, Z. 2010. A bicarbonate cofactor modulates 1,4-dihydroxy-2-naphthoyl-coenzyme a synthase in menaquinone biosynthesis of *Escherichia coli*. *J Biol Chem,* 285**,** 30159-69.

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, C. L., PECHONICK, E., PARK, S. D., HAVEMANN, G. D., LEAL, N. A. & BOBIK, T. A. 2001. Functional genomic, biochemical, and genetic characterization of the *Salmonella pduO* gene, an ATP:cob(I)alamin adenosyltransferase gene. *J Bacteriol,* 183**,** 1577-84.

JOHNSON, E. F. & MUKHOPADHYAY, B. 2005. A new type of sulfite reductase, a novel coenzyme F420-dependent enzyme, from the methanarchaeon *Methanocaldococcus jannaschii*. *J Biol Chem,* 280**,** 38776-86.

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.

JOLLEY, K. A., MADDOCKS, D. G., GYLES, S. L., MULLAN, Z., TANG, S. L., DYALL-SMITH, M. L., HOUGH, D. W. & DANSON, M. J. 2000. 2-Oxoacid dehydrogenase multienzyme complexes in the halophilic Archaea? Gene sequences and protein structural predictions. *Microbiology (Reading),* 146 (Pt 5)**,** 1061-69.

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

KAJIWARA, Y., SANTANDER, P. J., ROESSNER, C. A., PEREZ, L. M. & SCOTT, A. I. 2006. Genetically engineered synthesis and structural characterization of cobalt-precorrin 5A and -5B, two new intermediates on the anaerobic pathway to vitamin B12: definition of the roles of the CbiF and CbiG enzymes. *J Am Chem Soc,* 128**,** 9971-8.

KAMINSKAS, E., KIMHI, Y. & MAGASANIK, B. 1970. Urocanase and N-formimino-L-glutamate formiminohydrolase of *Bacillus subtilis*, two enzymes of the histidine degradation pathway. *J Biol Chem,* 245**,** 3536-44.

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.

KANG, B. S. & KIM, Y. M. 1999. Cloning and molecular characterization of the genes for carbon monoxide dehydrogenase and localization of molybdopterin, flavin adenine dinucleotide, and iron-sulfur centers in the enzyme of *Hydrogenophaga pseudoflava*. *J Bacteriol,* 181**,** 5581-90.

KARATZA, P. & FRILLINGOS, S. 2005. Cloning and functional characterization of two bacterial members of the NAT/NCS2 family in *Escherichia coli*. *Mol Membr Biol,* 22**,** 251-61.

KARDINAHL, S., SCHMIDT, C. L., HANSEN, T., ANEMULLER, S., PETERSEN, A. & SCHAFER, G. 1999. The strict molybdate-dependence of glucose-degradation by the thermoacidophile *Sulfolobus acidocaldarius* reveals the first crenarchaeotic molybdenum containing enzyme - an aldehyde oxidoreductase. *European Journal of Biochemistry,* 260**,** 540-48.

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.

KATO, C., KURIHARA, T., KOBASHI, N., YAMANE, H. & NISHIYAMA, M. 2004. Conversion of feedback regulation in aspartate kinase by domain exchange. *Biochem Biophys Res Commun,* 316**,** 802-8.

KAUR, N., GAUTAM, A., KUMAR, S., SINGH, A., SINGH, N., SHARMA, S., SHARMA, R., TEWARI, R. & SINGH, T. P. 2011. Biochemical studies and crystal structure determination of dihydrodipicolinate synthase from *Pseudomonas aeruginosa*. *Int J Biol Macromol,* 48**,** 779-87.

KAWAI, S., MORI, S., MUKAI, T., SUZUKI, S., YAMADA, T., HASHIMOTO, W. & MURATA, K. 2000. Inorganic Polyphosphate/ATP-NAD kinase of Micrococcus flavus and Mycobacterium tuberculosis H37Rv. *Biochem Biophys Res Commun,* 276**,** 57-63.

KEIGHTLEY, J. A., ZIMMERMANN, B. H., MATHER, M. W., SPRINGER, P., PASTUSZYN, A., LAWRENCE, D. M. & FEE, J. A. 1995. Molecular genetic and protein chemical characterization of the cytochrome ba3 from *Thermus thermophilus* HB8. *J Biol Chem,* 270**,** 20345-58.

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. 1976. A ferredoxin from halobacteria. *FEBS Letters,* 67**,** 320-22.

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.

KERSCHER, L., OESTERHELT, D., CAMMACK, R. & HALL, D. O. 1976. A new plant-type ferredoxin from halobacteria. *Eur J Biochem,* 71**,** 101-7.

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.

KIM, K., PARK, J. & RHEE, S. 2007. Structural and functional basis for (S)-allantoin formation in the ureide pathway. *J Biol Chem,* 282**,** 23457-64.

KLEIN, A. R., BERK, H., PURWANTINI, E., DANIELS, L. & THAUER, R. K. 1996. Si-face stereospecificity at C5 of coenzyme F420 for F420-dependent glucose-6-phosphate dehydrogenase from *Mycobacterium smegmatis* and F420-dependent alcohol dehydrogenase from *Methanoculleus thermophilicus*. *Eur J Biochem,* 239**,** 93-7.

KLEINE, T., LOCKHART, P. & BATSCHAUER, A. 2003. An *Arabidopsis* protein closely related to *Synechocystis* cryptochrome is targeted to organelles. *Plant J,* 35**,** 93-103.

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.

KOIKE-TAKESHITA, A., KOYAMA, T. & OGURA, K. 1997. Identification of a novel gene cluster participating in menaquinone (vitamin K2) biosynthesis. Cloning and sequence determination of the 2-heptaprenyl-1,4-naphthoquinone methyltransferase gene of *Bacillus stearothermophilus*. *J Biol Chem,* 272**,** 12380-3.

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.

KUHNER, M., HAUFSCHILDT, K., NEUMANN, A., STORBECK, S., STREIF, J. & LAYER, G. 2014. The alternative route to heme in the methanogenic archaeon *Methanosarcina barkeri*. *Archaea,* 2014**,** Article ID 327637.

KUNOW, J., SCHWOËRER, B., STETTER, K. O. & THAUER, R. K. 1993. A F420-dependent NADP reductase in the extremely thermophilic sulfate-reducing *Archaeoglobus fulgidus*. *Archives of Microbiology,* 160**,** 199-205.

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.

LEE, T.-X., METZGER, S. U., CHO, Y. S., WHITMARSH, J. & KALLAS, T. 2001. Modification of inhibitor binding sites in the cytochrome bf complex by directed mutagenesis of cytochrome b6 in *Synechococcus* sp. PCC 7002. *Biochimica et Biophysica Acta (BBA) - Bioenergetics,* 1504**,** 235-47.

LEE, Y., LEE, D. H., KHO, C. W., LEE, A. Y., JANG, M., CHO, S., LEE, C. H., LEE, J. S., MYUNG, P. K., PARK, B. C. & PARK, S. G. 2005. Transthyretin-related proteins function to facilitate the hydrolysis of 5-hydroxyisourate, the end product of the uricase reaction. *FEBS Lett,* 579**,** 4769-74.

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.

LEMKE, C., YEUNG, M. & HOWELL, P. L. 1999. Expression, purification, crystallization and preliminary X-ray analysis of *Escherichia coli* argininosuccinate synthetase. *Acta Crystallogr D Biol Crystallogr,* 55**,** 2028-30.

LI, H., GRAUPNER, M., XU, H. & WHITE, R. H. 2003. CofE catalyzes the addition of two glutamates to F420-0 in F420 coenzyme biosynthesis in *Methanococcus jannaschii*. *Biochemistry,* 42**,** 9771-8.

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.

LIN, Y. K., MYHRMAN, R., SCHRAG, M. L. & GELB, M. H. 1988. Bacterial N-succinyl-L-diaminopimelic acid desuccinylase. Purification, partial characterization, and substrate specificity. *J Biol Chem,* 263**,** 1622-7.

LIU, J., KRULWICH, T. A. & HICKS, D. B. 2008. Purification of two putative type II NADH dehydrogenases with different substrate specificities from alkaliphilic *Bacillus pseudofirmus* OF4. *Biochim Biophys Acta,* 1777**,** 453-61.

LOC'H, J., BLAUD, M., RETY, S., LEBARON, S., DESCHAMPS, P., BAREILLE, J., JOMBART, J., ROBERT-PAGANIN, J., DELBOS, L., CHARDON, F., ZHANG, E., CHARENTON, C., TOLLERVEY, D. & LEULLIOT, N. 2014. RNA mimicry by the fap7 adenylate kinase in ribosome biogenesis. *PLoS Biol,* 12**,** e1001860.

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.

MARTINEZ-RODRIGUEZ, S., GARCIA-PINO, A., LAS HERAS-VAZQUEZ, F. J., CLEMENTE-JIMENEZ, J. M., RODRIGUEZ-VICO, F., GARCIA-RUIZ, J. M., LORIS, R. & GAVIRA, J. A. 2012. Mutational and structural analysis of L-N-carbamoylase reveals new insights into a peptidase M20/M25/M40 family member. *J Bacteriol,* 194**,** 5759-68.

MASHHADI, Z., XU, H., GROCHOWSKI, L. L. & WHITE, R. H. 2010. Archaeal RibL: a new FAD synthetase that is air sensitive. *Biochemistry,* 49**,** 8748-55.

MASHHADI, Z., XU, H. & WHITE, R. H. 2009. An Fe^{2+}-dependent cyclic phosphodiesterase catalyzes the hydrolysis of 7,8-dihydro-D-neopterin 2',3'-cyclic phosphate in methanopterin biosynthesis. *Biochemistry,* 48**,** 9384-92.

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

MATTAR, S. & ENGELHARD, M. 1997. Cytochrome ba3 from *Natronobacterium pharaonis*--an archaeal four-subunit cytochrome-c-type oxidase. *Eur J Biochem,* 250**,** 332-41.

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.

MAURER, S., LUDT, K. & SOPPA, J. 2018. Characterization of copy number control of two *Haloferax volcanii* replication origins using deletion mutants and haloarchaeal artificial chromosomes. *J Bacteriol,* 200.

MCCREADY, S. & MARCELLO, L. 2003. Repair of UV damage in *Halobacterium salinarum*. *Biochem Soc Trans,* 31**,** 694-8.

MILLER, M. J. & GENNIS, R. B. 1983. The purification and characterization of the cytochrome d terminal oxidase complex of the *Escherichia coli* aerobic respiratory chain. *J Biol Chem,* 258**,** 9159-65.

MIYAZAKI, J., KOBASHI, N., NISHIYAMA, M. & YAMANE, H. 2001. Functional and evolutionary relationship between arginine biosynthesis and prokaryotic lysine biosynthesis through alpha-aminoadipate. *J Bacteriol,* 183**,** 5067-73.

MOON, S., KIM, J., KOO, J. & BAE, E. 2019. Structural and mutational analyses of psychrophilic and mesophilic adenylate kinases highlight the role of hydrophobic interactions in protein thermal stability. *Struct Dyn,* 6**,** 024702.

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.

MORII, H., KIYONARI, S., ISHINO, Y. & KOGA, Y. 2009. A novel biosynthetic pathway of archaetidyl-myo-inositol via archaetidyl-myo-inositol phosphate from CDP-archaeol and D-glucose 6-phosphate in methanoarchaeon *Methanothermobacter thermautotrophicus* cells. *J Biol Chem,* 284**,** 30766-74.

MORII, H. & KOGA, Y. 2003. CDP-2,3-Di-O-geranylgeranyl-sn-glycerol:L-serine O-archaetidyltransferase (archaetidylserine synthase) in the methanogenic archaeon *Methanothermobacter thermautotrophicus*. *J Bacteriol,* 185**,** 1181-9.

MOSHIRI, F., CHAWLA, A. & MAIER, R. J. 1991. Cloning, characterization, and expression in *Escherichia coli* of the genes encoding the cytochrome d oxidase complex from *Azotobacter vinelandii*. *J Bacteriol,* 173**,** 6230-41.

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.

NAM, S. E., KIM, A. C. & PAETZEL, M. 2012. Crystal structure of *Bacillus subtilis* signal peptide peptidase A. *J Mol Biol,* 419**,** 347-58.

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

NEELON, K., ROBERTS, M. F. & STEC, B. 2011. Crystal structure of a trapped catalytic intermediate suggests that forced atomic proximity drives the catalysis of mIPS. *Biophys J,* 101**,** 2816-24.

NEWTON, W. A., MORINO, Y. & SNELL, E. E. 1965. Properties of crystalline tryptophanase. *J Biol Chem,* 240**,** 1211-8.

NG, S. Y., CHABAN, B., VANDYKE, D. J. & JARRELL, K. F. 2007. Archaeal signal peptidases. *Microbiology (Reading),* 153**,** 305-14.

NOCEK, B., EVDOKIMOVA, E., PROUDFOOT, M., KUDRITSKA, M., GROCHOWSKI, L. L., WHITE, R. H., SAVCHENKO, A., YAKUNIN, A. F., EDWARDS, A. & JOACHIMIAK, A. 2007. Structure of an amide bond forming F(420):gamma-glutamyl ligase from *Archaeoglobus fulgidus* -- a member of a new family of non-ribosomal peptide synthases. *J Mol Biol,* 372**,** 456-69.

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.

O'TOOLE, G. A., TRZEBIATOWSKI, J. R. & ESCALANTE-SEMERENA, J. C. 1994. The cobC gene of *Salmonella typhimurium* codes for a novel phosphatase involved in the assembly of the nucleotide loop of cobalamin. *Journal of Biological Chemistry,* 269**,** 26503-11.

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

ODA, M., SUGISHITA, A. & FURUKAWA, K. 1988. Cloning and nucleotide sequences of histidase and regulatory genes in the *Bacillus subtilis hut* operon and positive regulation of the operon. *J Bacteriol,* 170**,** 3199-205.

OHNISHI, H., KITAMURA, H., MINOWA, T., SAKAI, H. & OHTA, T. 1992. Molecular cloning of a glucoamylase gene from a thermophilic *Clostridium* and kinetics of the cloned enzyme. *Eur J Biochem,* 207**,** 413-8.

OPPERMANN, F. B., SCHMIDT, B. & STEINBUCHEL, A. 1991. Purification and characterization of acetoin:2,6-dichlorophenolindophenol oxidoreductase, dihydrolipoamide dehydrogenase, and dihydrolipoamide acetyltransferase of the *Pelobacter carbinolicus* acetoin dehydrogenase enzyme system. *J Bacteriol,* 173**,** 757-67.

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.

OTAKA, E., HIGO, K. & ITOH, T. 1984. Yeast ribosomal proteins. VIII. Isolation of two proteins and sequence characterization of twenty-four proteins from cytoplasmic ribosomes. *Mol Gen Genet,* 195**,** 544-6.

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.

PALMER, D. R., GARRETT, J. B., SHARMA, V., MEGANATHAN, R., BABBITT, P. C. & GERLT, J. A. 1999. Unexpected divergence of enzyme function and sequence: "N-acylamino acid racemase" is o-succinylbenzoate synthase. *Biochemistry,* 38**,** 4252-8.

PAN, X., CAO, D., XIE, F., XU, F., SU, X., MI, H., ZHANG, X. & LI, M. 2020. Structural basis for electron transport mechanism of complex I-like photosynthetic NAD(P)H dehydrogenase. *Nat Commun,* 11**,** 610.

PENNINGTON, J. M., KEMP, M., MCGARRY, L., CHEN, Y. & STROUPE, M. E. 2020. Siroheme synthase orients substrates for dehydrogenase and chelatase activities in a common active site. *Nat Commun,* 11**,** 864.

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.

PETROVA, T. E., BOYKO, K. M., NIKOLAEVA, A. Y., STEKHANOVA, T. N., GRUZDEV, E. V., MARDANOV, A. V., STROILOV, V. S., LITTLECHILD, J. A., POPOV, V. O. & BEZSUDNOVA, E. Y. 2018. Structural characterization of geranylgeranyl pyrophosphate synthase GACE1337 from the hyperthermophilic archaeon *Geoglobus acetivorans*. *Extremophiles,* 22**,** 877-88.

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.

PFRIMER, P., DE MORAES, L. M., GALDINO, A. S., SALLES, L. P., REIS, V. C., DE MARCO, J. L., PRATES, M. V., BLOCH, C., JR. & TORRES, F. A. 2010. Cloning, purification, and partial characterization of *Bacillus subtilis* urate oxidase expressed in *Escherichia coli*. *J Biomed Biotechnol,* 2010**,** 674908.

PHILLIPS, G., GROCHOWSKI, L. L., BONNETT, S., XU, H., BAILLY, M., BLABY-HAAS, C., EL YACOUBI, B., IWATA-REUYL, D., WHITE, R. H. & DE CRECY-LAGARD, V. 2012. Functional promiscuity of the COG0720 family. *ACS Chem Biol,* 7**,** 197-209.

PHILLIPS, R. S. & GOLLNICK, P. D. 1989. Evidence that cysteine 298 Is in the active site of tryptophan indole-lyase. *Journal of Biological Chemistry,* 264**,** 10627-32.

PHILMUS, B., DECAMPS, L., BERTEAU, O. & BEGLEY, T. P. 2015. Biosynthetic versatility and coordinated action of 5'-deoxyadenosyl radicals in deazaflavin biosynthesis. *J Am Chem Soc,* 137**,** 5406-13.

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.

PORAT, I., WATERS, B. W., TENG, Q. & WHITMAN, W. B. 2004. Two biosynthetic pathways for aromatic amino acids in the archaeon *Methanococcus maripaludis*. *J Bacteriol,* 186**,** 4940-50.

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.

PURWANTINI, E. & MUKHOPADHYAY, B. 2013. Rv0132c of *Mycobacterium tuberculosis* encodes a coenzyme F420-dependent hydroxymycolic acid dehydrogenase. *PLoS One,* 8**,** e81985.

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

QU, Q., LEE, S. J. & BOOS, W. 2004. Molecular and biochemical characterization of a fructose-6-phosphate-forming and ATP-dependent fructokinase of the hyperthermophilic archaeon *Thermococcus litoralis*. *Extremophiles,* 8**,** 301-8.

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.

RAFFAELLI, N., EMANUELLI, M., PISANI, F. M., AMICI, A., LORENZI, T., RUGGIERI, S. & MAGNI, G. 1999. Identification of the archaeal NMN adenylyltransferase gene. *Molecular & Cellular Biochemistry,* 193**,** 99-102.

RAFFAELLI, N., PISANI, F. M., LORENZI, T., EMANUELLI, M., AMICI, A., RUGGIERI, S. & MAGNI, G. 1997. Characterization of nicotinamide

mononucleotide adenylyltransferase from thermophilic archaea. *J Bacteriol,* 179**,** 7718-23.

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.

RAUT, P., GLASS, J. B. & LIEBERMAN, R. L. 2021. Archaeal roots of intramembrane aspartyl protease siblings signal peptide peptidase and presenilin. *Proteins,* 89**,** 232-41.

RAUX, E., LEECH, H. K., BECK, R., SCHUBERT, H. L., SANTANDER, P. J., ROESSNER, C. A., SCOTT, A. I., MARTENS, J. H., JAHN, D., THERMES, C., RAMBACH, A. & WARREN, M. J. 2003. Identification and functional analysis of enzymes required for precorrin-2 dehydrogenation and metal ion insertion in the biosynthesis of sirohaem and cobalamin in *Bacillus megaterium*. *Biochem J,* 370**,** 505-16.

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.

REDDY, S. G., SACCHETTINI, J. C. & BLANCHARD, J. S. 1995. Expression, purification, and characterization of *Escherichia coli* dihydrodipicolinate reductase. *Biochemistry,* 34**,** 3492-501.

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

REN, H., WANG, L., BENNETT, M., LIANG, Y., ZHENG, X., LU, F., LI, L., NAN, J., LUO, M., ERIKSSON, S., ZHANG, C. & SU, X. D. 2005. The crystal structure of human adenylate kinase 6: An adenylate kinase localized to the cell nucleus. *Proc Natl Acad Sci U S A,* 102**,** 303-8.

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.

RIVLIN, A. A., CHAN, Y. L. & WOOL, I. G. 1999. The contribution of a zinc finger motif to the function of yeast ribosomal protein YL37a. *J Mol Biol,* 294**,** 909-19.

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.

ROESSNER, C. A., WARREN, M. J., SANTANDER, P. J., ATSHAVES, B. P., OZAKI, S.-I., STOLOWICH, N. J., IIDA, K. & SCOTT, A. I. 1992. Expression of 9 *Salmonella typhimurium* enzymes for cobinamide synthesis. Identification of the 11-methyl and 20-methyl transferases of corrin biosynthesis. *FEBS Letters,* 301**,** 73-78.

ROMISCH-MARGL, W., EISENREICH, W., HAASE, I., BACHER, A. & FISCHER, M. 2008. 2,5-diamino-6-ribitylamino-4(3H)-pyrimidinone 5'-phosphate synthases of fungi and archaea. *FEBS J,* 275**,** 4403-14.

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.

RUEPP, A., MULLER, H. N., LOTTSPEICH, F. & SOPPA, J. 1995. Catabolic ornithine transcarbamylase of halobacterium halobium (salinarium) - purification, characterization, sequence determination, and evolution. *Journal of Bacteriology,* 177**,** 1129-36.

SANTANDER, P. J., KAJIWARA, Y., WILLIAMS, H. J. & SCOTT, A. I. 2006. Structural characterization of novel cobalt corrinoids synthesized by enzymes of the vitamin B12 anaerobic pathway. *Bioorg Med Chem,* 14**,** 724-31.

SANTANDER, P. J., ROESSNER, C. A., STOLOWICH, N. J., HOLDERMAN, M. T. & SCOTT, A. I. 1997. How corrinoids are synthesized without oxygen: nature's first pathway to vitamin B12. *Chemistry & Biology,* 4**,** 659-66.

SATO, S., NAKADA, Y., KANAYA, S. & TANAKA, T. 1988. Molecular cloning and nucleotide sequence of *Thermus thermophilus* HB8 *trpE* and *trpG*. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression,* 950**,** 303-12.

SCHADT, H. S., SCHADT, S., OLDACH, F. & SUSSMUTH, R. D. 2009. 2-Amino-2-deoxyisochorismate is a key intermediate in *Bacillus subtilis* p-aminobenzoic acid biosynthesis. *J Am Chem Soc,* 131**,** 3481-3.

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**,** e00976-20.

SCHMIDT, D. M., HUBBARD, B. K. & GERLT, J. A. 2001. Evolution of enzymatic activities in the enolase superfamily: functional assignment of unknown proteins in *Bacillus subtilis* and *Escherichia coli* as L-Ala-D/L-Glu epimerases. *Biochemistry,* 40**,** 15707-15.

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.

SCHOLZEN, T. & ARNDT, E. 1991. Organization and nucleotide sequence of ten ribosomal protein genes from the region equivalent to the spectinomycin operon in the archaebacterium *Halobacterium marismortui*. *Mol Gen Genet,* 228**,** 70-80.

SCHUBERT, H. L., ROSE, R. S., LEECH, H. K., BRINDLEY, A. A., HILL, C. P., RIGBY, S. E. & WARREN, M. J. 2008. Structure and function of SirC from *Bacillus megaterium*: a metal-binding precorrin-2 dehydrogenase. *Biochem J,* 415**,** 257-63.

SCHULLER, J. M., BIRRELL, J. A., TANAKA, H., KONUMA, T., WULFHORST, H., COX, N., SCHULLER, S. K., THIEMANN, J., LUBITZ, W., SETIF, P., IKEGAMI, T., ENGEL, B. D., KURISU, G. & NOWACZYK, M. M. 2019. Structural adaptations of photosynthetic complex I enable ferredoxin-dependent electron transfer. *Science,* 363**,** 257-60.

SCHULTZ, A. C., NYGAARD, P. & SAXILD, H. H. 2001. Functional analysis of 14 genes that constitute the purine catabolic pathway in *Bacillus subtilis* and evidence for a novel regulon controlled by the PucR transcription activator. *J Bacteriol,* 183**,** 3293-302.

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. Available: https://www.biorxiv.org/content/biorxiv/early/2021/01/21/2021.01.21.427637.full.pdf [Accessed January 21, 2021].

SCOTT, J. W. & RASCHE, M. E. 2002. Purification, overproduction, and partial characterization of beta-RFAP synthase, a key enzyme in the methanopterin biosynthesis pathway. *J Bacteriol,* 184**,** 4442-8.

SELBY, C. P. & SANCAR, A. 2006. A cryptochrome/photolyase class of enzymes with single-stranded DNA-specific photolyase activity. *Proc Natl Acad Sci U S A,* 103**,** 17696-700.

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.

SERINO, L., REIMMANN, C., BAUR, H., BEYELER, M., VISCA, P. & HAAS, D. 1995. Structural genes for salicylate biosynthesis from chorismate in *Pseudomonas aeruginosa*. *Mol Gen Genet,* 249**,** 217-28.

SHAHEEN, N., KOBAYASHI, K., TERAZONO, H., FUKUSHIGE, T., HORIUCHI, M. & SAHEKI, T. 1994. Characterization of human wild-type and mutant argininosuccinate synthetase proteins expressed in bacterial cells. *Enzyme Protein,* 48**,** 251-64.

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**,** 172.

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.

SHIMA, S., WARKENTIN, E., GRABARSE, W., SORDEL, M., WICKE, M., THAUER, R. K. & ERMLER, U. 2000. Structure of coenzyme F(420) dependent

methylenetetrahydromethanopterin reductase from two methanogenic archaea. *J Mol Biol,* 300**,** 935-50.

SHIMIZU, T., TOMITA, T., KUZUYAMA, T. & NISHIYAMA, M. 2016. Crystal structure of the LysY.LysW complex from *Thermus thermophilus*. *J Biol Chem,* 291**,** 9948-59.

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.

SIMMS, S. A., VOIGE, W. H. & GILVARG, C. 1984. Purification and characterization of succinyl-CoA: tetrahydrodipicolinate N-succinyltransferase from *Escherichia coli*. *J Biol Chem,* 259**,** 2734-41.

SISIGNANO, M., MORBITZER, D., GATGENS, J., OLDIGES, M. & SOPPA, J. 2010. A 2-oxoacid dehydrogenase complex of *Haloferax volcanii* is essential for growth on isoleucine but not on other branched-chain amino acids. *Microbiology,* 156**,** 521-9.

SLOCK, J., STAHLY, D. P., HAN, C. Y., SIX, E. W. & CRAWFORD, I. P. 1990. An apparent *Bacillus subtilis* folic acid biosynthetic operon containing *pab*, an amphibolic *trpG* gene, a third gene required for synthesis of para-aminobenzoic acid, and the dihydropteroate synthase gene. *J Bacteriol,* 172**,** 7211-26.

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

SPATH, B., SCHUBERT, S., LIEBEROTH, A., SETTELE, F., SCHUTZ, S., FISCHER, S. & MARCHFELDER, A. 2008. Two archaeal tRNase Z enzymes: similar but different. *Arch Microbiol,* 190**,** 301-8.

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.

STORBECK, S., ROLFES, S., RAUX-DEERY, E., WARREN, M. J., JAHN, D. & LAYER, G. 2010. A novel pathway for the biosynthesis of heme in Archaea: genome-based bioinformatic predictions and experimental evidence. *Archaea,* 2010**,** 175050.

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.

STROUPE, M. E., LEECH, H. K., DANIELS, D. S., WARREN, M. J. & GETZOFF, E. D. 2003. CysG structure reveals tetrapyrrole-binding features and novel regulation of siroheme biosynthesis. *Nat Struct Biol,* 10**,** 1064-73.

SU, X., LIN, Z., CHEN, W., JIANG, H., ZHANG, S. & LIN, H. 2012. Chemogenomic approach identified yeast YLR143W as diphthamide synthetase. *Proc Natl Acad Sci U S A,* 109**,** 19983-7.

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.

SUVARNA, K., STEVENSON, D., MEGANATHAN, R. & HUDSPETH, M. E. 1998. Menaquinone (vitamin K2) biosynthesis: localization and characterization of the menA gene from *Escherichia coli*. *J Bacteriol,* 180**,** 2782-7.

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

TACHIBANA, A., YANO, Y., OTANI, S., NOMURA, N., SAKO, Y. & TANIGUCHI, M. 2000. Novel prenyltransferase gene encoding farnesylgeranyl diphosphate synthase from a hyperthermophilic archaeon, *Aeropyrum pernix* - Molecular

evolution with alteration in product specificity. *European Journal of Biochemistry,* 267**,** 321-28.

TAKAO, M., KOBAYASHI, T., OIKAWA, A. & YASUI, A. 1989. Tandem arrangement of photolyase and superoxide dismutase genes in *Halobacterium halobium*. *J Bacteriol,* 171**,** 6323-9.

TANAKA, M., OGAWA, N., IHARA, K., SUGIYAMA, Y. & MUKOHATA, Y. 2002. Cytochrome aa(3) in *Haloferax volcanii*. *J Bacteriol,* 184**,** 840-5.

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.

TE BRÖMMELSTROET, B. W., HENSGENS, C. M., KELTJENS, J. T., VAN DER DRIFT, C. & VOGELS, G. D. 1990. Purification and properties of 5,10-methylenetetrahydromethanopterin reductase, a coenzyme F420-dependent enzyme, from *Methanobacterium thermoautotrophicum* strain delta H. *Journal of Biological Chemistry,* 265**,** 1852-57.

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.

TRZEBIATOWSKI, J. R., O'TOOLE, G. A. & ESCALANTE-SEMERENA, J. C. 1994. The cobT gene of *Salmonella typhimurium* encodes the NaMN: 5,6-dimethylbenzimidazole phosphoribosyltransferase responsible for the synthesis of N1-(5-phospho-alpha-D-ribosyl)-5,6-dimethylbenzimidazole, an intermediate in the synthesis of the nucleotide loop of cobalamin. *J Bacteriol,* 176**,** 3568-75.

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.

UTHMAN, S., BAR, C., SCHEIDT, V., LIU, S., TEN HAVE, S., GIORGINI, F., STARK, M. J. & SCHAFFRATH, R. 2013. The amidation step of diphthamide biosynthesis in yeast requires DPH6, a gene identified through mining the DPH1-DPH5 interaction network. *PLoS Genet,* 9**,** e1003334.

VAN OOYEN, J. & SOPPA, J. 2007. Three 2-oxoacid dehydrogenase operons in *Haloferax volcanii*: expression, deletion mutants and evolution. *Microbiology,* 153**,** 3303-13.

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.

VAUPEL, M. & THAUER, R. K. 1995. Coenzyme F420-dependent N5,N10-methylenetetrahydromethanopterin reductase (Mer) from *Methanobacterium thermoautotrophicum* strain Marburg. Cloning, sequencing, transcriptional analysis, and functional expression in *Escherichia coli* of the *mer* gene. *Eur J Biochem,* 231**,** 773-8.

VAUPEL, M., VORHOLT, J. A. & THAUER, R. K. 1998. Overproduction and one-step purification of the N5,N10-methenyltetrahydromethanopterin cyclohydrolase (Mch) from the hyperthermophilic *Methanopyrus kandleri*. *Extremophiles,* 2**,** 15-22.

VENCES-GUZMAN, M. A., GEIGER, O. & SOHLENKAMP, C. 2008. *Sinorhizobium meliloti* mutants deficient in phosphatidylserine decarboxylase accumulate phosphatidylserine and are strongly affected during symbiosis with alfalfa. *J Bacteriol,* 190**,** 6846-56.

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*.

WANNER, C. & SOPPA, J. 2002. Functional role for a 2-oxo acid dehydrogenase in the halophilic archaeon *Haloferax volcanii*. *J Bacteriol,* 184**,** 3114-21.

WERNER, A. K., MEDINA-ESCOBAR, N., ZULAWSKI, M., SPARKES, I. A., CAO, F. Q. & WITTE, C. P. 2013. The ureide-degrading reactions of purine ring catabolism employ three amidohydrolases and one aminohydrolase in *Arabidopsis*, soybean, and rice. *Plant Physiol,* 163**,** 672-81.

WERNER, A. K., ROMEIS, T. & WITTE, C. P. 2010. Ureide catabolism in *Arabidopsis thaliana* and *Escherichia coli*. *Nat Chem Biol,* 6**,** 19-21.

WESTENBERG, D. J., BRAUNE, A., RUPPERT, C., MULLER, V., HERZBERG, C., GOTTSCHALK, G. & BLAUT, M. 1999. The F420H2-dehydrogenase from *Methanolobus tindarius*: cloning of the *ffd* operon and expression of the genes in *Escherichia coli*. *FEMS Microbiol Lett,* 170**,** 389-98.

WHITE, P. J. & KELLY, B. 1965. Purification and properties of diaminopimelate decarboxylase from *Escherichia coli*. *Biochem J,* 96**,** 75-84.

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.

WISEMAN, J. S. & NICHOLS, J. S. 1984. Purification and properties of diaminopimelic acid epimerase from *Escherichia coli*. *J Biol Chem,* 259**,** 8907-14.

WOODSON, J. D. & ESCALANTE-SEMERENA, J. C. 2004. CbiZ, an amidohydrolase enzyme required for salvaging the coenzyme B12 precursor cobinamide in archaea. *Proc Natl Acad Sci U S A,* 101**,** 3591-6.

WOODSON, J. D., PECK, R. F., KREBS, M. P. & ESCALANTE-SEMERENA, J. C. 2003a. The *cobY* gene of the archaeon *Halobacterium* sp. strain NRC-1 is required for de novo cobamide synthesis. *J Bacteriol,* 185**,** 311-6.

WOODSON, J. D., ZAYAS, C. L. & ESCALANTE-SEMERENA, J. C. 2003b. A new pathway for salvaging the coenzyme B12 precursor cobinamide in archaea requires cobinamide-phosphate synthase (CbiB) enzyme activity. *J Bacteriol,* 185**,** 7193-201.

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.

XI, H., SCHNEIDER, B. L. & REITZER, L. 2000. Purine catabolism in *Escherichia coli* and function of xanthine dehydrogenase in purine salvage. *J Bacteriol,* 182**,** 5332-41.

XU, Z., JIANG, W. H., JIAO, R. S. & YANG, Y. L. 2002. [Cloning, sequencing and high expression in *Escherichia coli* of D-hydantoinase gene from *Burkholderia pickettii*]. *Sheng Wu Gong Cheng Xue Bao,* 18**,** 149-54.

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, J., XU, L. X., CHERNEY, M. M., RAUX-DEERY, E., BINDLEY, A. A., SAVCHENKO, A., WALKER, J. R., CUFF, M. E., WARREN, M. J. & JAMES, M. N. 2006. Crystal structure of the vitamin B12 biosynthetic cobaltochelatase, CbiXS, from *Archaeoglobus fulgidus*. *J Struct Funct Genomics,* 7**,** 37-50.

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.

YOSHIDA, A., TOMITA, T., FUJIMURA, T., NISHIYAMA, C., KUZUYAMA, T. & NISHIYAMA, M. 2015. Structural insight into amino group-carrier protein-mediated lysine biosynthesis: crystal structure of the LysZ.LysW complex from *Thermus thermophilus*. *J Biol Chem,* 290**,** 435-47.

YU, Y., LIANG, Y. H., BROSTROMER, E., QUAN, J. M., PANJIKAR, S., DONG, Y. H. & SU, X. D. 2006. A catalytic mechanism revealed by the crystal structures of the imidazolonepropionase from *Bacillus subtilis*. *J Biol Chem,* 281**,** 36929-36.

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.

ZAYAS, C. L. & ESCALANTE-SEMERENA, J. C. 2007. Reassessment of the late steps of coenzyme B12 synthesis in *Salmonella enterica*: evidence that dephosphorylation of adenosylcobalamin-5'-phosphate by the CobC phosphatase is the last step of the pathway. *J Bacteriol,* 189**,** 2210-8.

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.

ZHANG, F., SCHEERER, P., OBERPICHLER, I., LAMPARTER, T. & KRAUSS, N. 2013. Crystal structure of a prokaryotic (6-4) photolyase with an Fe-S cluster and a 6,7-dimethyl-8-ribityllumazine antenna chromophore. *Proc Natl Acad Sci U S A,* 110**,** 7217-22.

ZHANG, P., BATTCHIKOVA, N., PAAKKARINEN, V., KATOH, H., IWAI, M., IKEUCHI, M., PAKRASI, H. B., OGAWA, T. & ARO, E. M. 2005. Isolation, subunit composition and interaction of the NDH-1 complexes from *Thermosynechococcus elongatus* BP-1. *Biochem J,* 390**,** 513-20.

ZHU, X., DZIKOVSKI, B., SU, X., TORELLI, A. T., ZHANG, Y., EALICK, S. E., FREED, J. H. & LIN, H. 2011. Mechanistic understanding of *Pyrococcus horikoshii* Dph2, a [4Fe-4S] enzyme required for diphthamide biosynthesis. *Mol Biosyst,* 7**,** 74-81.

ZHU, X., KIM, J., SU, X. & LIN, H. 2010. Reconstitution of diphthine synthase activity in vitro. *Biochemistry,* 49**,** 9649-57.

ZIMMERMANN, B. H., NITSCHE, C. I., FEE, J. A., RUSNAK, F. & MUNCK, E. 1988. Properties of a copper-containing cytochrome ba3: a second terminal oxidase from the extreme thermophile *Thermus thermophilus*. *Proc Natl Acad Sci U S A,* 85**,** 5779-83.