Chapter 3

Formation of the ether lipids archaeatidylglycerol and archaeatidylethanolamine in *Escherichia coli*

Biochemical Journal 2015, 470, 343-355

Antonella Caforio¹ Samta Jain¹,⁴, Peter Fodran², Melvin Siliakus³, Adriaan Minnaard², John van der Oost³ and Arnold J. M. Driessen¹

¹Department of Molecular Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9747 AG Groningen, The Netherlands; The Zernike Institute for Advanced Materials, University of Groningen, 9747 AG Groningen, The Netherlands

²Stratingh Institute for Chemistry, University of Groningen, 9747 AG Groningen, The Netherlands

³Department of Microbiology, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands

⁴Present address: Department of Medicine, Section of Infectious Diseases, Boston University School of Medicine, 02118 Boston, Massachusetts, United States of America
Ether lipid biosynthesis

Abstract

In archaea, the membrane phospholipids consist of isoprenoid hydrocarbon chains that are ether-linked to a sn-glycerol-1-phosphate backbone. This unique structure is believed to be vital for the adaptation of these microorganisms to extreme environments, but it also reflects an evolutionary marker that distinguishes archaea from bacteria and eukaryotes. CDP-archaeol is the central precursor for polar head group attachment. We examined various bacterial enzymes involved in the attachment of L-serine and glycerol as polar head groups for their promiscuity in recognizing CDP-archaeol as a substrate. Using a combination of mutated bacterial and archaeal enzymes, archaetidylethanolamine (AE) and archaetidylglycerol (AG) could be produced in vitro using nine purified enzymes while starting from simple building blocks. The ether lipid pathway constituted by a set of archaeal and bacterial enzymes was introduced into E. coli, which resulted in the biosynthesis of AE and AG. This is a further step in the reprogramming of E. coli for ether lipid biosynthesis.
Introduction

The cytoplasmic membrane is an essential constituent of cells. It forms a barrier that separates the cytosol from the external milieu. In conjunction with energy transducing complexes and transporter proteins, the phospholipid bilayer allows homeostasis of the intracellular concentration of nutrients and other metabolites within the cell [126]. The lipid composition of the cytoplasmic membrane differs between organisms and is one of the elements that distinguishes archaea from bacteria and eukarya. This marked diversity of the lipid composition between archaea and all other organisms is also termed “the lipid divide” that may finds its origin in the early stages of the evolution of life [88,89,91,127].

Archaeal lipids are composed of highly branched isoprenoid chains ether-linked to a glycerol-1-phosphate (G1P) backbone, compared to fatty acid chains ester-linked to the glycerol-3-phosphate (G3P) backbone as typically found in Bacteria and Eukarya. Besides these main characteristics, archaeal membranes display a further diversity in their lipid composition consisting of different modifications of the two major structures: sn-2,3-diphytanylglycerol diether, called archaeol and sn-2,3-diphytanylglycerol tetraether lipid, known as caldarchaeol [57,75,127]. The biosynthetic pathway leading to the formation of archaeal lipids has been studied in some detail [2,5,6,10,21,37] and most of the enzymes involved in the biosynthesis have been identified and characterized. However, the entire pathway is not completely understood, nor is it clear how caldarchaeol is formed. The isoprenoid building blocks isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) are synthesized via the mevalonate pathway in archaea [10] and are combined through sequential condensation reactions catalyzed by geranylgeranyl diphosphate (GGPP) synthase and farnesylgeranyl diphosphate (FGPP) synthase, depending on the length of the isoprenoid chain product [19,44]. G1P in archaea and G3P in bacteria and eukarya are formed by similar reactions although the enzymes involved, i.e., glycerol-1-phosphate dehydrogenase (G1PDH) [28,29] and glycerol-3-phosphate dehydrogenase (G3PDH), are not evolutionarily related and belong to different protein families [2]. The elongated isoprenoid chain and G1P are subsequently
Ether lipid biosynthesis

linked together through ether linkages by two prenyltransferases. A
cytosolic protein geranylgeranylglyceryl phosphate synthase (GGGP
synthase) selectively attaches the isoprenoid chain to the G1P [35,36]
leading to the first ether bond formation. Next, the di-O-
geranylgeranylphosphate synthase (DGGGP synthase) [38,39,41]
catalyzes the second ether bond formation linking another isoprenoid
chain to the lipid precursor yielding DGGGP. The following step concerns
the activation of DGGGP via a CTP-transferring reaction by CDP-archaeol
synthase (CarS) that was recently discovered [41]. CDP-archaeol is an
important intermediate for the successive steps in lipid biosynthesis
where the CDP group is replaced by a polar head group. Serine,
ethanolamine, glycerol and myo-inositol are common polar head groups
found throughout the three domains of life. The enzymes involved in the
initial replacement of CMP from a CDP-alcohol with different types of polar
head groups share mostly a common mode of action among eukarya,
bacteria and archaea (Figure 1). A different mechanism to produce
phosphatidylethanolamine (PE) has been described for Eukarya that
involves a direct replacement of the CDP group with ethanolamine
[128,129].

The E. coli membrane is composed of 70-80% of PE, 20-25% of
phosphatidylglycerol (PG) and 5% or less of cardiolipin
(diphosphatidylglycerol) as the major phospholipid components
[126,130,131]. PG represents the main anionic phospholipid and is
important for various cellular processes such as the initiation of DNA
synthesis [126,132–134] and protein translocation [135–137], whereas PE
is the main zwitterionic lipid. In order to re-program E. coli for the
synthesis of archaeal ether lipids, it will be essential to produce the two
archaeal lipids archaetidylethanolamine (AE) and archaetidylglycerol (AG)
to accommodate at least the polar head group composition. For polar head
group modification, the branch point in bacteria lies with CDP-
diacylglycerol (CDP-DAG), produced by CDP-diacylglycerol synthase
encoded by the cdsA gene [102]. CDP-DAG acts as a substrate for two sets
of enzymes. For the formation of PE, the CDP group of CDP-DAG is replaced
with L-serine by phosphatidylserine synthase (Pss) leading to the
production of phosphatidylserine (PS). Next, PS is decarboxylated by
**Figure 1** | Polar head group attachment in bacteria and archaea. The scheme represents the enzymatic steps of the lipid biosynthetic pathway for polar head group attachment in Archaea and Bacteria. The archaeal and the bacterial enzymes involved in the replacement of CDP-group with L-serine or glycerol from CDP-archaeol or CDP-diacylglycerol are indicated.

Phosphatidylserine decarboxylase (Psd) converting it into PE. In the other biosynthetic pathway the CDP group of CDP-DAG is initially replaced by a glycerol-3-phosphate (G3P) moiety by PG synthase (Pgs) leading to the formation of phosphatidylglycerol phosphate (PGP). The enzyme phosphatidylglycerol phosphatase (Pgp) removes the phosphate resulting in the formation of PG [70,126,130,131]. In archaea, the formation of the corresponding archaeal lipids, AE and AG, seem to take place via very similar mechanisms (Figure 1). The two biosynthetic branches diverge from the CDP-archaeol towards the formation of AE or AG. The former is produced by the sequential action of two enzymes, archaetidylserine synthase (Ass) [47] and archaetidylserine decarboxylase (Asd) which replace the CDP group with L-serine whereupon a decarboxylation
Ether lipid biosynthesis

reaction yields the AE. Likewise, AG synthesis involves the enzyme AG phosphate (AGP) synthase (Ags) which attaches a G3P to CDP-archaeol forming AGP which is subsequently dephosphorylated into AG by the action of archaetidylglycerol phosphatase (Agp)[7,93].

The bacterial and archaeal synthase enzymes involved in polar head group attachment all contain a well conserved domain, identified as D-G-x(3)-D-x(3)-D in the PROSITE database, that classifies these proteins as members of the CDP-alcohol phosphatidyltransferase family (Supplementary Figure S1) [7,43]. Previous bioinformatics analysis [7,8,93] revealed a wide distribution of these enzymes in bacteria and archaea suggesting the existence of an ancestral enzyme in the last universal common ancestor (LUCA) [43] able to produce both archaeal- and bacterial-like lipids. Therefore, this could indicate substrate promiscuity among these enzymes and possibly that bacterial enzymes are able to catalyze polar head group attachment to the CDP-archaeol. In this study, we have tested various bacterial enzymes for the formation of unsaturated AE and AG from CDP-archaeol and in vitro reconstituted the entire pathways employing simple building blocks and up to nine purified enzymes of archaeal and bacterial origin. In addition, we have introduced the archaeal ether lipid biosynthetic pathway for AE and AG synthesis into E. coli making use of the substrate promiscuity of bacterial enzymes catalyzing the final polar head group conversion steps.

Results

Archaetidylglycerol formation by bacterial PgsA and PgpA enzymes

PG synthesis in E. coli proceeds via two enzymatic steps. The first reaction is catalyzed by phosphatidylglycerophosphate synthase. E. coli contains two individual genes that encode for this enzyme activity [138,139]. The pgsA gene exclusively functions in phosphatidylglycerol formation [140], whereas the pgsB gene is also involved in the synthesis of the Lipid A core of Lipopolysaccharide (LPS) [103]. The resulting PGP is dephosphorylated to PG. In E. coli, three main PGPs are found, encoded by the pgpA, pgpB and pgpC genes [141]. PgpA exhibits a narrow substrate
Figure 2 | *In vitro* demonstration of AG biosynthesis involving the *E. coli* PgsA and PgpA.  
(A) Coomassie Blue-stained SDS-PAGE gels showing the Ni-NTA purified proteins PgsA (21 kDa) and PgpA (20.5 kDa) from *E. coli*. *In vitro* reactions using (B) DGGGP or (C) PA (C<sub>18:1</sub>) and the purified proteins as specified, to test the substrate specificity of the *E. coli* PgsA and PgpA. Total ion counts from LC-MS data were normalized using DDM as internal standard. The data are the average of three experiments ± SE. (D) Schematic representation of the *in vitro* reactions. The purified enzymes used in the experiments are indicated above the arrows and the mass spectra from the LC-MS runs of the corresponding products are shown: DGGGP (m/z= 715.51 [M-H]-), CDP<sub>ol</sub> (m/z= 1020.54 [M-H]-), AGP (m/z= 869.51 [M-H]-) and AG (m/z= 789.55 [M-H]-) in (B) and PA 18:1 (m/z= 699.50 [M-H]-), CDP<sub>DAG</sub> (m/z= 1004.54 [M-H]-), PGP (m/z= 853.50 [M-H]-) and PG (m/z= 773.54 [M-H]-) in (C).
Ether lipid biosynthesis

specificity dephosphorylating only PGP [142] while PgpB also shows phosphatase activity towards DAG pyrophosphate [147], phosphatidic acid and lysophosphatidic acid [144,145]. Bioinformatics analysis [7,8] revealed the occurrence of Pgs homologs in archaea that belong to the CDP-alcohol phosphatidylintransferase family (Supplementary Figure S1), suggesting functional and structural conservation of these enzymes among bacteria and archaea. The same analysis revealed a weak sequence conservation of Pgp in archaea, limited to PgpA and PgpB only (Supplementary Figure S2). To investigate the ability of bacterial enzymes to accept CDP-archaeol as a substrate for AG formation, the pgsA and pgpA genes of E. coli, were cloned and overexpressed in E. coli BL21. Both overexpressed proteins localized to the membrane fraction after cell fractionation [141]. Upon membrane solubilization with the detergent n-dodecyl-β-D-maltopyranoside (DDM), PgsA and PgpA could be purified by Ni-NTA affinity chromatography (Figure 2A). The activity of the enzymes was analyzed in in vitro coupled reactions in detergent solution using LC-MS for detection (Figure 2D). The activity of the E. coli PgsA towards the archaeal substrates CDP-archaeol (CDP-ol) was tested using chemically synthetised DGGGP (unsaturated archaetidic acid) as a substrate that was converted into CDP-ol in the presence of the A. fulgidus CarS and CTP as described previously [41] (Figure 2B, lane 2). Upon the addition of the E. coli PgsA and G3P, the formation of AGP could be demonstrated (m/z= 869.51 [M–H]) (Figure 2B, lane 3). In the subsequent reaction the conversion of AGP into AG (m/z= 789.55 [M–H]) was observed when also the E. coli PgpA was included in the reaction (Figure 2B, lane 4). The E. coli PgsA was unable to use archaeal DGGGP as substrate (Figure 2B, lane 5). As a control, the activities of PgsA and PgpA were also tested towards the bacterial substrate CDP-DAG. The latter was produced by incubating the purified E. coli CdsA protein with its substrates PA (C18:1) and CTP (Figure 2C, lane 2). The formation of PGP (m/z= 853.50 [M–H]) was detected only in the presence of the E. coli PgsA and G3P (Figure 2C, lane 3), which was further converted into the final product PG (m/z= 773.54 [M–H]) upon the addition of the E. coli PgpA (Figure 2C, lane 4). These results demonstrate that the E. coli PgsA and PgpA recognizes and
converts the archaeal substrate CDP-archaeol and AGP respectively, forming the archaeal polar lipid AG.

**Archaeidylethanolamine formation by bacterial PssA and Psd enzymes**

Like PG synthesis, PE synthesis also requires the action of two enzymes: Pss for the synthesis of phosphatidylserine (PS) which is further decarboxylated to PE by Psd [146]. Two different subclasses of Pss exist: Pss-I, a cytoplasmic protein present mainly in Gram-negative bacteria such as *E. coli*, and Pss-II, a membrane protein that is found in Gram-positive bacteria, yeast [147] and archaea [148]. The Pss-II enzymes contain a highly conserved domain present in CDP-alcohol phosphatidyltransferases [148]. Previously, Ass activity was identified in the membrane fraction of *Methanothermobacter thermautotrophicus* [47]. Secondary structure analysis of the archaeal members of the CDP-alcohol phosphatidyltransferase family indicates the presence of eight conserved transmembrane domains (TMDs) comparable to the corresponding bacterial protein family (Figure 3A). Interestingly, the PssA sequence of *B. subtilis* is substantially smaller than the other members of this family (only 177 amino acids instead of 451 amino acids of the *E. coli* Pss-I), but it shows the core of five TMDs (Figure 3B) but lacking two C-terminal TMDs. A membrane fraction of *B. subtilis* incubated with the different archaea-like substrates showed AS formation whereas such activity could not be demonstrated with *E. coli* membranes [47] suggesting a more narrow substrate specificity of the *E. coli* Pss-I than Pss-II enzymes. Therefore, the *pssA* gene of *B. subtilis* was cloned and overexpressed in *E. coli* BL21 strain under the control of T7 promoter. The protein was solubilized from the membrane with DDM and purified by Ni-NTA affinity chromatography (Figure 3C). Coupled *in vitro* reactions were performed using DGGGP as initial substrate along with the *A. fulgidus* CarS and CTP. Products were extracted with *n*-butanol and analyzed by LC-MS (Figure 3F). In the presence of the *B. subtilis* PssA, AS (m/z= 802.53 [M-H]-) formation was observed in the presence of L-serine and Mg$^{2+}$ (Figure 3D, lane 3). PssA showed no activity towards DGGGP indicating the strict requirement of a
Ether lipid biosynthesis

CDP-activated intermediate for the reaction (Figure 3D, lane 4). Moreover, no AS was formed in presence of EDTA (Figure 3D, lane 5) which chelates divalent cations, consistent with a requirement of Mg\textsuperscript{2+} for enzymatic activity [149]. As a control, formation of phosphatidylserine by B. subtilis PssA was demonstrated in the presence of PA (C\textsubscript{18:1}), CTP and the E. coli CdsA (Figure 3D, lane 6). To examine the activity of the E. coli PssA towards CDP-archaeol, the enzyme was also overexpressed and purified (Figure 3C). In the presence of CDP-archaeol as substrate, no AS formation was observed demonstrating that the E. coli PssA (the Pss-I enzyme) indeed does not recognize the archaeal precursor (Figure 3E, lane 1). The enzyme, however, converted CDP-DAG into PS (m/z= 786.53 [M-H]-) in a coupled enzyme assay using PA as substrate (Figure 3E, lane 4).

In archaea and bacteria, AE and PE are produced by a decarboxylation reaction of L-serine. In E. coli, the Psd is encoded by the psd gene that specifies a membrane-associated pro-enzyme which undergoes an autocatalytic internal cleavage [150] leading to two subunits, the α subunit containing a pyruvoyl prosthetic group and a β subunit [151,152]. Previous bioinformatics analysis [7] identified an archaeal hypothetical protein as a potential Asd showing sequence similarity to the bacterial Psd (Supplementary Figure S3). Moreover, a similar operon conservation of the two genes pss/psd, typical of several bacterial species, was found in some archaea [7]. However, the archaeal Asd has not been biochemically characterized. Because of the general mechanism of the decarboxylation reaction [128], the possibility exists that the endogenous Psd of E. coli is able to recognize AS. Therefore, the E. coli psd gene was overexpressed in E. coli BL21 strain under the control of T7 promoter. Membranes bearing overexpressed levels of Psd were solubilized with DDM and the enzyme was purified by Ni-NTA affinity chromatography (Figure 4A). The α-subunit showed a slower migration on SDS-PAGE as expected on the basis of its predicted molecular mass. This is probably due to the presence of a covalently attached pyruvoyl prosthetic group that may affect the folding state in SDS-PAGE [151]. The enzymatic activity of the E. coli Psd was tested by in vitro coupled reactions as described above and product formation was detected by LC-MS (Figure 4C). Conversion of AS to AE (m/z= 758.55 [M-H]-) by the E. coli Psd was observed in the presence of
Figure 3 | In vitro AS synthesis by the B. subtilis Pssa. (A) Averaged hydropathy profile of the bacterial Pss proteins (blue line) and archaeal ones (red line) is aligned highlighting their conservations in the two kingdoms. The bacterial hydrophobicity profile is based on a multiple sequence alignment of 260 sequences sharing between 40% and 80% sequence identity. Likewise the archaeal hydrophobicity plot is based on a multiple sequence alignment of 38 archaeal sequences having a sequence identity between 30% and 70%. The membrane topology prediction is depicted above the plot. (B) Hydropathy profile alignment of B. subtilis PssA (purple line) and the averaged hydropathy profile of the bacterial Pss protein family (blue line). The conserved transmembrane domains (TMDs) are indicated by Roman numbers. (C) Coomassie Blue-stained SDS-PAGE gels showing the Ni-NTA purified proteins B. subtilis PssA (21 kDa) and E. coli PssA (53.6 kDa). Specificity of the bacterial B. subtilis PssA (D) and E. coli PssA (E) towards DGGGP and PA (C18:1) as assessed by means of an in vitro assay using the purified enzymes. Total ion counts from LC-MS data were normalized using DDM as internal standard. The data are the average of three experiments ± SE. (F) Schematic representation of the performed in vitro reactions. The mass spectra from the LC-MS runs of the two products AS (m/z = 802.54 [M-H]-) and PS (m/z = 786.53 [M-H]-).
Ether lipid biosynthesis

Mg$^{2+}$, L-serine, *B. subtilis* PssA, *A. fulgidus* CarS, CTP and DGGGP (Figure 4B, lane 4).

**In vitro** reconstitution of archaetidylethanolamine and archaetidylglycerol formation

Previously, we have described the **in vitro** reconstitution of CDP-archaeol formation starting from the precursors IPP and DMAPP/FPP [41] using two bacterial enzymes and three archaeal enzymes that were overexpressed and purified from *E. coli* upon codon optimization. Due to the wider substrate specificity described above, the extension of this pathway for the reconstitution of the formation of AE and AG requires four additional bacterial enzymes described in the previous section.

The ether lipid biosynthetic pathway reconstitution [41] starts from the two isoprenoid building blocks IPP and FPP which undergo several cycles of condensation leading to isoprenoid chains with the required C$_{20}$ length [16]. GGPP (m/z = 449.19 [M-H]) formation was detected by LC-MS with

![Image](image_url)

**Figure 4** | In vitro AE formation involving the *E. coli* Psd. (A) Coomassie Blue-stained SDS-PAGE gel showing the Ni-NTA purified protein *E. coli* Psd. The α-subunit with a molecular weight of 14.4 kDa and the β-subunit of 28.6 kDa. (B) In vitro reactions using the chemically synthesized substrate DGGGP, and the purified proteins as specified. Total ion counts from LC-MS data were normalized using DDM as internal standard. The data are the average of three experiments ± SE. (C) The mass spectra from the LC-MS runs of the products AE (m/z = 758.55 [M-H]).
an *E. coli* IspA [25] mutant as described previously ([Figure 5A, lane 2]) [41]. Despite the unique feature of archaeal G1P dehydrogenases to synthesize the G1P, some bacterial enzymes are also able to perform this reaction [32]. The enzyme AraM (G1PDH) from *B. subtilis* [31] was used that produces the required glycerophosphate backbone which can be attached to the GGPP via the first ether bond leading to the synthesis of GGGP (m/z= 443.26 [M-H]-). *M. maripaludis* GGGP synthase (MmarC7_1004) [36] was used to catalyze the latter reaction which in combination with IspA and G1PDH and the substrates IPP, FPP, DHAP and NADH, leads to the conversion of GGPP into GGGP ([Figure 5A, lane 3]). DGGGP production (m/z= 715.51 [M-H]-) was observed in a subsequent reaction with the *A. fulgidus* DGGGP synthase (AF0404) [38] ([Figure 5A, lane 4]). Next, DGGGP was converted into CDP-archaeol (m/z= 1020.54 [M-H]-) by the *A. fulgidus* CarS (AF1740) in the presence of CTP [41] ([Figure 5A, lane 5]). CDP-archaeol is the precursor for the formation of AS (m/z= 802.51 [M-H]-) in the presence of the *B. subtilis* PssA and L-serine ([Figure 5A, lane 6]), which was further converted into AE (m/z= 758.54 [M-H]-) ([Figure 5A, lane 7]) by the *E. coli* Psd. In another reaction, the CDP-archaeol was converted into AGP (m/z= 869.51 [M-H]-) by the *E. coli* PgsA in the presence of G3P ([Figure 5A, lane 8]). AGP was then converted into AG (m/z= 789.54 [M-H]-) by the addition of the *E. coli* PgpA to the *in vitro* reactions ([Figure 5A, lane 9]). Taken together, the reactions described here employing purified enzymes represent the *in vitro* reconstitution of the entire archaeal lipid pathway using a set of archaeal and bacterial enzymes.

**Archaetidylglycerol and archaetidylethanolamine formation in *E. coli***

To reconstruct the entire archaeal ether lipid biosynthetic pathway into *E. coli*, a system of four compatible expression vectors was used to co-express seven ether lipid genes into *E. coli*. In this system, the vectors containing one or two genes each, as listed in the **Table 1**, allowed the simultaneous expression of three archaeal enzymes (*M. maripaludis* GGGPS, *A. fulgidus* DGGGPS and *A. fulgidus* CarS) and four bacterial
Ether lipid biosynthesis

**Figure 5** | *In vitro* reconstitution of the archaeal ether lipid pathway. *In vitro* reactions were performed using a combination of purified enzymes as specified and the substrates IPP, FPP, DHAP and NADH in presence of Mg$^{2+}$ and 0.2% of DDM. Each product was detected by LC-MS: FPP (m/z = 381 [M-H]), GGPP (m/z = 449.19 [M-H]), GGGP (m/z = 443.26 [M-H]), DGGGP (m/z = 715.51 [M-H]), CDP-archaeol (m/z = 1020.54 [M-H]), AS (m/z = 802.51 [M-H]), AE (m/z = 758.54 [M-H]), AGP (m/z = 869.51 [M-H]) and AG (m/z = 789.54 [M-H]). The total ion counts were normalized using DDM as internal standard. The graph represents average of two experiments ± SE.
enzymes (*E. coli* Idi, *E. coli* mutant IspA, *B. subtilis* AraM and *B. subtilis* PssA). All genes were expressed under the control of the T7 promoter and induced with 0.25 mM of ITPG for 3.5 h. Total lipids were extracted from the *E. coli* membrane fraction using the Bligh and Dyer method [153] and analyzed by LC-MS. Four different engineered *E. coli* strains were compared, containing a different combination of archaeal lipid enzymes: (1) control harboring only the empty vectors; (2) – five ether lipid enzymes (*E. coli* Idi, *E. coli* mutant IspA, *B. subtilis* AraM, *M. maripaludis* GGGPS and *A. fulgidus* DGGGPS); (3) – six ether lipid enzymes (*E. coli* Idi, *E. coli* mutant IspA, *B. subtilis* AraM, *M. maripaludis* GGGPS, *A. fulgidus* DGGGPS and *A. fulgidus* CarS) and (4) – seven ether lipid enzymes (*E. coli* Idi, *E. coli* mutant IspA, *B. subtilis* AraM, *M. maripaludis* GGGPS, *A. fulgidus* DGGGPS, *A. fulgidus* CarS and *B. subtilis* PssA). When the total phospholipid composition of *E. coli* was compared for the different strains, no major differences were observed for the four differently engineered *E. coli* strains (Figure 6A). However, a detailed analysis showed the presence of a peak corresponding to AE (unsaturated) in *E. coli* strain 4 (Figure 6B, lane 4) which contains all the ether lipid genes required for synthesis compared to the other strains that lack the *B. subtilis* PssA (Figure 6B, lane 3) or both the *A. fulgidus* CarS and *B. subtilis* PssA (Figure 6B, lane 2) or the control (Figure 6B, lane 1). Another archaeal specific product digeranylgeranylglyceryl phosphoglycerol (DGGGP-Gro) (m/z = 789.5 [M-H]-) was detected in the engineered *E. coli* strains, which is absent in the control strain (Figure 6B, lane 1). This compound was also detected in a previous study [40] and measured as the unsaturated form of AG. This lipid was observed in the *E. coli* strains that contain the basic set of five ether lipid genes up to the genes required for the formation of DGGGP (Figure 6B, lane 2). However, an enhanced production of unsaturated AG was observed when the strain also contains the *A. fulgidus* CarS (Figure 6B, lane 3). Upon the introduction of the AS-producing *B. subtilis* PssA enzyme strain 4, the amount of AG decreased and instead, AE was produced (Figure 6B, lane 4). Formation of unsaturated AG has been attributed to an endogenous reaction in *E. coli* which directly attaches the glycerol moiety to DGGGP [40], as it was assumed that the *E. coli* PgsA and PgpA enzymes are not able to accept the archaeal intermediate as a
substrate. As shown in the previous section, these two enzymes do recognize CDP-archaeol and convert it into AG. Therefore, the presence of unsaturated AG in the *E. coli* strain expressing the ether lipid genes is due to the low, but significant activity of the endogenous CDP-DAG synthase towards DGGGP. To confirm this hypothesis, an *in vitro* reaction was performed where DGGGP was incubated in presence of the purified CDP-DAG synthase (*E. coli* CdsA) in the presence of CTP. Under those conditions, a significant fraction of the DGGGP was converted into CDP-archaeol (**Figure 6C, lane 2**) demonstrating that CdsA is active with the archaeal substrate. When the *E. coli* PgsA and PgpA and the substrate G3P were added to the reaction, the CDP-archaeol was converted into AG (**Figure 6C, lane 3**). Taken together, these data demonstrate that the low activity of the endogenous *E. coli* CdsA (*in vitro* less than 1% compared to CarS in our previous work [41]) towards DGGGP results in substantial AG production *in vivo*. It is concluded that with a limited set of genes the archaeal lipids AG and AE can be produced in *E. coli*.

**Discussion**

For several decades, *E. coli* has been used for metabolic engineering such as the improvement of isoprenoid or carotenoid production [95,154–156]. Several attempts have been made to reconstruct the archaeal ether lipid biosynthetic pathway into *E. coli*. Gunsalus *et al.* [21] demonstrated production of DGGGP upon the overexpression of the endogenous *E. coli* Idi and expression of four enzymes (G1PDH, GGPPS, GGGPS and DGGGPS) from the hyperthermophilic archaeon *Archaeoglobus fulgidus*. Likewise, Yokoi *et al.* [40] employed ether lipid genes from the mesophilic archaeon *Methanosarcina acetivorans* to produce DGGGP and another compound that was identified as the PG-type derivative of DGGGP, named DGGGP-Gro. Isobe *et al.* [53] in addition expressed the *M. acetivorans* geranylgeranyl reductase for double bond reduction in conjunction with the four aforementioned archaeal enzymes [40]. They observed the *in vivo* production of archaeal lipids with a fully saturated isoprenoid chain only when the cofactor *M. acetivorans* ferredoxin was co-expressed.
Figure 6 | Archaeal lipid production in *E. coli*. (A) Total lipid analysis of four different engineered *E. coli* strains containing a different combination of ether lipid enzymes. The lipids were extracted from the membrane fraction and analyzed by LC-MS. The total ion counts of the several PE and PG species and of the archaeal lipids were normalized using Eicosane (m/z = 281.55 [M-H]) as internal standard. The species as classified according to the number of carbon atoms and number of unsaturated bonds in the acyl chains at the sn-1 and sn-2 positions. (B) Archaeal lipids production by different strains of *E. coli* upon the expression of the ether lipid biosynthetic genes. Results are the average of three biological replicates ± SE. (C) *In vitro* reactions using DGGGP and the purified *E. coli* CdsA, PgsA and PgpA proteins to determine the ability of CdsA to convert DGGGP into CDP-archaeol. Total ion counts from LC-MS data were normalized using DDM as internal standard. The data are the average of three experiments ± SE.
In this study, we aimed to produce the two major archaeal-like lipids AE and AG in *E. coli* as these lipids have the same polar head group as the major phospholipids PE and PG present in the *E. coli* membrane. However, this required a further understanding of the enzymatic steps required for polar head group attachment. Importantly, we have previously shown the production of CDP-archaeol *in vitro* by also including the CarS into the pathway that yielded DGGGP. Expression of CarS provides a means to substantially increase the production of the endogenously produced DGGGP-Gro that corresponds to AG (Figure 6B). During polar head group attachment, the CDP group from the CDP-activated precursor is replaced by a different polar head group. In bacteria and in archaea, the reactions involved in this process are very similar and mediated by the enzymes belonging to the CDP-alcohol phosphatidyltransferase superfamily. Given the high sequence homology among archaeal and bacterial phosphatidyltransferase [7,8], we investigated the substrate promiscuity of the enzymes PgsA and PgpA that are involved in PG formation, and their ability to recognize CDP-archaeol. Using *in vitro* reactions, purified *E. coli* PgsA was able to produce AGP from CDP-archaeol which in turn was produced by the *A. fulgidus* CarS from DGGGP, CTP and G3P. By means of the *E. coli* phosphatase PgpA, the AGP was readily converted into AG. This demonstrates a high substrate promiscuity of these bacterial enzymes and alleviates the need to introduce archaeal enzymes into *E. coli* to perform these reactions. Indeed, when the ether lipid biosynthesis pathway up to the formation of CDP-archaeol is introduced into *E. coli*, AG formation is observed. However, CarS is not essential, as, even in its absence, some AG can be formed. The origin of this AG was previously unknown, but we now show that this is due to a low activity of the endogenous *E. coli* CdsA for DGGGP resulting in the formation of CDP-archaeol that is further converted by *E. coli* PgsA and PgpA into AG.

In contrast to AG formation, AE formation has not been observed previously in *E. coli* which must imply that the *E. coli* Psses are unable to recognize CDP-archaeol in order to produce the intermediate AS that by decarboxylation should be further converted into AE. Indeed, the purified *E. coli* PssA was inactive with this substrate. Structural analysis of the bacterial and archaeal members of CDP-alcohol phosphatidyltransferase
family indicated a high level of secondary structure conservation of these enzymes with eight transmembrane segments. The *B. subtilis* PssA is a truncated version of these enzymes, being significantly shorter but still containing the highly conserved protein core. Importantly, this enzyme was previously shown to be active with the archaeal-like substrates [47]. Indeed, the purified *B. subtilis* PssA catalyzed the production of AS from CDP-archaeol and L-serine. The next step, the decarboxylation of the serine moiety, seems less specific, as purified *E. coli* Psd mediated the decarboxylation of AS with the concomitant formation of AE. Thus, for the production of AE in *E. coli* cells, only the *B. subtilis* PssA needs to be introduced (Figure 6C).

Having established the exact sequence of reactions needed for the archaeal ether lipid biosynthetic pathway [41] including the polar head group attachment, the reconstitution of AE and AG formation in *vitro* could be realized. Using a combination of archaeal and bacterial enzymes, and a breakdown in individual reactions, the synthesis of the unsaturated archaeal lipids AE and AG from the initial building blocks IPP, FPP and DHAP was achieved with nine purified enzymes. This defined the conditions needed for the reconstitution of AE and AG formation in *vivo*, using *E. coli* as a host. Since three endogenous *E. coli* enzymes (Psd, PgsA and PgpA) recognize the archaeal precursors, the *in vivo* reconstitution depends on the expression of seven ether lipid genes. This includes overexpression of the *E. coli* Idi to boost IPP formation, and expression of a mutant IspA to generate GGPP; the *B. subtilis* AraM for G1P formation, and *B. subtilis* PssA for AS formation, the key ether lipid biosynthetic proteins, the *M. maripaludis* GGGPS, the *A. fulgidus* DGGGPS and the *A. fulgidus* CarS. Although the conversion reactions appear efficient *in vitro*, the amounts of AE and AG produced *in vivo* are still low and compared to the total *E. coli* lipid content they are likely less than 1%. However, the expression of CarS elevated the levels three-fold as compared to a previous report on AG formation [40].

The work described here represents a unique strategy to synthesize archaeal ether lipids in bacteria. Although the levels are still low, it is important to realize that the production of AE and AG as reported in this study, is performed in the presence of a fully functional phospholipid...
ether lipid biosynthesis. High-level AE and AG production likely also requires the up-regulation of the entire pathway leading to IPP production. In addition, the pathway resulting in the reduction of the isoprenoid chains needs to be introduced to produce the saturated archaeatidyl compounds. Future studies should be directed towards a gradual down-regulation of the endogenous ester-bonded phospholipid biosynthetic pathway so that the endogenous lipids can be replaced by ether lipids. Such a bacterial strain could be used as an experimental model to examine the impact of the ‘lipid divide’ on the physiology and robustness of bacteria.

Experimental procedures

Bacterial strain and cloning procedures

*Escherichia coli* and *Bacillus subtilis* genomic DNA was used as template for the amplification of genes encoding the bacterial enzymes. *E. coli* DH5α (Invitrogen) was used for cloning. The primers and the plasmids used in this study are listed in Table 1 and 2. *E. coli* BL21 (DE3) or Lemo21 (DE3) [122] was used as protein overexpression host strain and grown in aerobic condition at 37°C in LB medium supplemented with the required antibiotics, Kanamycin (50 μg/ml), chloramphenicol (34 μg/ml), streptomycin (50 μg/ml) and ampicillin (50 μg/ml) in conjunction with 0.2% glucose added when necessary.

Expressions and purification of ether lipid enzymes

The bacterial proteins *B. subtilis* PssA and *E. coli* Psd were expressed in *E. coli* BL21 strain and induced with 1 mM of IPTG. *E. coli* PssA was induced with 0.5 mM of IPTG in the same over-expression strain *E. coli* BL21, whereas *E. coli* PgsA and PgpA were expressed in *E. coli* Lemo strain and induced with 0.4 mM of IPTG and 0.5 mM of L-rhamnose. After 2.5 hours of induction the cytoplasmic and membrane fractions were separated as described in a previous study [41]. The inner membrane vesicles (IMVs) of *E. coli* expressing the membrane proteins (*B. subtilis* PssA and *E. coli* Psd, PgsA and PgpA) were isolated as previously described [117]. The *E. coli* IMVs harboring the *B. subtilis* PssA and *E. coli* Psd were resuspended in buffer A (50 mM Tris-HCl pH 7.5, 300 mM NaCl and 10% glycerol) and 0.5
mg/ml of IMVs were solubilized in 2% of DDM (n-dodecyl-β-D-maltopyranoside) detergent for 1 hour at 4 °C. A centrifugation (173,400 xg) step of 30 min at 4 °C removed the insolubilized materials and the supernatant was incubated with Ni-NTA (Ni²⁺-nitrilotriacetic acid) beads (Sigma) for 1 hour at 4 °C. The Ni-NTA beads were washed ten times with 40 column volumes (CV) of buffer B (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 10% glycerol and 0.2% DDM) supplemented with 20 mM imidazole, and the proteins were eluted three times with 0.5 CV of buffer B supplemented with 250 mM imidazole. The *E. coli* IMVs (1 mg/ml) containing the *E. coli* PgsA and PgpA were resuspended in buffer C (50 mM Tris-HCl pH 7.5, 150 mM NaCl and 10% glycerol). The solubilization steps were performed as above. The *E. coli* PgsA-bound beads were washed five times with 40 CV of buffer D (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% glycerol and 0.2% DDM) supplemented with 10 mM imidazole and eluted three times with 0.5 CV of buffer D supplemented with 250 mM imidazole. The Ni-NTA beads containing *E. coli* PgpA were washed ten times with 40 CV of buffer D supplemented with 10 mM imidazole and the protein was eluted with 0.5 CV of buffer D supplemented with 300 mM imidazole.

For the purification of the soluble protein *E. coli* PssA, the cytoplasmic fraction was incubated with Ni-NTA beads in buffer C overnight at 4 °C. The beads were washed three times with 40 CV of buffer C supplemented with 10 mM imidazole, once with 40 CV of buffer C supplemented with 60 mM Imidazole end eluted with 2 CV of buffer C supplemented with 300 mM imidazole. The purity of the eluted proteins were assessed on 12% SDS-PAGE gel stained with Coomassie Brilliant Blue and the protein concentration was determined by measuring the Absorbance at 280 nm. The other cytosolic proteins (*E. coli* isopentenyl diphosphate isomerase (Idi), *E. coli* mutant farnesyl diphosphate synthase (IspA), *B. subtilis* glycerol-1-phosphate dehydrogenase (AraM) and *Methanococcus maripaludis* GGGPS) and the membrane proteins (*Archaeoglobus fulgidus* DGGGPS, *A. fulgidus* CarS and *E. coli* CDP-diacylglycerol synthase (CdsA)) used in the present study were expressed and purified as described previously [41].
Ether lipid biosynthesis

**In vitro assays for archaeal lipids production**

*In vitro* reactions were performed in 100 μl of assay buffer containing a final concentration of 50 mM Tris-HCl pH7.5, 10 mM MgCl₂, 52.5 mM NaCl, 87.5 mM Imidazole, 0.07% DDM and 3.5% glycerol. Where specified, 100 μM synthetic DGGGP, 100 μM PA (phosphatidic acid) (C₁₈:₁), 2 mM CTP, 10 mM G3P, 10 mM L-serine, 20 mM EDTA and the indicated amount of purified enzymes were added to the reaction mixture. For the *in vitro* reconstitution of AE and AG, 100 μl of reaction volume was used containing the following assay buffer: 50 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 15 mM NaCl, 25 mM Imidazole, 0.02% DDM, 1% glycerol, 0.5 mM DTT, 0.1 mM farnesyl pyrophosphate (FPP), 0.1 mM IPP, 2 mM nicotinamide adenine dinclueotide (NADH) and 2 mM dihydroxyacetone phosphate (DHAP). Reactions were incubated at 37 °C for 1 hour as described previously [41] and the products were extracted two times with 0.3 ml of *n*-butanol. Extracted lipids were evaporated under a stream of nitrogen gas and resuspended in 50 μl of methanol for the LC-MS analysis.

**In vivo archaeal lipids synthesis**

Engineered *E. coli* strains were aerobically grown at 37 °C in 200 ml of LB medium supplemented with required antibiotics [kanamycin (25 μg/ml), chloramphenicol (17 μg/ml), streptomycin (25 μg/ml) and ampicillin (25 μg/ml), 0.2% of glucose and 1 mM NiCl₂]. The cells were induced with 0.25 mM IPTG and after 3.5 hours, the total membrane fractions were isolated as previously described [41]. The internal standard Eicosane (20 μM) was added to the total membrane fractions (8 mg/ml) and lipids were extracted by means of the Bligh and Dyer method [153]. The chloroform extractable lipid fraction was washed with the aqueous phase of a blank Bligh and Dyer extraction solution and evaporated under a stream of nitrogen gas. The evaporated samples were resuspended with 0.3 ml 1:2 chloroform:MeOH, evaporated under a stream of nitrogen and finally resuspended in 100 μl methanol for LC-MS analysis.
**Table 1. Expression vectors used in the present study**

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRSF-Duet-1</td>
<td>Cloning and expression vector (Kan(^{\text{R}})), T7 promoter</td>
<td>Novagen</td>
</tr>
<tr>
<td>pET-Duet-1</td>
<td>Cloning and expression vector (Amp(^{\text{R}})), T7 promoter</td>
<td>Novagen</td>
</tr>
<tr>
<td>pCDF-Duet-1</td>
<td>Cloning and expression vector (Str(^{\text{R}})), T7 promoter</td>
<td>Novagen</td>
</tr>
<tr>
<td>pACYC-Duet-1</td>
<td>Cloning and expression vector (Cm(^{\text{R}})), T7 promoter</td>
<td>Novagen</td>
</tr>
<tr>
<td>pSJ122</td>
<td>Synthetic gene encoding codon optimized DGGGP synthase from <em>A. fulgidus</em> with N-terminal His-tag and</td>
<td><em>Jain et al.</em> [41]</td>
</tr>
<tr>
<td></td>
<td>redesigned ribosome binding site AGGACGTTAACAT cloned into pRSF-Duet vector using the primers 41 and 42</td>
<td></td>
</tr>
<tr>
<td>pSJ135</td>
<td>PCR product of <em>ispA</em> gene with N-terminal His-tag from <em>E. coli K12</em> genomic DNA containing a double</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>mutation Y79H and S140T. PCR product of <em>idi</em> gene with His-tag at the N-terminal from <em>E. coli K12</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td>genomic DNA. Both genes were cloned into pCDF-Duet vector using the primers 62, 63, 24 and 57</td>
<td></td>
</tr>
<tr>
<td>pSJ138</td>
<td>Synthetic gene encoding codon optimized GGGP synthase from <em>M. maripaludis</em> with N-terminal His-tag,</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>PCR product of <em>araM</em> with C-terminus His-tag from <em>B. subtilis</em> genomic DNA. Both genes were cloned</td>
<td></td>
</tr>
<tr>
<td></td>
<td>into pET-Duet vector using the primers 70, 71, 11 and 12</td>
<td></td>
</tr>
<tr>
<td>pSJ140</td>
<td>Synthetic gene encoding codon optimized DGGGP synthase from <em>A. fulgidus</em> with N-terminal His-tag and</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>redesigned ribosome binding site AGGACGTTAACAT. Synthetic gene encoding codon optimized CDP-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>archaeol synthase from <em>A. fulgidus</em> with C-</td>
<td></td>
</tr>
</tbody>
</table>
Ether lipid biosynthesis

terminus His-tag. Both synthetic genes are cloned into pRSF-Duet vector using the primers 32, 20, 84 and 86

<table>
<thead>
<tr>
<th>Primers name</th>
<th>Primer sequence 5' —&gt; 3'</th>
<th>Restriction site</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>GCGCGAATTTCATGCATCACCACCACC</td>
<td>EcoRI</td>
</tr>
<tr>
<td>12</td>
<td>GCGCAAGCTTTTCATTTTTTGACAGC</td>
<td>HindIII</td>
</tr>
<tr>
<td>20</td>
<td>GCGCTCGAGGACAGGGTTCTGCCGACTGGAAG</td>
<td>XhoI</td>
</tr>
<tr>
<td>24</td>
<td>GATATACATGGCGAGCCATCCACATC</td>
<td>Ncol</td>
</tr>
<tr>
<td>32</td>
<td>GCGCCATATGCTGGATCTGATTCTGAA</td>
<td>Ndel</td>
</tr>
<tr>
<td>41</td>
<td>GACCAAGCTTTGCCGCATAATGC</td>
<td>HindIII</td>
</tr>
<tr>
<td>42</td>
<td>GATGCCTGAGTTAGAATGCACCGGCG</td>
<td>XhoI</td>
</tr>
<tr>
<td>57</td>
<td>GCGCGAATTTCATTTTTTAGGATGATGTAG</td>
<td>EcoRI</td>
</tr>
<tr>
<td>62</td>
<td>CACTCATTAAATCATGATGATTTCACCACGCAATGG</td>
<td>blunt</td>
</tr>
</tbody>
</table>

Table 2. Oligonucleotide primers used in the study.
LC-MS analysis

The samples from *in vitro* reactions were analyzed using an Accela1250 high-performance liquid chromatography system coupled with an ESI-MS Orbitrap Exactive (Thermo Fisher Scientific). A volume of 5 μl of each sample was used for the analysis. The LC-MS method parameters used in this study were the same as described previously [41]. The capillary and the tube lens voltage were set to -75 V and -190 V, respectively. For the samples from *in vivo* experiments, a sample volume of 5 μl was injected into a Shim-pack XR-ODS/C8/Phenyl column with dimension 3.0 mm I.D. x 75 mm (Shimadzu) operating at 55 °C with a flow rate of 400 μl/min. Mobile phase A [10 mM ammonium formate with 0.1% formic acid in water:acetonitrile (40:60 v/v)] and mobile phase B [(10 mM ammonium formate with 0.1% formic acid in acetonitrile:isopropanol (10:90 v/v)] were used as follows: initial condition started with 40% B, a linear gradient was started in 2 min from 40% to 43% B, gradient from 43% to
Ether lipid biosynthesis

50% B in 0.1 min, 54% B isocratic for the next 9.9 min, linear gradient from 54% to 70% B in 0.1 min, 99% B isocratic for the following 5.9 min, gradient from 99% to 40% B in 0.1 min and 40% B isocratic for the last 1.9 min [157]. The MS settings used for this analysis were the same as described above. The Thermo XCalibur processing software was used for the data analysis and Genesis algorithm for automated peak detection and integration was applied to this analysis.

Author Contribution
Antonella Caforio, Samta Jain and Arnold Driessen conceived and designed the research. Antonella Caforio cloned the genes, purified the enzymes and performed the experiments. Melvin Siliakus assisted in the cloning of genes and John van der Oost coordinated the pathway design. Adriaan Minnard design the DGGGP synthesis, which was performed by Peter Fodran. The manuscript was written by the contribution of all the authors.

Acknowledgments
This project was carried out within the research program of the biobased ecologically balanced sustainable industrial chemistry (BE-BASIC). We thanks Oleksander Salo for technical assistance and Juke Lolkema for the assistance with the hydrophobicity analysis.
### Supplementary Information

| Prok A. coli | 1 | EF- | | | |
| Prok B. subtilis | 1 | EF- | | | |
| Prok C. pasteurianum | 1 | EF- | | | |
| Prok C. saccharolyticum | 1 | EF- | | | |
| Prok S. coelicolor | 1 | EF- | | | |
| Prok S. pneumoniae | 1 | EF- | | | |
| Prok N. formicicum | 1 | EF- | | | |
| Prok G. amamii | 1 | EF- | | | |
| Prok T. carboxydivorans | 1 | EF- | | | |
| Prok F. succinogenes | 1 | EF- | | | |
| Prok S. venezuelae | 1 | EF- | | | |
| Prok B. subtilis | 1 | EF- | | | |
| Prok F. succinogenes | 1 | EF- | | | |
| Prok N. formicicum | 1 | EF- | | | |
| Prok S. albus | 1 | EF- | | | |
| Prok S. solfataricus | 1 | EF- | | | |
| Prok M. amylolytica | 1 | EF- | | | |
| Prok S. coelicolor | 30 | EF- | | | |
| Prok S. pneumoniae | 5 | EF- | | | |
| Prok N. formicicum | 5 | EF- | | | |
| Prok G. amamii | 13 | EF- | | | |
| Prok N. stamnogena | 14 | EF- | | | |
| Prok T. carboxydivorans | 23 | EF- | | | |
| Prok F. succinogenes | 24 | EF- | | | |
| Prok S. venezuelae | 18 | EF- | | | |
| Prok N. mediterranei | 55 | EF- | | | |
| Prok B. subtilis | 6 | EF- | | | |
| Prok C. tetani | 6 | EF- | | | |
| Prok S. succinogenum | 18 | EF- | | | |
| Prok N. albus | 13 | EF- | | | |
| Prok A. baumannii | 8 | EF- | | | |
| Prok M. amylolytica | 6 | EF- | | | |

<p>| Prok K. hispanicus | 7 | PLAsT- | | | |
| Prok N. solfataricus | 15 | PLAsT- | | | |
| Prok M. amylolytica | 7 | PLAsT- | | | |
| Prok K. hispanicus | 4 | EF- | | | |
| Prok B. subtilis | 2 | EF- | | | |
| Prok C. thermodenitrificans | 4 | EF- | | | |
| Prok C. carbonum | 12 | EF- | | | |
| Prok C. saccharolyticum | 2 | EF- | | | |
| Prok S. coelicolor | 59 | EF- | | | |
| Prok S. pneumoniae | 5 | EF- | | | |
| Prok N. formicicum | 5 | EF- | | | |
| Prok G. amamii | 21 | EF- | | | |
| Prok N. stamnogena | 19 | EF- | | | |
| Prok T. carboxydivorans | 16 | EF- | | | |
| Prok F. succinogenes | 22 | EF- | | | |
| Prok S. venezuelae | 19 | EF- | | | |
| Prok N. mediterranei | 93 | EF- | | | |
| Prok B. subtilis | 4 | EF- | | | |
| Prok C. tetani | 6 | EF- | | | |
| Prok S. succinogenum | 18 | EF- | | | |
| Prok N. albus | 13 | EF- | | | |
| Prok A. baumannii | 8 | EF- | | | |
| Prok M. amylolytica | 6 | EF- | | | |
| Prok N. hispanicus | 7 | EF- | | | |
| Prok S. solfataricus | 21 | EF- | | | |
| Prok M. amylolytica | 7 | EF- | | | |
| Prok K. hispanicus | 27 | EF- | | | |
| Prok B. subtilis | 30 | EF- | | | |
| Prok C. thermodenitrificans | 29 | EF- | | | |
| Prok C. saccharolyticum | 25 | EF- | | | |
| Prok S. coelicolor | 57 | EF- | | | |
| Prok S. pneumoniae | 28 | EF- | | | |
| Prok N. formicicum | 20 | EF- | | | |
| Prok G. amamii | 46 | EF- | | | |
| Prok N. stamnogena | 44 | EF- | | | |
| Prok T. carboxydivorans | 60 | EF- | | | |
| Prok F. succinogenes | 58 | EF- | | | |
| Prok S. venezuelae | 49 | EF- | | | |
| Prok N. albus | 124 | EF- | | | |
| Prok B. subtilis | 23 | EF- | | | |</p>
<table>
<thead>
<tr>
<th><strong>Ether lipid biosynthesis</strong></th>
</tr>
</thead>
</table>

| **Phas C. tetani** | 27 | **---** | **---** |
| **Phas S. aureofaciens** | 45 | **---** | **---** |
| **Phas D. oleovorans** | 37 | **---** | **---** |
| **Phas A. baumannii** | 32 | **---** | **---** |
| **Phas N. phosphatidicum** | 31 | **---** | **---** |
| **Phas R. hispanica** | 31 | **---** | **---** |
| **Phas S. solfataricus** | 46 | **---** | **---** |
| **Phas R. amylovorica** | 31 | **---** | **---** |

| **Phas E. coli** | 65 | **---** | **---** |
| **Phas R. subtilis** | 76 | **---** | **---** |
| **Phas R. parasuis** | 65 | **---** | **---** |
| **Phas C. thermophillus** | 74 | **---** | **---** |
| **Phas C. saccharobutylicum** | 64 | **---** | **---** |
| **Phas S. coelicolor** | 138 | **---** | **---** |
| **Phas S. pneumoniae** | 67 | **---** | **---** |
| **Phas N. formicium** | 94 | **---** | **---** |
| **Phas G. anhangari** | 83 | **---** | **---** |
| **Phas M. stadtmanae** | 81 | **---** | **---** |
| **Phas T. carboxydutrophus** | 98 | **---** | **---** |
| **Phas R. marilli** | 98 | **---** | **---** |
| **Phas S. islandicus** | 86 | **---** | **---** |
| **Phas R. mediterranei** | 172 | **---** | **---** |
| **Phas R. subtilis** | 61 | **---** | **---** |
| **Phas C. tetani** | 65 | **---** | **---** |
| **Phas S. aureofaciens** | 95 | **---** | **---** |
| **Phas D. oleovorans** | 75 | **---** | **---** |
| **Phas A. baumannii** | 70 | **---** | **---** |
| **Phas M. formicium** | 68 | **---** | **---** |
| **Phas R. hispanica** | 66 | **---** | **---** |
| **Phas S. solfataricus** | 63 | **---** | **---** |
| **Phas R. amylovorica** | 66 | **---** | **---** |

| **Phas E. coli** | 104 | **---** | **---** |
| **Phas R. subtilis** | 111 | **---** | **---** |
| **Phas R. parasuis** | 104 | **---** | **---** |
| **Phas C. thermophillus** | 115 | **---** | **---** |
| **Phas C. saccharobutylicum** | 101 | **---** | **---** |
| **Phas S. coelicolor** | 176 | **---** | **---** |
| **Phas S. pneumoniae** | 104 | **---** | **---** |
| **Phas M. formicium** | 104 | **---** | **---** |
| **Phas G. anhangari** | 119 | **---** | **---** |
| **Phas M. stadtmanae** | 119 | **---** | **---** |

| **Phas T. carboxydutrophus** | 135 | **---** | **---** |
| **Phas F. fumarii** | 143 | **---** | **---** |
| **Phas S. islandicus** | 123 | **---** | **---** |
| **Phas R. mediterranei** | 209 | **---** | **---** |
| **Phas R. subtilis** | 99 | **---** | **---** |
| **Phas C. tetani** | 104 | **---** | **---** |
| **Phas D. oleovorans** | 113 | **---** | **---** |
| **Phas A. baumannii** | 105 | **---** | **---** |
| **Phas H. phosphatidicum** | 107 | **---** | **---** |
| **Phas H. hispanica** | 114 | **---** | **---** |
| **Phas S. solfataricus** | 114 | **---** | **---** |
| **Phas R. amylovorica** | 114 | **---** | **---** |

| **Phas S. coli** | 147 | **---** | **---** |
| **Phas R. subtilis** | 155 | **---** | **---** |
| **Phas A. parasuis** | 147 | **---** | **---** |
| **Phas C. thermophillus** | 158 | **---** | **---** |
| **Phas C. saccharobutylicum** | 147 | **---** | **---** |
| **Phas S. coelicolor** | 215 | **---** | **---** |
| **Phas S. pneumoniae** | 145 | **---** | **---** |
| **Phas M. formicium** | 135 | **---** | **---** |
| **Phas G. anhangari** | 158 | **---** | **---** |
| **Phas M. stadtmanae** | 154 | **---** | **---** |
| **Phas T. carboxydutrophus** | 180 | **---** | **---** |
| **Phas R. fumarii** | 189 | **---** | **---** |
| **Phas S. islandicus** | 167 | **---** | **---** |
| **Phas R. mediterranei** | 255 | **---** | **---** |
| **Phas R. subtilis** | 139 | **---** | **---** |
| **Phas C. tetani** | 145 | **---** | **---** |
| **Phas D. oleovorans** | 159 | **---** | **---** |
| **Phas A. baumannii** | 152 | **---** | **---** |
| **Phas M. formicium** | 154 | **---** | **---** |
| **Phas H. hispanica** | 158 | **---** | **---** |
| **Phas S. solfataricus** | 164 | **---** | **---** |
| **Phas R. amylovorica** | 156 | **---** | **---** |
| **Phas S. coli** | 173 | **---** | **---** |
| **Phas R. subtilis** | 183 | **---** | **---** |
| **Phas A. parasuis** | 173 | **---** | **---** |
| **Phas C. thermophillus** | 185 | **---** | **---** |
| **Phas C. saccharobutylicum** | 186 | **---** | **---** |
| **Phas S. coelicolor** | 241 | **---** | **---** |
Multiple sequence alignment of archaeal and bacterial PgsA and PssA. The multiple sequence alignment indicates the presence of a conserved domain typical of the CDP-alcohol phosphatidyltransferase superfamily. Species list: Bacteria: Escherichia coli, Bacillus subtilis, Haemophilus parasuis, Chloracidobacterium therophilum, Clostridium saccharobutylicum, Streptomyces coelolor, Streptococcus pneumonia, Clostridium tetani, Streptomyces aureofaciens, Desulfococcus oleovorans, and Acinetobacter baumannii; Archaea: Methanobacterium formicicum, Geoglobus ahangari, Methanosphaera stadtmannae, Thermofilum carboxyditrophus, Pyrolobus fumarii, Sulfolobus islandicus, Haloarcula mediterranei, Methanobacterium formicicum, Haloarcula hispanica, Sulfolobus solfataricus, and Haloarcula amylolytica.
Ether lipid biosynthesis

A

E. coli 1
B. mycoides 1
V. cholerae 1
D. acetiphilus 1
H. borinquense 1
N. gari 1
uncultured marine 1

E. coli 50
B. mycoides 61
V. cholerae 43
D. acetiphilus 34
H. borinquense 26
N. gari 36
uncultured marine 39

E. coli 104
B. mycoides 115
V. cholerae 97
D. acetiphilus 88
H. borinquense 86
N. gari 96
uncultured marine 92

E. coli 154
B. mycoides 168
V. cholerae 152
D. acetiphilus 136
H. borinquense 145
N. gari 155
uncultured marine 145

B

E. coli 1
S. sanguinis 1
D. gobiensis 1
A. sulfaticalidus 1
N. aegyptia 1
N. mazei 1
N. occitans 1

E. coli 32
S. sanguinis 45
D. gobiensis 50
A. sulfaticalidus 57
N. aegyptia 31
M. mazeli 9
N. occultans 32

E. coli 70
S. sanguinis 83
D. gobiensis 88
A. sulfaticalidus 91
N. aegyptia 66
M. mazeli 36
N. occultans 85

E. coli 95
S. sanguinis 103
D. gobiensis 108
A. sulfaticalidus 111
Figure S2] Multiple sequence alignment of a diverse group of bacterial and archaeal PgpA and PgpB proteins. (A) Alignment of bacterial and archaeal PgpA protein sequences showing a high degree of conservation. Species list: Bacteria: *Escherichia coli*, *Bacillus mycoides*, *Vibrio cholera*, and *Denitrovibrio acetiphilus*; Archaea: *Halogeometricum borinquense*, *Natrinema gari*, and an uncultured marine archaeon (Euryarchaeota). (B) Alignment of bacterial and archaeal PgpB protein sequences. Species list: Bacteria: *Escherichia coli*, *Streptococcus sanguinis*, and *Deinococcus gobiensis*; Archaea: *Archaeoglobus sulfaticallidus*, *Natrialba aegyptia*, *Methanosarcina mazi*, and *Natronococcus occultus*. 

<table>
<thead>
<tr>
<th>Specie</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>N. aegyptia</em></td>
<td>91</td>
</tr>
<tr>
<td><em>M. mazi</em></td>
<td>56</td>
</tr>
<tr>
<td><em>N. occultus</em></td>
<td>184</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>126</td>
</tr>
<tr>
<td><em>S. sanguinis</em></td>
<td>117</td>
</tr>
<tr>
<td><em>D. gobiensis</em></td>
<td>122</td>
</tr>
<tr>
<td><em>A. sulfaticallidus</em></td>
<td>129</td>
</tr>
<tr>
<td><em>N. aegyptia</em></td>
<td>106</td>
</tr>
<tr>
<td><em>M. mazi</em></td>
<td>70</td>
</tr>
<tr>
<td><em>N. occultus</em></td>
<td>184</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>183</td>
</tr>
<tr>
<td><em>S. sanguinis</em></td>
<td>152</td>
</tr>
<tr>
<td><em>D. gobiensis</em></td>
<td>156</td>
</tr>
<tr>
<td><em>A. sulfaticallidus</em></td>
<td>159</td>
</tr>
<tr>
<td><em>N. aegyptia</em></td>
<td>140</td>
</tr>
<tr>
<td><em>M. mazi</em></td>
<td>99</td>
</tr>
<tr>
<td><em>N. occultus</em></td>
<td>218</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>235</td>
</tr>
<tr>
<td><em>S. sanguinis</em></td>
<td>206</td>
</tr>
<tr>
<td><em>D. gobiensis</em></td>
<td>208</td>
</tr>
<tr>
<td><em>A. sulfaticallidus</em></td>
<td>209</td>
</tr>
<tr>
<td><em>N. aegyptia</em></td>
<td>190</td>
</tr>
<tr>
<td><em>M. mazi</em></td>
<td>151</td>
</tr>
<tr>
<td><em>N. occultus</em></td>
<td>270</td>
</tr>
<tr>
<td>Ether lipid biosynthesis</td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td></td>
</tr>
</tbody>
</table>
| **E. coli** | 1 | \text{GNSPKK-S-}
| **Streptomyces sp.** | 1 | \text{-QYIPKL-WLTRLQGASKRA-GLT-} |
| **B. flexus** | 1 | \text{-KEDK-}
| **N. exilis** | 1 | \text{-HCQAXH-1RXIR-AEDFQGK-} |
| **M. siciliae** | 1 | \text{-N32-3FVYIENGHHK-} |
| **M. paludis** | 1 | \text{-JVEK-7-} |
| **Salinarumae sp.** | 1 | \text{-MVAP-0-} |
| **E. coli** | 35 | 
| **Streptomyces sp.** | 27 | \text{-MYDAAATANVVPYPADTPKELCDKDNCGQGKRLGQTCDDFREYMMHFPQKENYEYFS-} |
| **B. flexus** | 17 | \text{-KHEIQ-} |
| **N. exilis** | 58 | \text{-IVYVNLIMLDEUERKENHVTFI-} |
| **M. siciliae** | 32 | \text{-KVKAHKVNLQIIPIDTFDDQVMKNGATIEDLGNFQAWYAMNRTL-} |
| **M. paludis** | 12 | \text{-} |
| **Salinarumae sp.** | 7 | \text{-} |
| **E. coli** | 35 | \text{-} |
| **Streptomyces sp.** | 84 | \text{-WHINYRNAAYMVFVTCGPGKVLSDQTHLQOMQMDERPNSH-} |
| **B. flexus** | 30 | \text{-IS-} |
| **N. exilis** | 95 | \text{-PQVH-} |
| **M. siciliae** | 89 | \text{-WLYMKNEAGLEFVSTDPQMNLMFYYVEILDQQRHDSFA-} |
| **M. paludis** | 12 | \text{-} |
| **Salinarumae sp.** | 7 | \text{-} |
| **E. coli** | 44 | \text{-YIK-} |
| **Streptomyces sp.** | 134 | \text{-VSK-AGCR-} |
| **B. flexus** | 141 | \text{-QG-} |
| **N. exilis** | 145 | \text{-DESPFRSVLHTENGWOCASSYNYK-} |
| **M. siciliae** | 136 | \text{-} |
| **M. paludis** | 22 | \text{-} |
| **Salinarumae sp.** | 22 | \text{-} |
| **E. coli** | 77 | \text{-} |
| **Streptomyces sp.** | 154 | \text{-} |
| **B. flexus** | 73 | \text{-} |
| **N. exilis** | 205 | \text{-} |
| **M. siciliae** | 168 | \text{-} |
| **M. paludis** | 49 | \text{-} |
| **Salinarumae sp.** | 49 | \text{-} |
| **E. coli** | 134 | \text{-} |
| **Streptomyces sp.** | 223 | \text{-} |
| **B. flexus** | 130 | \text{-} |
| **N. exilis** | 264 | \text{-} |
| **M. siciliae** | 227 | \text{-} |
| **M. paludis** | 75 | \text{-} |
| **Salinarumae sp.** | 75 | \text{-} |
Figure S3. Multiple sequence alignment of a selection of diverse archaeal and bacterial Psd proteins. Species list: Bacteria: Escherichia coli, Streptomyces sp., Bacillus flexus, and Nocardia exalbida; Archaea: two sequences from Methanosarcina siciliae, Methanobacterium paludis and a Salinarchaeum sp.