FEMS Microbes, 2, 2021, xtab012

https://doi.org/10.1093/femsmc/xtab012
Advance Access Publication Date: 24 August 2021
Research Article

RESEARCH ARTICLE – Microbes & Metabolism

Archaeal pseudomurein and bacterial murein cell wall biosynthesis share a common evolutionary ancestry

Bishwa P. Subedi[1,2,†], William F. Martin[3], Vincenzo Carbone[1], Eduardus C. Duin[4], Bryan Cronin[4], Julia Sauter[1], Linley R. Schofield[1], Andrew J. Sutherland-Smith[2] and Ron S. Ronimus[1,*]

[1]AgResearch Ltd. Grasslands, Tennent Drive, Palmerston North 4442, New Zealand, [2]Massey University, Tennent Drive, Palmerston North 4442, New Zealand, [3]Institute for Molecular Evolution, Heinrich-Heine University, Düsseldorf Universitätsstraße 1, D-40225, Germany and [4]Department of Chemistry and Biochemistry, Auburn University, Auburn, AL 36849, USA

*Corresponding author: AgResearch Ltd. Grasslands, Tennent Drive, Palmerston North 4442, New Zealand. E-mail: ron.ronimus@agresearch.co.nz

[†]Present address: Monash Institute of Pharmaceutical Sciences, Monash University, Victoria 3052, Australia.

One sentence summary: Gene cluster and phylogenetic analyses of pseudomurein-containing methanogens, and a structure for a pseudomurein peptide ligase, suggest that murein and pseudomurein biosynthetic pathways share a common evolutionary history.

[‡]Bishwa P. Subedi, https://orcid.org/0000-0001-5297-0867
[#]Andrew J. Sutherland-Smith, https://orcid.org/0000-0001-9485-1428
[$]Ron S. Ronimus, https://orcid.org/0000-0001-8669-5840

ABSTRACT

Bacteria near-universally contain a cell wall sacculus of murein (peptidoglycan), the synthesis of which has been intensively studied for over 50 years. In striking contrast, archaeal species possess a variety of other cell wall types, none of them closely resembling murein. Interestingly though, one type of archaeal cell wall termed pseudomurein found in the methanogen orders Methanobacteriales and Methanopyrales is a structural analogue of murein in that it contains a glycan backbone that is cross-linked by a L-amino acid peptide. Here, we present taxonomic distribution, gene cluster and phylogenetic analyses that confirm orthologues of 13 bacterial murein biosynthesis enzymes in pseudomurein-containing methanogens, most of which are distantly related to their bacterial counterparts. We also present the first structure of an archaeal pseudomurein peptide ligase from *Methanothermus fervidus* DSM1088 (Mfer336) to a resolution of 2.5 Å and show that it possesses a similar overall tertiary three domain structure to bacterial MurC and MurD type murein peptide ligases. Taken together the data strongly indicate that murein and pseudomurein biosynthetic pathways share a common evolutionary history.

Keywords: pseudomurein; murein; murein peptide ligase; cell wall; evolution; *Methanothermus fervidus*

INTRODUCTION

The development of cell walls was a vital step in the early evolution of life approximately 4 billion years ago and is one of the main features distinguishing archaea from bacteria (Kandler and König 1993; Kandler 1993; Martin *et al.* 2008; Albers and Meyer 2011; Tashiro *et al.* 2017). The most important functions of cell walls are to maintain cell shape and protect against osmotic pressure; they can also provide attachment sites for cell proteins

Received: 27 March 2021; **Accepted:** 19 August 2021

and biomolecules (Boucher *et al.* 2003; Bouhss *et al.* 2008; Dramsi *et al.* 2008; Egan *et al.* 2017).

The biosynthetic pathway of bacterial murein (peptidoglycan; Fig. 1) has been studied for over 50 years and is the target of numerous important antibiotics (Mandelstam and Rogers 1959; Schneider and Sahl 2010; Bugg *et al.* 2011). The biosynthesis of the murein pentapeptide monomer, starting with GlmS (glutamine-fructose-6-phosphate transaminase (isomerizing; EC 2.6.1.16) to its export, requires ~13 enzymes, dependent on species (Bouhss *et al.* 2008). The murein glycan backbone contains alternating *N*-acetyl-D-glucosamine (NAG) and *N*-acetyl-D-muramic acid (NAM) residues linked by β(1–4) bonds. Bacterial murein biosynthesis occurs in three phases, intracellular, membrane-associated and final extracellular cross-linking (Scheffers and Pinho 2005). In brief, enzymes GlmS, GlmM and bifunctional GlmU (Table S1, Supporting Information) synthesize UDP-NAG which is converted to UDP-NAM by MurA and MurB (Bugg *et al.* 2011). The cross-linking peptide is then added through sequential amino acid addition by peptide ligases MurC, MurD, MurE and MurF. The NAM-pentapeptide moiety is subsequently transferred to the membrane undecaprenyl-C_{55} carrier (UppP) by the integral membrane protein MraY to form Lipid I, and a NAG moiety is added by MurG to form Lipid II. A flippase (MurJ; Ruiz 2015) enables the extracellular export of the cell wall monomer with transglycosylases and transpeptidases (penicillin-binding proteins; PBPs) ultimately producing mature murein (Mohammadi *et al.* 2011; Caveney, Li and Strynadka 2018; Kumar *et al.* 2019). The composition of murein can show differences in the peptide cross-link, the main one being catalyzed by MurE which adds either L-Lys in Gram-positive bacteria or *meso*-diaminopimelic acid (*m*-DAP) in the case of Gram-negative bacteria (Barreteau *et al.* 2008).

From an evolutionary perspective, it is significant that no archaea possess a murein cell wall, but instead possess a variety of different cell wall types including sulfated-heteropolysaccharides, methanochondroitin, glutaminylglycan, proteinaceous sheaths, halomucin, pseudomurein and glycoprotein surface layers (e.g. S-layers; Kandler and König 1993; Albers and Meyer 2011), with the latter being the most common (Claus and König 2010; Klingl, Pickl and Flechsler 2019). Due to their broad taxonomic distribution in archaea and the lack of complexity, S-layers have been suggested to have been the first archaeal cell wall to develop (Albers and Meyer 2011). In some genera such as *Sulfolobus* the S-layer can be the only cell wall component present (Albers and Meyer 2011; Gambelli *et al.* 2019). By contrast, bacteria only rarely lost their murein-based cell wall (and did not invent S-layers as substitutes) (Albers and Meyer 2011; Meyer and Albers 2014).

Pseudomurein of archaea is a structural analogue of murein, comprising a peptide cross-linked glycan backbone, but the chemistry shows major differences (Fig. 1; Formanek 1985). These include the presence of a methanogen-specific aminosugar *N*-acetyl-L-talosaminuronic acid (NAT), β(1–3) bonds in the glycan backbone and isopeptide bonds (ε and γ) and use of only L-amino acids in the peptide stem (Fig. 1; Hartmann and König 1990, 1994; König, Hartmann and Kärcher 1993). Pseudomurein peptide biosynthesis has been proposed to proceed through the biosynthesis of the unusual UDP-N^{α}-glutamyl-γ-phosphate precursor that is formed in three steps, not through UDP-*N*-acetylmuramic acid as in bacteria (Hartmann and König 1994; König, Hartmann and Kärcher 1993; Kandler and König 1998; Leahy *et al.* 2010). Analogous to murein biosynthesis, pseudomurein peptide stem amino acids are ligated to the UDP-N^{α}-glutamyl-γ-phosphate in a series of ATP-dependent

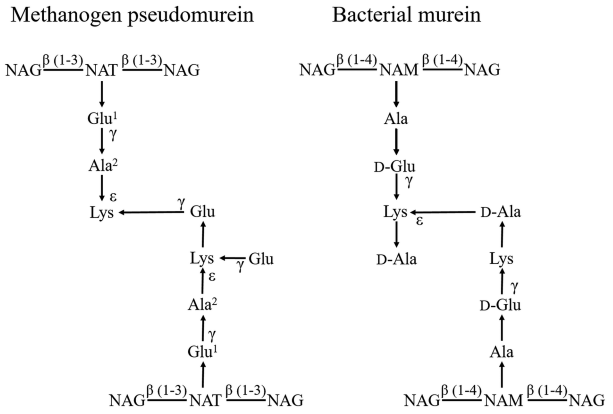

Figure 1. Comparison of the structure of methanogen pseudomurein and bacterial murein. NAG, *N*-acetylglucosamine; NAT, *N*-acetyltalosaminuronic acid; NAM, *N*-acetylmuramic acid. Methanogen pseudomurein contains; L-amino acids, β (Kandler and König 1993; Kandler 1993; Tashiro *et al.* 2017) bonds and *N*-acetyltalosaminuronic acid in the glycan chain, and several unusual isopeptide bonds. 1, Glu can be replaced by Asp; 2, Ala can be replaced by Thr or Ser, in some strains (Kandler and König 1993). Bacterial murein contains; D- and L-amino acids, and β (Kandler and König 1993; Kandler 1993; Albers and Meyer 2011; Tashiro *et al.* 2017) bonds and *N*-acetylmuramic acid in the glycan chain. The figure was adapted and modified from Schofield *et al.* (Schofield *et al.* 2015).

reactions to produce UDP-L-Glu-γ-L-Ala-ε-L-Lys-L-Ala with an additional glutamic acid residue attached to the L-Lys residue via a γ bond (Hartmann and König 1994). During this process, the growing pseudomurein UDP-peptide becomes linked to the glycan backbone at an early cytoplasmic biosynthesis step whereas in bacteria this occurs at the membrane in a reaction catalyzed by MurG. It is then exported to outside the cell for final cross-linking involving analogues of MraY, UppP, UppS, flippase and transglutaminase (Leahy *et al.* 2010). Pseudomurein has a very limited phylogenetic distribution, being only found in some methanogenic archaea represented by the orders Methanobacteriales and Methanopyrales (Hartmann and König 1990, 1994; König, Hartmann and Kärcher 1993; Albers and Meyer 2011).

Owing to the differences between pseudomurein and murein biosynthesis and chemistry, it has been proposed that the evolution of the two cell wall biosynthesis pathways occurred independently of each other (Kandler and König 1993; Albers and Meyer 2011; Errington 2013). However, four murein peptide ligase homologues were identified in the first methanogen genomes sequenced for pseudomurein-containing *Methanothermobacter thermautotrophicus* ΔH and *Methanopyrus kandleri* (Smith *et al.* 1997; Slesarev *et al.* 2002). Similarly, orthologues of most murein biosynthesis proteins were also identified, including key proteins such as MurG and an *O*-antigen teichoic acid transporter-like protein. Leahy *et al.* (2010) were the first to specifically focus on the proteins involved in pseudomurein biosynthesis and proposed a pathway mainly based on the biochemical pathway proposed by Hartmann *et al.* (Hartmann and König 1994; König, Hartmann and Kärcher 1993; Kandler and König 1998), including the possible use of transglutaminases to aid cross-linking of pseudomurein. In the latter paper, no attempt was made to assign individual Mur ligase family proteins to a particular ligase step or class, examine gene clustering patterns or perform phylogenetic analyses which could help unravel the evolution of pseudomurein synthesis. Methanogenesis is often considered to represent one of life's most ancient metabolisms (Decker, Jungermann and Thauer 1970; Ueno *et al.* 2006; Wong *et al.* 2007, 2016; Martin *et al.* 2008; Di Giulio 2009; Kelly, Wickstead and

Gull 2011; Petitjean *et al.* 2015; Weiss *et al.* 2018; Borrel *et al.* 2019). While methanogens were indicated to be contained solely within the Euryarchaea (Bapteste, Brochier and Boucher 2005; Evans *et al.* 2015; Borrel *et al.* 2019), recent studies have identified novel archaea that possess traits of methanogenic metabolism in non-Euryarchaeal phyla including Bathyarchaeota and the Verstraetearchaeota within the TACK superphylum (Evans *et al.* 2015; Castelle and Banfield 2018; Berghuis *et al.* 2019). This broader taxonomic distribution supports an early origin for methanogenesis in the Archaea (Borrel *et al.* 2019). Elucidating the evolutionary interconnections between pseudomurein and murein biosynthesis has the potential to provide insights into one of life's most critical early cellular achievements, the formation of cell walls.

In order to characterize the potential evolutionary relationships between the origins of pseudomurein and bacterial murein biosynthesis pathways we have built upon the initial findings of Leahy *et al.* (2010) by conducting further protein sequence searches using bacterial murein and putative archaeal pseudomurein biosynthesis enzymes, and performing gene clustering, taxonomic distribution and phylogenetic analyses on the basis of new structures and more thorough phylogenomic sampling. One key element to understanding how the evolution of the two cell wall types occurred is the origin of the archaeal peptide ligases which typically contain approximately 13–20% amino acid sequence identity over the full-length of the proteins to their respective bacterial orthologues. One confounding factor in the sequence analysis of peptide ligases is that the bacterial ligases fall into two subclasses due to major differences in their N-terminal domains (MurC/D and MurE/F; Smith 2006). Since enzyme structure is more conserved than amino acid sequence (Chothia and Lesk 1986; Illergård, Ardell and Elofsson 2009), we pursued the structural determination of an archaeal peptide ligase. We present here the first structure of a pseudomurein peptide ligase (Mfer336) from the hyperthermophilic *Methanothermus fervidus*, which reveals that it shares a similar overall 3-dimensional structure to bacterial MurC/D peptide ligases. The new taxonomic distribution, phylogenetic, gene clustering analyses and structural data provide important insights into the biosynthesis of pseudomurein and the evolutionary origins of cell walls in bacteria and archaea.

EXPERIMENTAL PROCEDURES

Sequence and phylogenetic analyses

Protein sequence searches used BLASTP and PSI-BLASTP with the non-redundant database at NCBI or targeted particular microbial genomes (Altschul *et al.* 1997). Firmicute sequences such as from *Bacillus subtilis subsp. subtilis str.* 168 (Lake and Sinsheimer 2013) (the original sequences used for the searches are shown in Table S2, Supporting Information), were used for initial PSI-BLASTP (0.005 threshold) searches in Archaea (iterated until no further sequences were identified). Once archaeal sequences were identified, these were also used to broaden the searches (summarized in Tables S1 and S2, Supporting Information). Representative peptide ligase sequences for a wide phylogenetic survey of bacteria and pseudomurein-containing methanogens from KEGG were used for construction of the phylogenetic trees using MEGA7 (Kumar, Stecher and Tamura 2016). Amino acid alignments were produced using Muscle within MEGA7 and trees generated using the Maximum Likelihood, Neighbor-Joining, Minimum Evolution and Maximum Parsimony methods. Bootstrapping (100 iterations) was used to examine the robustness of the branching patterns.

Cloning and expression of Mfer336 pseudomurein peptide ligase

The *M. fervidus* (DSM 1088) gene *Mfer336* was codon-optimized for expression in *Escherichia coli* using the vector pET15b (Genscript, Singapore). Expression and enzyme purification were performed essentially as described by Zhang *et al.* (2017) and utilized nickel nitrilotriacetic acid (Ni^{2+}-NTA) chromatography (Jena Bioscience, Germany). The final buffer for performing crystallization experiments was 20 mM MOPS (3-(*N*-morpholino)propanesulfonic acid), pH 7.0, 500 mM KCl and 2 mM TCEP (tris(2-carboxyethyl)phosphine).

Crystallization, X-ray diffraction and data collection

Crystallization experiments were performed using freshly prepared Ni^{2+}-NTA purified Mfer336 protein with sparse-matrix or grid-screening plates using the sitting drop method. 75 μL of crystal screening condition was pipetted into plate reservoirs, 1.0 μL of purified Mfer336 was placed into the crystallization well to which 1.0 μL of reservoir solution was added. Crystal condition, 0.2 M lithium sulfate, 0.1 M BIS-Tris, pH 5.5, 25% w/v PEG3350 (JCSG-plus screen; Molecular Dimensions, UK) produced thin-plate crystals. The crystal condition was optimized using Silver Bullets crystallization additive screen (Hampton Research, Aliso Viejo, CA, USA) in a 1:9 ratio. Plates were incubated at 21°C and growth was monitored. The best Mfer336 protein crystal was produced using 0.33% (w/v) 1,5-naphthalenedisulfonic acid disodium salt, 0.33% (w/v) 2,5-pyridinedicarboxylic acid, 0.33% (w/v) 3,5-dinitrosalicylic acid and 0.02 M HEPES, pH 6.8 and harvested in mother liquor with 25% (v/v) glycerol as cryo-protectant. X-ray diffraction data sets were collected at 100 K with an oscillation angle of 0.5° over a 360° range at the Australian Synchrotron beamline MX1 (McPhillips *et al.* 2002; Cowieson *et al.* 2015) with a detector distance of 200 mm using a double-crystal monochromator energy of 13 keV.

Phasing, model building, refinement and structure analyses

The crystals showed $P4_22_12$ symmetry with unit cell parameters $a = b = 97.8$ Å and $c = 138.2$ Å corresponding to a solvent content of 59.5% with a single molecule in the asymmetric unit. The X-ray data were anisotropic with diffraction beyond 2.3 Å in the best directions (*hk* plane) but only to 3.0 Å in the weakest direction (*l*-axis); data were processed to 2.5 Å. The Mfer336 structure was determined by molecular replacement as implemented in *MoRDa* (Vagin and Lebedev 2015) which led to a partial model based on the *Pseudomonas aeruginosa* MurF (PDB-4CVK_A) middle and C-terminal domains. *Buccaneer* (Cowtan 2006), within *CCP*4 (Hough and Wilson 2018), was used for further model building and sidechain fitting to the Mfer336 amino acid sequence. The N-terminal domain for Mfer336 could not be modelled by automated model building programs and was built manually in *Coot* (Emsley *et al.* 2010) and refined with *REFMAC*5 (Murshudov *et al.* 2011) and *PHENIX* (Adams *et al.* 2010) including automatic water placement. *MolProbity* (Davis *et al.* 2004) was used for model validation during model building. *Dali* was used to find the closest structural homologues (Holm and Laakso

2016) of Mfer336 (including individual N-terminal, middle and C-terminal domains).

RESULTS

Identification of core cell wall biosynthesis enzymes, taxonomic distribution and phylogenetic analyses

Sequences of core bacterial murein monomer biosynthesis enzymes, and those required for murein export and cross-linking, and archaeal enzymes indicated by gene clustering patterns to be required for pseudomurein formation, were used in BLASTP and/or PSI-BLASTP (Altschul *et al.* 1990, 1997; Van Heijenoort 1998; Smith 2006; Schneider and Sahl 2010; Bugg *et al.* 2011) searches of archaea. Properties of the core enzymes involved in murein and pseudomurein biosynthesis are summarized in Table S1 (Supporting Information; Van Heijenoort 1998; Smith 2006) and the taxonomic distribution of the enzymes in Archaea is summarized in Table S2 (Supporting Information).

To search for the presence of glycan subunit biosynthesis enzymes, *M. fervidus* GlmS, GlmM and GlmU sequences were used in BLASTP searches of bacterial and archaeal genomes resulting in conserved orthologues being widely found across both domains (>1600 hits in archaea for each of the three enzymes). The enzymes GlmS, GlmM and GlmU lead to the formation of UDP-*N*-acetylglucosamine and are found to be universally present, including in pseudomurein-containing methanogens. Significantly, when MurA and MurB sequences from the Gram-positive Firmicute *B. subtilis subsp. subtilis str.* 168 were used for searches of archaea no hits were identified in pseudomurein-containing methanogens, and only six and eight sequences (respectively) in other archaea (see Tables S1 and S2, Supporting Information). Thus, MurA and MurB which are required for the early steps in murein biosynthesis are absent in pseudomurein-containing methanogens. It is unclear what the roles of MurA-like (UDP-N-acetylglucosamine 1-carboxyvinyltransferase) and MurB-like proteins are in other archaea, but since murein has not been found in any archaea these MurA orthologues may simply function in aromatic amino acid synthesis, as MurA possesses a high degree of homology (circa 25%) with another enolpyruvyl transferase, 3-phosphoshikimate 1-carboxyvinyltransferase, involved in aromatic amino acid biosynthesis. Another possibility, due to the propensity for binding UDP, might be an unknown role in biosynthesis of the cell walls of these other archaea or their extracellular polysaccharides. MurB is a small dehydrogenase which could be re-deployed to undertake another role, perhaps in the cell wall synthesis in those archaea. The formation of NAT has been proposed to require epimerase and oxidoreductase activities (Hartmann and König 1994); pseudomurein-containing methanogens all contain orthologues of UDP-glucose 4-epimerase (GalE; EC 5.1.3.2) which may either catalyze the interconversion of UDP-glucose/UDP-*N*-acetylglucose to the respective galactose-based amino-sugars and possibly play a role in the formation of NAT (Supporting Information; Carbone *et al.* 2018).

To investigate enzymes that are likely to be involved in synthesis of the pseudomurein peptide stem, bacterial murein peptide ligases were used as search sequences of archaea which identified orthologues of the bacterial peptide ligases MurC, MurD and MurE in all pseudomurein-containing methanogens, and MurF in all methanogens (including pseudomurein-containing methanogens and in some of the newly described non-Euryarchaeal methanogens (Evans *et al.* 2015; Castelle and Banfield 2018; Berghuis *et al.* 2019; Table S2, Supporting Information). In all methanogens the peptide ligase most similar to bacterial MurF (now termed CfbE) was recently shown to catalyze a late step in the biosynthesis of the methanogen cofactor F_{430} (Zheng *et al.* 2016; Moore *et al.* 2017). We refer to the pseudomurein-containing methanogen peptide ligases as pMurC, pMurD1, pMurD2 and pMurE and these are typically found in two gene clusters (Fig. 2; discussed further below). In all pseudomurein-containing methanogens two MurD-like sequence are found (pMurD1 and pMurD2).

We next investigated murein cell wall biosynthesis enzymes with membrane-associated activities. Pseudomurein biosynthesis shows several differences in the order of reaction steps as formation of the disaccharide moiety of the glycopeptide monomer occurs before the transfer to membrane protein by MraY (Hartmann and König 1990) rather than afterwards as in bacteria. In bacteria, MraY is a membrane phospho-*N*-acetylmuramoyl-pentapeptide transferase (EC 2.7.8.13) that catalyzes the transfer of the phospho-*N*-acetylmuramoyl-pentapeptide to C_{55} undecaprenyl-phosphate at the intracellular face of the membrane. MraY sequences are found in most archaeal phyla (Table S2, Supporting Information) and are universally distributed in bacteria. Similar to the study by Leahy *et al.* (2010) MurG, UppP, UppS and flippase sequences were also found for pseudomurein-containing methanogens (Supporting Information). Thus, pseudomurein-containing methanogens possess orthologues of all the membrane-associated enzymes steps in the biosynthesis of murein (i.e. MurG, MraY, UppP, UppS and flippase). Interestingly, Leahy *et al.* (2010) also identified two hypothetical transmembrane proteins that are only found in pseudomurein-containing methanogens (corresponding to Mfer371 and Mfer259/Mfer218). Neither of these two genes are associated with the two main gene clusters for peptide ligases (described below) and it is unclear if these gene products have any role in pseudomurein biosynthesis.

Little or no evidence was found for archaeal orthologues of enzymes involved in modification of murein, such as FEM proteins that add glycine residues to the peptide stem, or final cross-linking (transglycosylases or transpeptidases) or murein recycling enzymes (Tables S1 and S2, Supporting Information). In contrast, the final extracellular cross-linking of pseudomurein has been postulated to be catalyzed by transglutaminases (Leahy *et al.* 2010), some of which may be involved in cell wall recycling (Supporting Information). When the phage endoisopeptidase PeiW, which contains a C-terminal transglutaminase domain, was used in PSI-BLASTP searches of archaea, >650 hits above threshold were found, the large majority (>460) in pseudomurein-containing methanogens (Makarova, Aravind and Koonin 1999). Many of these are annotated as adhesins in gut methanobacterial species which are likely involved in binding to other components and/or cells in the gut milieu (Leahy *et al.* 2010).

Identification of pseudomurein biosynthesis gene clusters and comparison to the *dcw* bacterial division and cell wall biosynthesis gene cluster

We identified two separate gene clusters encoding enzymes likely involved in pseudomurein peptide stem formation pMurC, pMurD1, pMurD2 and pMurE (Hartmann and König 1990; Hartmann and König 1994; Claus and König 2010). These are present in all pseudomurein-containing methanogens and are not found in other methanogens. Gene Cluster 1

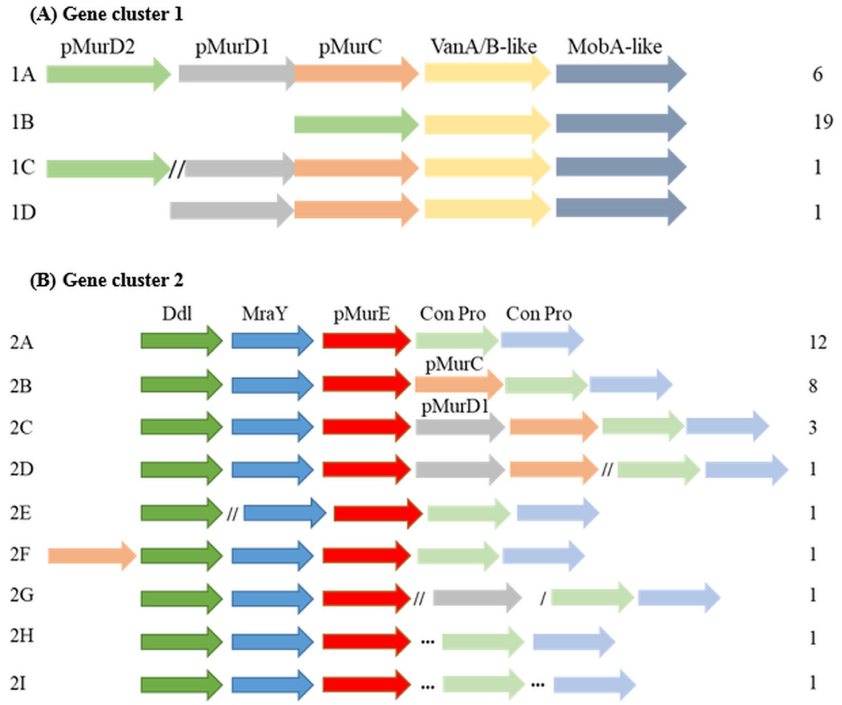

Figure 2. (A) Pseudomurein biosynthesis Gene Cluster 1. Genes are identified by the color arrow code below and the number of occurrences in a sample of 27 pseudomurein-containing archaeal genomes is on the right. Gene cluster types are classed by letters on the left margin. The overlapping grey/orange arrows indicate the genes have been annotated as being fused/overlapping;/, one gene separated;//, two genes separated. The arrowhead of each gene represents its direction in the genome. *Methanothermobacter thermautotrophicus* ΔH and *Methanothermobacter marburgensis* display the Type 1A gene clustering pattern. Species within each of the 1A–1D cluster types are shown in Supporting Information. The *Methanothermus fervidus* enzymes which show the cluster type 1C gene order are Mfer340, Mfer337, Mfer336, Mfer335 and Mfer334, respectively (left to right). **(B)** Pseudomurein biosynthesis Gene Cluster 2. Genes are shown for 29 pseudomurein-containing genomes. Gene cluster types are classed by letters on the left margin and the number of instances that these occur in the right-hand margin (/, one gene separated;//, two genes separated; …, not closely associated with pMurE in genome). The proteins in *M. fervidus* that show the cluster type 2A gene order are Mfer760 (D-Ala-D-Ala ligase), Mfer761 (MraY), Mfer762 (pMurE), Mfer763 and Mfer764, respectively (left to right). The latter two proteins are highly conserved pseudomurein-containing methanogen transmembrane domain-containing proteins. The species in each of the 2A–2I cluster types are shown in Supp. Information. Ddl, D-Ala-D-Ala ligase; Con Pro, conserved pseudomurein-containing methanogen genes encoding proteins with predicted transmembrane domains.

(Fig. 2A) contains two ORFs invariably present in pseudomurein-containing methanogens (exemplified by *Mfer334* and *Mfer335*; Fig. 2 cluster Type 1C) and at least one pseudomurein peptide ligase (C/D type). Mfer335-related sequences are restricted to pseudomurein-containing methanogens and the protein is predicted to possess a membrane insertion domain close to the N-terminus. Mfer334 shows distant similarity to MocA (EC 2.7.7.76; a nucleotidyl transferase), MobA-like and GlmU-like proteins (nucleotidyl transferases EC 2.7.7.77 and EC 2.7.7.23, respectively) (Jagtap *et al.* 2012; Marchler-Bauer *et al.* 2017).

Gene Cluster 2 (Fig. 2B) always includes three genes encoding proteins implicated in cell wall biosynthesis: MraY (*Mfer761*), pMurE (*Mfer762*) related to bacterial MurE peptide ligase and a D-Ala-D-Ala ligase or CarB (*Mfer760*). In some cases, pMurC and/or pMurD1 genes are also present. Two other highly conserved genes (*Mfer763* and *Mfer764*) which encode proteins with transmembrane-containing domains are always found downstream of the gene encoding pMurE, but the protein sequences share no homology with any annotated proteins in the non-redundant database. These enzymes could represent examples of *de novo* archaeal evolutionary solutions to cell wall biosynthesis. The D-Ala-D-Ala ligase/CarB protein is a member of the ATP-grasp family (Makarova, Aravind and Koonin 1999). In bacteria, D-Ala-D-Ala ligase catalyzes the formation of the D-Ala-D-Ala dipeptide prior to its addition to the growing murein peptide stem (Bugg *et al.* 2011). The role of Mfer760 in pseudomurein biosynthesis is not clear, but one possibility is in the three-step formation (Hartmann and König 1994) of the starting substrate UDP-N^{α}-glutamyl-γ-phosphate (Supporting Information).

Although Gene Clusters 1 and 2 account for most of the key activities in the biosynthesis of the pseudomurein peptide monomer, there are other gene clusters in pseudomurein-containing methanogens. For example, GlmM and GlmU are usually associated with each other, along with PGM (phosphoglucomutase EC 5.4.2.2). In addition, in approximately half of pseudomurein-containing methanogens, orthologues to bacterial cell wall modification/amidation enzymes MurT and a CobB/CobQ domain-containing glutamine transferase (GatD) are associated with each other (Münch *et al.* 2012).

The composition and order of the genes in pseudomurein Gene Clusters 1 and 2 differs largely from that of the division and cell wall (*dcw*) cluster of peptidoglycan biosynthesis genes in bacteria (Tamames *et al.* 2001; Figure S1, Supporting Information). The two gene pairings that are the most conserved in the *dcw* gene cluster are *murD-ftsW* and *ftsW-murG* (Tamames *et al.* 2001). PSI-BLASTP searches were performed using the *B. subtilis subsp. subtilis str.* 168 sequences and it was found that pseudomurein-containing methanogens do not contain orthologues of FtsL, FtsI, FtsQ, FtsW and MraW, and FtsA orthologues are only rarely found (e.g. in only four species; not shown). In addition, whereas in the *E. coli* and *B. subtilis dcw* gene cluster patterns all four Mur peptide ligase-encoding genes can be found either in the cluster or close by, in pseudomurein-containing methanogens the pMur genes are almost always split

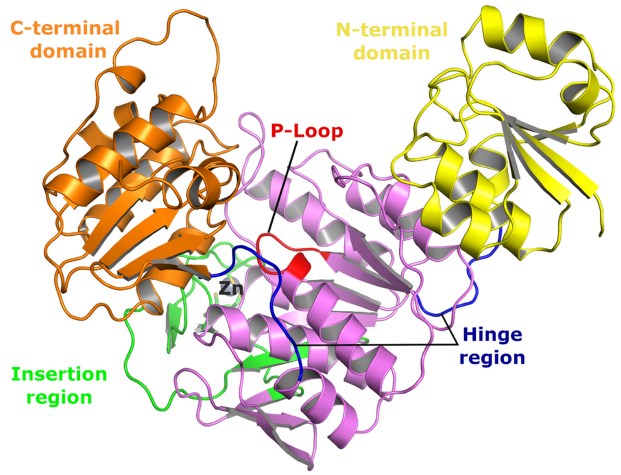

Figure 3. Cartoon representation of the Mfer336 peptide ligase structure. Each domain is colored (yellow for N-terminal, violet and green for middle domain and orange for C-terminal) with the connecting loops, presumed hinge regions, colored in blue. The insertion region (green) in the middle domain, between β10 and β13, contains a zinc-binding site formed by four conserved cysteine residues represented as sticks and bound zinc shown as a grey sphere. The signature ATPase P-loop conserved for the Mur-family ligases is shown in red.

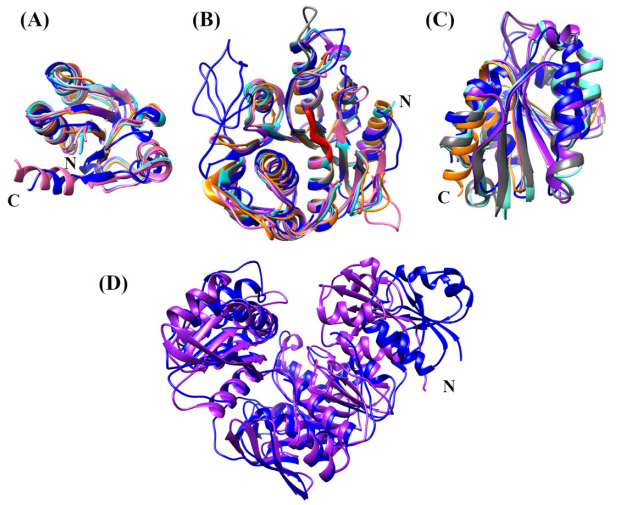

Figure 4. Domain structural analysis comparison of Mfer336 and bacterial MurC peptide ligases (Holm and Laakso 2016). The Mfer336 structure is shown as blue, *Thermotoga maritima* MurC 1J6U orange, *Haemophilus influenzae* MurC 1P3D purple, *E. coli* MurC 2F00 cyan, *Yersinia pestis* MurC 4HV4 grey and *Pseudomonas aeruginosa* MurC 5VVW pink, for each domain. **(A)** The Mfer336 N-terminal domain compared to bacterial MurC N-terminal domains. **(B)** The middle domain comparison with the conserved P-loop shown in red. The Mfer336 middle domain contains a Zn^{2+}-ribbon insertion region on the left. **(C)** The C-terminal domain comparison. The structural alignment was calculated from the best-aligning pair of chains considering each domain of the Mfer336 structure as a reference structure using Chimera (Pettersen *et al.* 2004). The structure of 5VVW does not contain the C-terminal domain. **(D)** Tertiary structural comparison of Mfer336 (blue) with *H. influenzae* MurC (purple). The superposition was calculated based on the middle domain.

into two clusters, and MurG orthologue-encoding genes are not associated with either Gene Cluster 1 or Gene Cluster 2.

The crystal structure of an archaeal pseudomurein peptide ligase

Whereas most of the enzymes involved in pseudomurein synthesis can be identifed reliably from sequence analyses, the assignment of pseudomurein peptide ligases to subclasses is more problematic. Given the large differences in the chemistry of the peptide cross-linking between murein and pseudomurein, the roles of peptide ligases are key to understanding how pseudomurein developed. While phylogenetic analysis has been unable to clearly implicate the individual activities of the pMurs, a crystal structure offered the potential for more definitve assignments. The crystal structure of the *M. fervidus* peptide ligase Mfer336 was determined at 2.5 Å resolution. The structure comprises three domains with a central domain flanked by N- and C-terminal domains (Fig. 3) revealing a similar fold to murein peptide ligases. The Mfer336 N-terminal domain (residues 1–102) is a Rossmann-like fold with a central five-stranded β-sheet and alternating helices and is presumed to bind the UDP-amino acid precursor UDP-N^{α}-glutamyl-γ-phosphate (by homology to the bacterial Mur ligase N-terminal domain that binds the UDP-MurNAc substrate). The central Mfer336 domain (108–348) is a P-loop-containing ATP-binding domain with a core of an eight-stranded predominantly parallel (7/8 strands) β-sheet flanked by three helices on either side and an inserted Zn^{2+}-ribbon motif unique to archaeal peptide ligases (Figures S2 and S4, Supporting Information). The C-terminal domain (355–491) is another Rossmann-like fold of a six-stranded β-sheet domain. The three domains are linked by short hinged regions (König, Kandler and Hammes 1989; Coutinho *et al.* 2003; Pettersen *et al.* 2004; Lairson *et al.* 2008; Lombard *et al.* 2013), 349–354). A structure-based similarity search (Holm and Laakso 2016) showed that Mfer336 is similar to bacterial murein peptide ligases with the highest structural similarity to MurC. Clear structural homology is present for the Mfer336 middle and C-terminal domains with all bacterial Mur peptide ligase structures, but structural alignment is only possible for the Mfer336 N-terminal domain with bacterial Mur C/D-type ligases and not E/F-type structural orthologues (Table S5, Supporting Information). It is the N-terminal domain that distinguishes the MurC/D type from the MurE/F in bacteria (Smith 2006; Díaz-Mejía, Pérez-Rueda and Segovia 2007). The topology of the MurC/D N-terminal domain contains five parallel β-strands flanked by alternating α-helices. In contrast, the N-terminal domain of MurE and MurF contains five β-strands with one arranged in an antiparallel fashion distinct from the MurC/D type (Smith 2006; Díaz-Mejía, Pérez-Rueda and Segovia 2007). Mur peptide ligases have an ATPase P-loop at the interface of the middle and C-terminal domains (Figure S3A, Supporting Information), the UDP moiety of the growing nucleotide-peptide substrate binds to the N-terminal domain, and the C-terminal domain binds the amino acid to be added (Mol *et al.* 2003; Smith 2006).

One clear difference between the Mfer336 pMurC and murein bacterial peptide ligases is the presence of a Zn^{2+}-binding domain (see Figs 3 and 4; Figures S2 and S4, Supporting Information). The inserted (relative to bacterial sequences) amino acids between residues 239–289 in the middle domain extend from the conserved middle domain fold forming two antiparallel strands (β11 and β12) arranging four Cys residues (C257, C260, C279 and C281) into a tetrahedral Zn^{2+}-binding site (average Cys-Zn^{2+} bond length of 2.3 Å) forming a zinc-ribbon motif (Gamsjaeger *et al.* 2007; Figure S2, Supporting Information). The insertion and Cys residues are conserved among archaeal pMurC sequences (Figure S4, Supporting Information), implying that a homologous Zn^{2+}-ribbon is structurally conserved. Conserved Cys residues are present for some pMurD1 sequences.

The conservation of the Mfer336 pMurC overall tertiary structure, individual domain folds (N-terminal, middle- and C-terminal; Fig. 4A–C) and full-length protein (Fig. 4D) and P-loop with bacterial homologues suggests a similar catalytic mechanism. Bacterial Mur enzymes typically adopt an 'open' conformation in the apo state and undergo a conformational change, via domain rigid body motion (Bertrand *et al.* 2000), to a 'closed' conformation when the substrate binds (Smith 2006). The pMurC domains form a continuous surface suitable for a conserved reaction mechanism invoking conformational movements of the three domains. The N-terminal domain is oriented so that the growing UDP-peptide substrate could be activated for ligation of the additional amino acid by the middle domain ATPase site in an analogous mechanism to that described for bacterial Mur ligases (Smith 2006). Due to the lack of a suitable substrate, the peptide ligase activity for Mfer336 could not be confirmed.

Phylogenetic and sequence analysis was performed to investigate which step Mfer336 could catalyse in pseudomurein biosynthesis, with an emphasis on the bacterial MurC/D family. A phylogenetic comparison using representative sequences from bacterial MurC-F from representative diverse bacterial phyla and pseudomurein pMurC, pMurD1, pMurD2 and pMurE, and CfbE classes was generated using the Maximum Likelihood method (Fig. 5; associated traditional Newick tree in Figure S5, Supporting Information; implemented in MEGA7; Kumar, Stecher and Tamura 2016). An alignment for representative sequences in Fig. 5 from each clade is shown in Figure S6 (Supporting Information). The tree in Fig. 5 incorporates sequences with a wide taxonomic diversity with the protein sequence % identities for the sequence comparison for Mfer336 with MurC proteins ranging from 12.3% (*Spirochaeta thermophila*) to 19.8% (*Caldicellulosiruptor kristjanssonii*). For the Mfer336-MurD protein comparison the range was 13.0% (*Acetobacter aceti*) to 19.5% (*Thermoanaerobacter brockii*). It should be mentioned that most of the identity is associated with the middle domain, the largest and the most conserved of the three domains. All nine enzyme types branched separately from each other with strong bootstrap support (eight clades >90%, one 76% (CfbE)). By contrast, eight of the nine nodes between the individual enzyme clades were not resolved (i.e. ≤40% bootstrap support). There was a 43% bootstrap support for an association between the pMurC and pMurD1 clades. *M. kandleri* was the earliest diverging pseudomurein peptide ligase sequence in the CfbE, pMurE and pMurD2 clades, and was second in the pMurD1, (after the hyperthermophilic *M. fervidus* sequence). For pMurC, the *M. kandleri* sequence was the third deepest, after the *M. fervidus* sequence (Mfer336). A phylogenetic tree showing only pseudomurein peptide ligase sequences is presented in Figure S7 (Supporting Information) and reveals similar highly supported individual clades to Fig. 5 but without support for any interconnecting node (see Supporting Information).

DISCUSSION

One of the major differences between bacterial and archaeal cells is found in their cell walls. Murein cell walls are near-universally found in bacterial species (Errington 2013); in marked contrast, archaea contain a variety of structurally unrelated cell wall types (Kandler and König 1993; Albers and Meyer 2011). This suggests that cell wall synthesis underwent independent evolutionary trajectories in the two domains. With respect to pseudomurein, the differences in biosynthesis and chemistry compared to those in murein have previously been considered sufficiently significant to lead to the proposal that these two pathways arose independently (Kandler and König 1993; Albers and Meyer 2011; Errington 2013). We have confirmed here, however, that the pseudomurein-containing methanogens in the archaeal orders Methanobacteriales and Methanopyrales possess orthologues of most bacterial murein biosynthesis enzymes (Leahy *et al.* 2010) and show that the key peptide biosynthesis encoding genes are present in well-conserved gene clusters. Furthermore, we have provided evidence that the putative *M. fervidus* pseudomurein peptide ligase Mfer336 pMurC is clearly related in structure to bacterial MurC/D class murein peptide ligases. These new data, when combined with our taxonomic distribution analysis and the identification of other proteins potentially involved, enables a fresh look at how the two pathways evolved.

Due to the number of bacterial homologues that are suggested to be involved in pseudomurein biosynthesis, it is instructive to first examine the development/assembly of the murein biosynthesis pathway itself from an evolutionary perspective. The production of UDP-*N*-acetylglucosamine by GlmS, GlmM and GlmU was possible by widely distributed universally present enzyme activities. A 4-aminosugar epimerase activity (GalE-like) is also near-universal in all microorganisms and would catalyze the interconversion of UDP-*N*-acetylglucosamine and UDP-*N*-acetylgalactosamine, the latter a common component of pseudomurein (Claus and König 2010). GalE-like UDP-glucose 4-epimerases are widely present in all organisms and were therefore likely to have been in the last universal common ancestor (LUCA; Ishiyama *et al.* 2004; Kanehisa *et al.* 2017). MurA is related to a key enzyme in the biosynthesis of aromatic amino acids (3-phosphoshikimate 1-carboxyvinyl transferase) which is widely distributed in bacteria and some archaea (methanogens and halophilic microorganisms) and is likely to have been the precursor of MurA (Krekel *et al.* 1999). MurB is a NADPH and FAD-dependent reductase, and Dali (Holm and Laakso 2016) structural analysis revealed that it is related to several other types of NAD and/or FAD-dependent enzymes, suggesting that it evolved from an oxidoreductase precursor in LUCA. The murein peptide ligases represent a key aspect of the evolution of the murein biosynthesis pathway. Interestingly, bacterial peptide ligases are distantly related to a folate biosynthesis enzyme (FolC or FPGS; folylpolyglutamate synthase; EC 6.3.2.17) which adds multiple glutamate residues in the final steps in folate biosynthesis (Smith 2006; Díaz-Mejía, Pérez-Rueda and Segovia 2007). FPGS possesses two of the murein peptide ligase domains, the central ATPase domain and the C-terminal Rossman dihydrofolate reductase (DHFR)-like fold domain. FPGS orthologues are widely present in archaea, including the DPANN and Euryarchaea, suggesting the FPGS was present in LUCA. It is possible that a common two-domain ancestor (containing the middle and C-terminal domains) could have diverged to bring forth FPGS, and murein peptide ligases by addition of the N-terminal domains of the MurC/D and MurE/F ligase types in separate events. For the membrane-associated steps, MurG is a Type B glycosyl transferase which are widely distributed in bacteria and archaea and are thought to be an ancient enzyme family, likely present in LUCA (Breton *et al.* 2006; Breton, Fournel-Gigleux and Palcic 2012). MraY is an integral membrane enzyme present in virtually all bacteria and in numerous archaea (Table S2, Supporting Information) and is also likely to have been in LUCA (Chung *et al.* 2013). Similarly, UppP and UppS are also widely present and have been suggested to have been early evolving enzyme activities (Lurie-Weinberger *et al.* 2012; Lombard 2016). The widespread presence of flippase in both

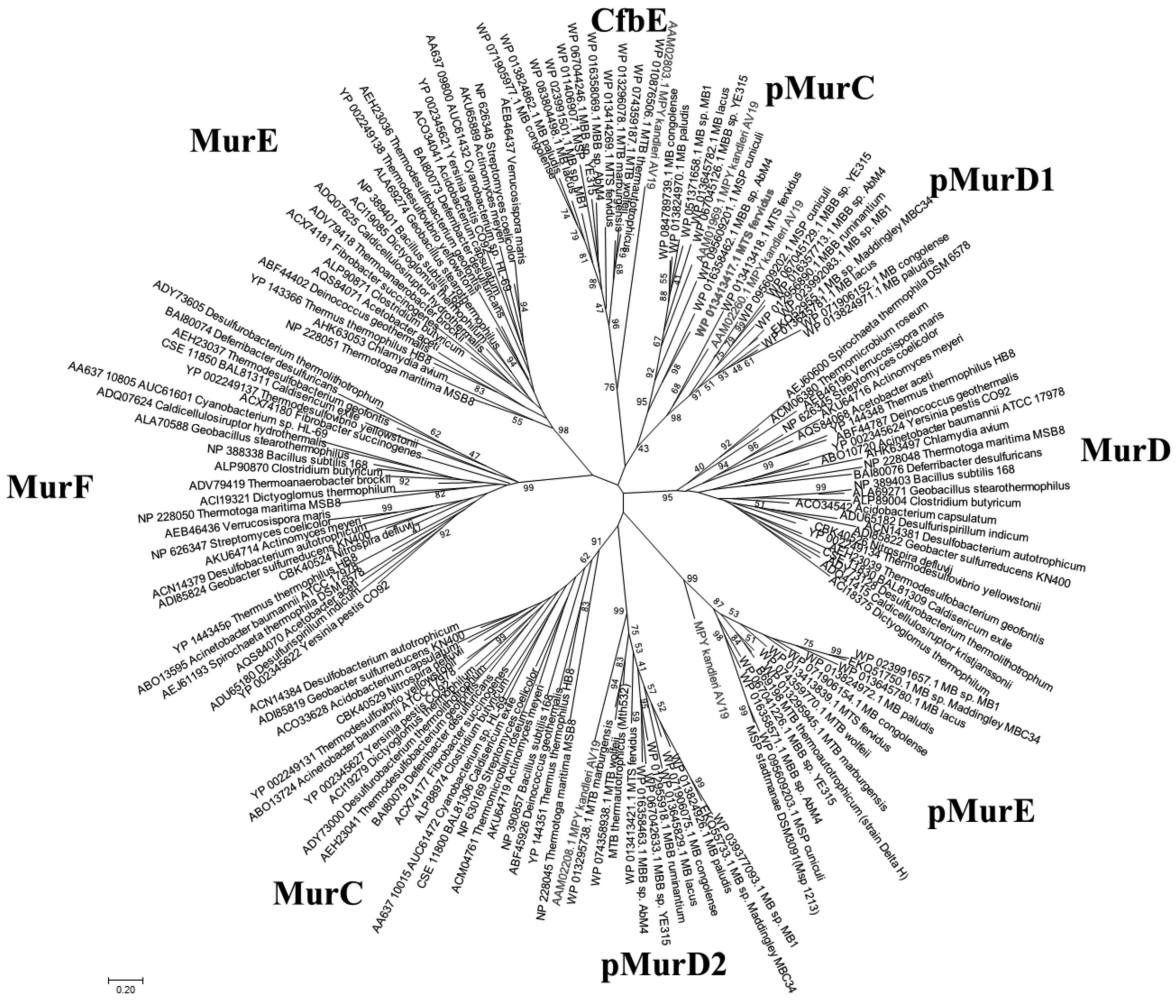

Figure 5. Molecular phylogenetic analysis of murein peptide ligases and their homologues in pseudomurein-containing methanogens (Kumar, Stecher and Tamura 2016). *Methanopyrus kandleri* AV19 sequences are deeply rooted in pseudomurein peptide ligase clades and are highlighted in blue. The Mfer336 sequence is shown in red font. The analysis includes 155 amino acid sequences (59 methanogen and 96 bacterial) and there were 187 conserved positions used in the final dataset. Support at nodes was calculated by non-parametric bootstrap on 100 pseudo-replicates of the original alignment. The tree was inferred using the Maximum Likelihood method based on the JTT matrix-based model as implemented in MEGA7 (Kumar *et al.* 2016). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Bootstrapping values at nodes showing ≤40% support were removed. An alignment for representative peptide ligases from each of the nine clades is shown in Figure S6 (Supporting Information).

bacteria and archaea suggests that a flippase was present in LUCA (Sham *et al.* 2014; Makarova, Galperin and Koonin 2015; Ruiz 2015; Teo and Roper 2015). Interestingly, the enzymes involved in murein extracellular cross-linking and recycling are not found in pseudomurein-containing methanogens, or in other archaea, indicating that these are specific to bacteria (Tables S2 and S3, Supporting Information).

Based on the taxonomic distribution, gene clustering, phylogenetic, predicted enzymatic activity and structural analysis data presented here, an evolutionary process underlying the origin of the pseudomurein biosynthesis pathway (Hartmann and König 1994) can be envisaged. The evidence suggests that the biosynthesis of the glycan-peptide monomer would use the universally distributed GlmS/GlmM/GlmU, a UDP-4-aminosugar epimerase, a MurG-like glycosyl transferase and peptide ligases MurC, MurD1/D2 and MurE (discussed further below). Evidence for archaeal orthologues of MraY, UppP, UppS and flippase proteins being involved in pseudomurein biosynthesis includes their presence in all pseudomurein-containing archaea, the isolation of undecaprenyl pseudomurein intermediates from pseudomurein-containing methanogens (indicating that undecaprenyl acts as the lipid carrier at the membrane (Hartmann and König 1990)), and the genomic clustering of MraY with pMurC-E. In addition, pseudomurein-containing methanogens are sensitive to bacitracin, which specifically targets bacterial UppP (Harris and Pinn 1985). Thus, at least 13 methanogen orthologues of bacterial enzymes can be confidently identified as having a role in pseudomurein biosynthesis. Significantly, genes for the individual enzymes that are likely to be directly involved in peptide stem biosynthesis (pMurC, pMurD1/D2, pMurE and subsequently MraY) and several other highly conserved genes, including two that are only found in pseudomurein-containing methanogens, are found in two conserved clusters. The colocation of these genes in conserved clusters is likely due to gains in functionality e.g. coordinated gene synthesis to match the cell's requirements. The enzymes involved in the biosynthesis of the unique starting UDP-N^{α}-glutamyl-γ-phosphate precursor are less clear, but clues can be

found in the two gene clusters. Conserved archaeal enzymes potentially with the appropriate catalytic activities are either Mfer335 of pseudomurein Gene Cluster 1 or the ATP-grasp protein of Gene Cluster 2 (the D-Ala-D-Ala ligase/CarB; Mfer760) for the first or third phosphoryl-transfer steps; and the GlmU-like nucleotidyl transferase (Mfer334) for the second step.

Due to the large differences in chemistry of the peptide component of pseudomurein, compared to murein (Fig. 1), it is useful to examine the evidence for the role of archaeal pMurC-pMurE peptide ligases in pseudomurein biosynthesis and their evolution more closely. The taxonomic distribution, gene clustering and sequence homologies presented here strongly indicate that the addition of amino acids to the starting substrate UDP-N^{α}-glutamyl-γ-phosphate during synthesis of the pseudomurein cross-linking peptide involves pMurC, pMurD1, pMurD2 and pMurE. The finding that all pseudomurein-containing methanogens contain at least four orthologues of the bacterial murein peptide ligases, that they are found in conserved gene clusters with other probable cell wall biosynthesis enzymes, and that the pseudomurein gene clusters are absent in non-pseudomurein-containing archaea indicates that these enzymes are likely involved in formation of the pseudomurein cross-linking peptide. The structure of Mfer336 (pMurC) shows the highest overall similarity to bacterial MurC/D structures. Based on the structural information available for published bacterial murein ligases and the observations derived from the Mfer336 structure, it is clear that all the peptide ligases in both cell wall biosynthesis pathways possess a highly conserved ATPase middle domain (with a signatory ATP-binding P-loop) and an amino acid-binding DHFR-like C-terminal domain. In the case of murein peptide ligases, the Rossmann-like fold N-terminal domains differ significantly between the MurC/D and MurE/F classes. Enzymes that incorporate Rossmann folds are widely present in both bacteria and archaea (Caetano-Anollés *et al.* 2009).

Like any pathway, the pseudomurein pathway cannot have arisen all at once. It likely emerged through the combination of preexisting components, potentially with the aid of several *de novo*-generated genes by the ancestors of pseudomurein-containing methanogens. Our focus in the following is on the origin of pseudomurein peptide ligases. Consistent with previous studies (Leahy *et al.* 2010; Nelson-Sathi *et al.* 2015), we detected archaeal peptide ligases using MurC-MurF bacterial peptide ligase protein sequence searches which overall possessed relatively low sequence identities (~13–20% amino acid identity over the full length of the proteins; Mol *et al.* 2003; Smith 2006). Much of the sequence identity was found in the relatively large, conserved ATP-binding middle domain, with less conservation being found for the N- and C-terminal domains (Figure S6, Supporting Information). In order to examine possible evolutionary links between cell wall peptide ligases we generated a Maximum Likelihood phylogenetic tree for the larger family of MurC-F/pMurC-F enzymes. Three central aspects emerge: the early presence of peptide ligases, independent cell wall evolution in bacteria and archaea after their divergence from LUCA, and their connections to other ancient pathways. For example, the phylogenetic tree (Fig. 5) of the (p)MurC-F peptide ligase family shows that the murein biosynthetic subtrees MurC, MurD, MurE and MurF contain only bacterial members, while the archaeal pMurC, pMurD1, pMurD2, MurE and MurF (CfbE) subtrees contain only archaeal members. The relative antiquity of different members of the (p)MurC-F family cannot be readily addressed with phylogeny because the short basal branches connecting the subtrees in Fig. 5 preclude unambiguous assignment of a root. This suggests that these gene families arose quite early in evolution. The lack of interleaving among bacterial/archaeal sequences within each of the nine families indicates that in this case, obvious lateral gene transfer across domains has not occurred for the peptide-ligase encoding genes sampled. Thus, the phylogeny indicates the establishment of four (five in the case of archaea) defined deeply diverging peptide ligase families that today are involved in (pseudo)murein synthesis. One interpretation of that finding is that LUCA possessed these activities and was able to generate amino acid polymers consisting mainly of glutamate, alanine, glycine and aspartate with a structural rather than catalytic function. This would best be viewed as an early cell wall synthetic capability in LUCA. Both the presence of peptide ligase function in LUCA, or alternatively an early gain of peptide ligase function in archaea, would have been followed by selective losses of this activity in archaeal lineages with other cell wall types.

The amino acids that occur in cell walls are connected to primordial metabolism in two ways. First, the main amino acids of the (pseudo)murein wall are Glu, Ala, Gly and Lys, in addition to Asp in some methanogens and *meso*-DAP in bacteria. Because Lys and *meso*-DAP are both derived from Asp, the amino acid content of murein and pseudomurein derives from Glu, Ala, Gly and Asp. These four amino acids are a recurrent theme in early evolution. They are consistently identified as the most ancient amino acids by different approaches to the origin of the genetic code (Trifonov 2004; Copley, Smith and Morowitz 2005; Wong 2005); they were recently shown to form without enzymes from the corresponding ketoacids using only Fe^{2+} as the catalyst and hydroxylamine as the amino donor (Muchowska, Varma and Moran 2019); they are the first acids that should arise in prebiotic metabolism if metabolism started out via the acetyl-CoA pathway (Preiner *et al.* 2020) generating the reductive amination products of acetate, pyruvate, oxaloacetate and α-ketoglutarate (Muchowska, Varma and Moran 2019) and they are central biosynthetic precursors of nucleic acid bases (Martin and Russell 2007). The accumulation in the cell wall of ancient amino acids that are very central to metabolism is consistent with an ancient nature of non-ribosomal peptide bond formation. A second noteworthy aspect concerns the presence of D-amino acids in peptidoglycan cell walls. At the very origin of metabolism, before there were free-living cells with genetically encoded pathways, the compounds from which the first pathways emerged must have been supplied by abiotic (geochemical) sources. Ribosomal peptide synthesis cannot accommodate D-amino acids, non-ribosomal peptide synthesis can. It is possible that D-amino acids in cell walls reflect a chemical relict from the habitat in which life arose, one that was abiotically supplied with D- and L-amino acids before cells learned to synthesize 20 (or 22) amino acids from H_2, CO_2 and N_2.

Independent evidence from methanogen biology underscores the early emergence of peptide ligases. For example, pMurF is identical (synonymous) with the F_{430} biosynthetic enzyme CfbE and is usually found in a gene cluster separate from those for pMurC-pMurE (Zheng *et al.* 2016). This is significant because CfbE/pMurF is essential for F_{430} biosynthesis, whereby it activates a carboxyl group by generating an acyl phosphate using ATP (Zheng *et al.* 2016; Moore *et al.* 2017). F_{430} is essential for methanogenesis because it is required as cofactor by methyl-CoM reductase, the methane generating enzyme. Methanogenesis was clearly ancestral to the methanogens and was very likely an early metabolism in archaea (Decker, Junger-

mann and Thauer 1970; Ueno *et al.* 2006; Martin and Russell 2007; Wong *et al.* 2007, 2016; Martin *et al.* 2008; Di Giulio 2009; Kelly, Wickstead and Gull 2011; Petitjean *et al.* 2015; Weiss *et al.* 2018; Berghuis *et al.* 2019; Borrel *et al.* 2019; Evans *et al.* 2019), placing pMurF (CfbE) in the archaeal common ancestor, consistent with the deep family diversification depicted in Fig. 5. In addition, the shared similarity of Mur/pMur peptide ligases with FPGS (Sun *et al.* 1998; Smith 2006), which adds glutamyl residues via non-ribosomal peptide bond formation in folate biosynthesis, connects murein synthesis with bacterial folate which is the backbone of acetogenesis. Acetogenesis is chemically related to methanogenesis and is a likely candidate for carbon and energy metabolism of the first bacteria (Martin 2020). Polyglutamate residues are also found in essential methanogen cofactors F_{420} and methanofuran (Li *et al.* 2003; Allen and White 2014).

An alternative to the above scenario for how the pseudomurein biosynthesis pathway developed is that a proportion of the genes for pseudomurein synthesis were acquired through early lateral gene transfer from bacteria that already possessed most if not all of the core murein cell wall biosynthesis machinery (including peptide ligases), near the separation of the bacterial and archaeal domains. The enzymes and their encoding genes that appear to show clear similarity to their bacterial counterparts (such as pMurC) could have been transferred singly, or perhaps more likely, as smaller clusters. Evidence for large lateral gene transfer events betwen bacteria and major phyla of archaea have been described (Groussin *et al.* 2015; Nelson-Sathi *et al.* 2015). Pseudomurein-containing methanogens would have then filled in the gaps in their biosynthetic pathway due to the lack of MurA and MurB (potentially utilizing *de novo* generated genes), extensively retooled their peptide ligases, developed an alternative extracellular cross-linking mechanism (via transglutaminases) and developed an alternative recycling system.

We have presented here an analysis of the genomic distribution and gene clustering of cell wall biosynthesis enzymes that are shared between the pseudomurein-containing methanogenic orders Methanobacteriales and Methanopyrales, and bacteria. We have also presented phylogenetic analysis and the first crystal structure for a pseudomurein peptide ligase. Analysis of the shared and differentiating features between the murein and pseudomurein biosynthesis pathways, including the clustering of key biosynthetic genes and the Mfer336 crystal structure allow us to conclude that pseudomurein biosynthesis and murein biosynthesis share a deep evolutionary history, shedding light on the early evolution of cell walls.

DATA AVAILABILITY

The atomic coordinates and structure factors (code 6VR7) have been deposited in the Protein Data Bank (http://wwpdb.org/). All other data are available in the manuscript and Supporting Information.

ACKNOWLEDGMENTS

We thank the New Zealand Synchrotron Group and the Australian Synchrotron (ANSTO) MX beamlines for their assistance and support. We thank Eric Altermann and Nik Palevich for critical review of the manuscript. We thank Professor Dr Bettina Siebers for arranging the MSc studentship for Julia Sauter.

SUPPLEMENTARY DATA

Supplementary data are available at *FEMSMC* online.

FUNDING

This research was funded by the Royal Society of New Zealand Marsden fund (AGR1301), ERC (EcolMetabOrigin 101018894) and the Pastoral Greenhouse Gas Research Consortium (grant number 17- MET5.4-V1).

Conflicts of interest. The authors declare that they have no conflicts of interest with the contents of this article.

REFERENCES

Adams PD, Afonine PV, Bunkóczi G *et al. PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* 2010;**66**:213–21.

Albers SV, Meyer BH. The archaeal cell envelope. *Nat Rev Microbiol* 2011;**9**:414–26.

Allen KD, White RH. Identification of structurally diverse methanofuran coenzymes in methanococcales that are both N-formylated and N-acetylated. *Biochemistry* 2014;**53**: 6199–210.

Altschul SF, Gish W, Miller W *et al.* Basic local alignment search tool. *J Mol Biol* 1990;**215**:403–10.

Altschul SF, Madden TL, Schäffer AA *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;**25**:3389–402.

Bapteste É, Brochier C, Boucher Y. Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. *Archaea* 2005;**1**:353.

Barreteau H, Kovač A, Boniface A *et al.* Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiol Rev* 2008;**32**:168–207.

Berghuis BA, Yu FB, Schulz F *et al.* Hydrogenotrophic methanogenesis in archaeal phylum Verstraetearchaeota reveals the shared ancestry of all methanogens. *Proc Natl Acad Sci* 2019;**116**:5037.

Bertrand JA, Fanchon E, Martin L *et al.* 'Open' structures of MurD: domain movements and structural similarities with folylpolyglutamate synthetase. *J Mol Biol* 2000;**301**:1257–66.

Borrel G, Adam PS, McKay LJ *et al.* Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea. *Nat Microbiol* 2019;**4**:603–13.

Boucher Y, Douady CJ, Papke RT *et al.* Lateral gene transfer and the origins of prokaryotic groups. *Annu Rev Genet* 2003;**37**:283–328.

Bouhss A, Trunkfield AE, Bugg TDH *et al.* The biosynthesis of peptidoglycan lipid-linked intermediates. *FEMS Microbiol Rev* 2008;**32**:208–33.

Breton C, Fournel-Gigleux S, Palcic MM. Recent structures, evolution and mechanisms of glycosyltransferases. *Curr Opin Struct Biol* 2012;**22**:540–9.

Breton C, Šnajdrová L, Jeanneau C *et al.* Structures and mechanisms of glycosyltransferases. *Glycobiology* 2006;**16**:29R–37R.

Bugg TDH, Braddick D, Dowson CG *et al.* Bacterial cell wall assembly: still an attractive antibacterial target. *Trends Biotechnol* 2011;**29**:167–73.

Caetano-Anollés G, Yafremava LS, Gee H *et al.* The origin and evolution of modern metabolism. *Int J Biochem Cell Biol* 2009;**41**:285–97.

Carbone V, Schofield LR, Sang C *et al.* Structural determination of archaeal UDP-N-acetylglucosamine 4-epimerase from

Methanobrevibacter ruminantium M1 in complex with the bacterial cell wall intermediate UDP-*N*-acetylmuramic acid. *Proteins Struct Funct Bioinf* 2018;**86**:1306–12.

Castelle CJ, Banfield JF. Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* 2018;**172**:1181–97.

Caveney NA, Li FK, Strynadka NC. Enzyme structures of the bacterial peptidoglycan and wall teichoic acid biogenesis pathways. *Curr Opin Struct Biol* 2018;**53**:45–58.

Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J* 1986;**5**:823–6.

Chung BC, Zhao J, Gillespie RA *et al.* Crystal structure of MraY, an essential membrane enzyme for bacterial cell wall synthesis. *Science* 2013;**341**:1012–6.

Claus H, König H. Cell envelopes of methanogens. In: *Prokaryotic Cell Wall Compounds: Structure and Biochemistry*, König H., Claus H., Varma A (eds), Berlin: Springer, 2010, 231–51.

Copley SD, Smith E, Morowitz HJ. A mechanism for the association of amino acids with their codons and the origin of the genetic code. *Proc Natl Acad Sci* 2005;**102**:4442–7.

Coutinho PM, Deleury E, Davies GJ *et al.* An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol* 2003;**328**:307–17.

Cowieson NP, Aragao D, Clift M *et al.* MX1: a bending-magnet crystallography beamline serving both chemical and macromolecular crystallography communities at the Australian Synchrotron. *J Synchrotron Radiat* 2015;**22**:187–90.

Cowtan K. The *Buccaneer* software for automated model building. 1. Tracing protein chains. *Acta Crystallogr Sect D Biol Crystallogr* 2006;**62**:1002–11.

Davis IW, Murray LW, Richardson JS *et al.* MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* 2004;**32**:W615–9.

Decker K, Jungermann K, Thauer RK. Energy production in anaerobic organisms. *Angewandte Chemie International Edition in English* 1970;**9**:138–58.

Di Giulio M. A methanogen hosted the origin of the genetic code. *J Theor Biol* 2009;**260**:77–82.

Díaz-Mejía JJ, Pérez-Rueda E, Segovia L. A network perspective on the evolution of metabolism by gene duplication. *Genome Biol* 2007;**8**:R26.

Dramsi S, Magnet S, Davison S *et al.* Covalent attachment of proteins to peptidoglycan. *FEMS Microbiol Rev* 2008;**32**:307–20.

Egan AJF, Cleverley RM, Peters K *et al.* Regulation of bacterial cell wall growth. *FEBS J* 2017;**284**:851–67.

Emsley P, Lohkamp B, Scott WG *et al.* Features and development of *Coot*. *Acta Crystallogr Sect D Biol Crystallogr* 2010;**66**:486–501.

Errington J. L-form bacteria, cell walls and the origins of life. *Open Biol* 2013;**3**:120143.

Evans PN, Boyd JA, Leu AO *et al.* An evolving view of methane metabolism in the Archaea. *Nat Rev Microbiol* 2019;**17**:219–32.

Evans PN, Parks DH, Chadwick GL *et al.* Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* 2015;**350**:434.

Formanek H. Three-dimensional models of the carbohydrate moieties of murein and pseudomurein. *Zeitschrift für Naturforschung C* 1985;**40**:555–61.

Gambelli L, Meyer BH, McLaren M *et al.* Architecture and modular assembly of *Sulfolobus* S-layers revealed by electron cryotomography. *Proc Natl Acad Sci* 2019;**116**:25278–86.

Gamsjaeger R, Liew CK, Loughlin FE *et al.* Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem Sci* 2007;**32**:63–70.

Groussin M, Boussau B, Szöllősi G *et al.* Gene acquisitions from bacteria at the origins of major archaeal clades are vastly overestimated. *Mol Biol Evol* 2016;**33**:305–10.

Harris JE, Pinn PA. Bacitracin-resistant mutants of a mesophilic *Methanobacterium* species. *Arch Microbiol* 1985;**143**:151–3.

Hartmann E, König H. A novel pathway of peptide biosynthesis found in methanogenic Archaea. *Arch Microbiol* 1994;**162**:430–2.

Hartmann E, König H. Comparison of the biosynthesis of the methanobacterial pseudomurein and the eubacterial murein. *Naturwissenschaften* 1990;**77**:472–5

Holm L, Laakso LM. Dali server update. *Nucleic Acids Res.* 2016;**44**:W351–5.

Hough MA, Wilson KS. From crystal to structure with *CCP4*. *Acta Crystallogr Sect D Struct Biol* 2018;**74**:67.

Illergård K, Ardell DH, Elofsson A. Structure is three to ten times more conserved than sequence - A study of structural response in protein cores. *Proteins Struct Funct Bioinf* 2009;**77**:499–508.

Ishiyama N, Creuzenet C, Lam JS *et al.* Crystal structure of WbpP, a genuine UDP-*N*-acetylglucosamine 4-epimerase from *Pseudomonas aeruginosa*: substrate specificity in UDP-hexose 4-epimerases. *J Biol Chem* 2004;**279**:22635–42.

Jagtap PKA, Soni V, Vithani N *et al.* Substrate-bound crystal structures reveal features unique to *Mycobacterium tuberculosis* *N*-Acetyl-glucosamine 1-phosphate uridyltransferase and a catalytic mechanism for acetyl transfer. *J Biol Chem* 2012;**287**:39524–37.

Kandler O, König H. Cell envelopes of archaea: structure and chemistry. In: Kates M, Kushner DJ, Matheson AT . (eds). *The Biochemistry of Archaea (Archaebacteria)*. Amsterdam: Elsevier, 1993, 223–59.

Kandler O, König H. Cell wall polymers in Archea (Archaebacteria). *Cell Mol Life Sci* 1998;**54**:305–8.

Kandler O. Cell wall biochemistry and three-domain concept of life. *Syst Appl Microbiol* 1993;**16**:501–9.

Kanehisa M, Furumichi M, Tanabe M *et al.* KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017;**45**:D353–61.

Kelly S, Wickstead B, Gull K. Archaeal phylogenomics provides evidence in support of a methanogenic origin of the Archaea and a thaumarchaeal origin for the eukaryotes. *Proc R Soc B Biol Sci* 2011;**278**:1009–18.

Klingl A, Pickl C, Flechsler J. Archaeal cell walls. In: Kuhn A. (ed), *Bacterial Cell Walls and Membranes*. Cham: Springer International Publishing, 2019, 471–93.

König H, Hartmann E, Kärcher U. Pathways and principles of the biosynthesis of methanobacterial cell wall polymers. *Syst Appl Microbiol* 1993;**16**:510–7.

König H, Kandler O, Hammes W. Biosynthesis of pseudomurein: isolation of putative precursors from *Methanobacterium thermoautotrophicum*. *Can J Microbiol* 1989;**35**:176–81.

Krekel F, Oecking C, Amrhein N *et al.* Substrate and inhibitor-induced conformational changes in the structurally related enzymes UDP-*N*-acetylglucosamine enolpyruvyl transferase (MurA) and 5-enolpyruvylshikimate 3-phosphate synthase (EPSPS). *Biochemistry* 1999;**38**:8864–78.

Kumar S, Rubino FA, Mendoza AG *et al.* The bacterial lipid II flippase MurJ functions by an alternating-access mechanism. *J Biol Chem* 2019;**294**:981–90.

Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016;**33**:1870–4.

Lairson LL, Henrissat B, Davies GJ *et al.* Glycosyl transferases: structures, functions, and mechanisms. *Annu Rev Biochem* 2008;**77**:521–55.

Lake JA, Sinsheimer JS. The deep roots of the Rings of Life. *Genome Biol Evolut* 2013;**5**:2440–8.

Leahy SC, Kelly WJ, Altermann E *et al.* The genome sequence of the rumen methanogen *Methanobrevibacter ruminantium* reveals new possibilities for controlling ruminant methane emissions. *PLoS ONE* 2010;**5**:e8926.

Li H, Xu H, Graham DE *et al.* Glutathione synthetase homologs encode α-L-glutamate ligases for methanogenic coenzyme F_{420} and tetrahydrosarcinapterin biosyntheses. *Proc Natl Acad Sci* 2003;**100**:9785–90.

Lombard J. Early evolution of polyisoprenol biosynthesis and the origin of cell walls. *PeerJ* 2016;**4**:e2626.

Lombard V, Golaconda Ramulu H, Drula E *et al.* The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 2014;**42**:D490–5.

Lurie-Weinberger MN, Peeri M, Tuller T *et al.* Extensive interdomain lateral gene transfer in the evolution of the human commensal *Methanosphaera stadtmanae*. *Front Genet* 2012;**3**:182.

Makarova KS, Aravind L, Koonin EV. A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases. *Protein Sci* 1999;**8**:1714–9.

Makarova KS, Galperin MY, Koonin EV. Comparative genomic analysis of evolutionarily conserved but functionally uncharacterized membrane proteins in archaea: prediction of novel components of secretion, membrane remodeling and glycosylation systems. *Biochimie* 2015;**118**:302–12.

Mandelstam J, Rogers HJ. The incorporation of amino acids into the cell-wall mucopeptide of staphylococci and the effect of antibiotics on the process. *Biochem J* 1959;**72**:654–62.

Marchler-Bauer A, Bo Y, Han L *et al.* CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 2017;**45**:D200–3.

Martin W, Baross J, Kelley D *et al.* Hydrothermal vents and the origin of life. *Nat Rev Microbiol* 2008;**6**:805–14.

Martin W, Russell MJ. On the origin of biochemistry at an alkaline hydrothermal vent. *Philos Trans R Soc B Biol Sci* 2007;**362**:1887–926.

Martin WF. Older than genes: the acetyl CoA pathway and origins. *Front Microbiol* 2020;**11**:817.

McPhillips TM, McPhillips SE, Chiu HJ *et al.* *Blu-Ice* and the *Distributed Control System*: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J Synchrotron Radiat* 2002;**9**:401–6.

Meyer BH, Albers S-V. *Archaeal Cell Walls* Chichester. John Wiley & Sons, Ltd, 2014, 1–13. eLS.

Mohammadi T, van Dam V, Sijbrandi R *et al.* Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane. *EMBO J* 2011;**30**:1425–32.

Mol CD, Brooun A, Dougan DR *et al.* Crystal structures of active fully assembled substrate- and product-bound complexes of UDP-N-acetylmuramic acid:l-alanine ligase (MurC) from *Haemophilus influenzae*. *J Bacteriol* 2003;**185**:4152–62.

Moore SJ, Sowa ST, Schuchardt C *et al.* Elucidation of the biosynthesis of the methane catalyst coenzyme F_{430}. *Nature* 2017;**543**:78–82.

Muchowska KB, Varma SJ, Moran J. Synthesis and breakdown of universal metabolic precursors promoted by iron. *Nature* 2019;**569**:104–7.

Münch D, Roemer T, Lee SH *et al.* Identification and *in vitro* analysis of the GatD/MurT enzyme-complex catalyzing lipid II amidation in *Staphylococcus aureus*. *PLoS Pathog* 2012;**8**:e1002509.

Murshudov GN, Skubák P, Lebedev AA *et al.* *REFMAC5* for the refinement of macromolecular crystal structures. *Acta Crystallogr Sect D Biol Crystallogr* 2011;**67**:355–67.

Nelson-Sathi S, Sousa FL, Roettger M *et al.* Origins of major archaeal clades correspond to gene acquisitions from bacteria. *Nature* 2015;**517**:77–80.

Petitjean C, Deschamps P, López-García P *et al.* Extending the conserved phylogenetic core of *Archaea* disentangles the evolution of the third domain of life. *Mol Biol Evol* 2015;**32**:1242–54.

Pettersen EF, Goddard TD, Huang CC *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 2004;**25**:1605–12.

Preiner M, Igarashi K, Muchowska KB *et al.* A hydrogen-dependent geochemical analogue of primordial carbon and energy metabolism. *Nat Ecol Evol* 2020;**4**:534–42.

Ruiz N. Lipid flippases for bacterial peptidoglycan biosynthesis. *Lipid Insights* 2015;**2015**:21–31.

Scheffers DJ, Pinho MG. Bacterial cell wall synthesis: new insights from localization studies. *Microbiol Mol Biol Rev* 2005;**69**:585–607.

Schneider T, Sahl HG. An oldie but a goodie - cell wall biosynthesis as antibiotic target pathway. *Int J Med Microbiol* 2010;**300**:161–9.

Schofield LR, Beattie AK, Tootill CM *et al.* Biochemical characterisation of phage pseudomurein endoisopeptidases PeiW and PeiP using synthetic peptides. *Archaea* 2015;**2015**:1.

Sham LT, Butler EK, Lebar MD *et al.* MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. *Science* 2014;**345**:220–2.

Slesarev AI, Mezhevaya KV, Makarova KS *et al.* The complete genome of hyperthermophile *Methanopyrus kandleri AV19* and monophyly of archaeal methanogens. *Proc Natl Acad Sci* 2002;**99**:4644–9.

Smith CA. Structure, function and dynamics in the *mur* family of bacterial cell wall ligases. *J Mol Biol* 2006;**362**:640–55.

Smith DR, Doucette-Stamm LA, Deloughery C *et al.* Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J Bacteriol* 1997;**179**:7135–55.

Sun X, Bognar AL, Baker EN *et al.* Structural homologies with ATP- and folate-binding enzymes in the crystal structure of folylpolyglutamate synthetase. *Proc Natl Acad Sci* 1998;**95**:6647–52.

Tamames J, González-Moreno M, Mingorance J *et al.* Bringing gene order into bacterial shape. *Trends Genet* 2001;**17**:124–6.

Tashiro T, Ishida A, Hori M *et al.* Early trace of life from 3.95 Ga sedimentary rocks in Labrador, Canada. *Nature* 2017;**549**:516–8.

Teo A, Roper D. Core steps of membrane-bound peptidoglycan biosynthesis: recent advances, insight and opportunities. *Antibiotics* 2015;**4**:495–520

Trifonov EN. The triplet code from first principles. *J Biomol Struct Dyn* 2004;**22**:1–11.

Ueno Y, Yamada K, Yoshida N *et al.* Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* 2006;**440**:516–9.

Vagin A, Lebedev A. MoRDa, an automatic molecular replacement pipeline. *Acta Crystallogr Sect A Found Adv* 2015; **71**:s19.

Van Heijenoort J. Assembly of the monomer unit of bacterial peptidoglycan. *Cell Mol Life Sci* 1998;**54**:300–4.

Weiss MC, Preiner M, Xavier JC *et al.* The last universal common ancestor between ancient Earth chemistry and the onset of genetics. *PLos Genet* 2018;**14**: e1007518.

Wong JTF, Chen J, Mat WK *et al.* Polyphasic evidence delineating the root of life and roots of biological domains. *Gene* 2007;**403**:39–52.

Wong JTF, Ng SK, Mat WK *et al.* Coevolution theory of the genetic code at age forty: pathway to translation and synthetic life. *Life* 2016;**6**:12.

Wong JTF. Coevolution theory of genetic code at age thirty. *Bioessays* 2005;**27**:416–25.

Zhang Y, Schofield LR, Sang C *et al.* Expression, purification, and characterization of (R)-sulfolactate dehydrogenase (ComC) from the rumen methanogen *Methanobrevibacter millerae* SM9. *Archaea* 2017;**2017**:1.

Zheng K, Ngo PD, Owens VL *et al.* The biosynthetic pathway of coenzyme F430 in methanogenic and methanotrophic archaea. *Science* 2016;**354**:339–42.