

# Identification of a gene involved in the biosynthesis pathway of the terminal sugar of the archaeallin *N*-linked tetrasaccharide in *Methanococcus maripaludis*

Yan Ding · Gareth M. Jones · Cedric Brimacombe ·  
Kaoru Uchida · Shin-Ichi Aizawa · Susan M. Logan ·  
John F. Kelly · Ken F. Jarrell 

Received: 1 September 2015 / Accepted: 28 October 2015 / Published online: 21 November 2015  
© Springer International Publishing Switzerland 2015

**Abstract** In *Methanococcus maripaludis*, the three archaeallins which comprise the archaeallum are modified at multiple sites with an *N*-linked tetrasaccharide with the structure of Sug-4- $\beta$ -ManNAc3NAcA6Thr-4- $\beta$ -GlcNAc3NAcA-3- $\beta$ -GalNAc, where Sug is a unique sugar (5*S*)-2-acetamido-2,4-dideoxy-5-*O*-methyl-L-erythro-hexos-5-ulo-1,5-pyranose, so far found exclusively in this species. In this study, a six-gene cluster *mmp1089–1094*, neighboring one of the genomic regions already known to contain genes involved with the archaeallin *N*-glycosylation pathway, was examined for its potential involvement in the archaeallin *N*-glycosylation or sugar biosynthesis pathway. The co-transcription of these six genes was demonstrated by RT-PCR. Mutants carrying an in-frame deletion in *mmp1090*, *mmp1091* or *mmp1092* were successfully generated. The  $\Delta$ *mmp1090* deletion mutant was archaeallated when examined by electron microscopy

and mass spectrometry analysis of purified archaealla showed that the archaeallins were modified with a truncated *N*-glycan in which the terminal sugar residue and the threonine linked to the third sugar residue were missing. Both gene annotation and bioinformatic analyses indicate that MMP1090 is a UDP-glucose 4-epimerase, suggesting that the unique terminal sugar of the archaeallin *N*-glycan might be synthesised from UDP-glucose or UDP-*N*-acetylglucosamine with an essential early step in synthesis catalysed by MMP1090. In contrast, no detectable phenotype related to archaeallin glycosylation was observed in mutants deleted for either *mmp1091* or *mmp1092* while attempts to delete *mmp1089*, *mmp1093* and *mmp1094* were unsuccessful. Based on its demonstrated involvement in the archaeallin *N*-glycosylation pathway, we designated *mmp1090* as *aglW*.

**Keywords** Archaea · Methanogens · *N*-linked glycosylation · Archaealla · Mass spectrometry · In-frame deletion

## Introduction

*N*-glycosylation is a prevalent protein co/post-translational modification which has been found in all three domains of life (Aebi 2013; Jarrell et al. 2014; Nothaft and Szymanski 2010). In the Archaea, bioinformatics analysis detected at least one copy of the predicted oligosaccharyltransferase (OST) gene (*aglB*),

Y. Ding · G. M. Jones · C. Brimacombe ·  
K. F. Jarrell (✉)  
Department of Biomedical and Molecular Sciences,  
Queen's University, Kingston K7L 3N6, Canada  
e-mail: jarrellk@queensu.ca

K. Uchida · S.-I. Aizawa  
Department of Life Sciences, Prefectural University of  
Hiroshima, Shobara, Hiroshima 727-0023, Japan

S. M. Logan · J. F. Kelly (✉)  
Human Health Therapeutics Portfolio, National Research  
Council, Ottawa, K1A 0R6, Canada  
e-mail: john.kelly@nrc-cnrc.gc.ca

encoding the most conserved enzyme in the *N*-glycosylation system, in 166 of 168 sequenced archaeal genomes, indicating the almost universal presence of this protein modification throughout the third domain (Kaminski et al. 2013). This is in contrast to the, so far, limited observation of *N*-glycosylation in Bacteria where the pathway seems restricted to most members of the Epsilon-subdivision of Proteobacteria (*Campylobacter*, *Wolinella*, *Helicobacter*) and a subset of the Deltaproteobacteria, including *Desulfovibrio* species (Nothaft and Szymanski 2010, Nothaft and Szymanski 2013). Archaea also have the most diverse *N*-glycan structures, with a great variety of sugar components (including unique sugars), modifications and glycosidic linkages, as well as a range of linking sugars, including simple hexoses (Jarrell et al. 2014).

Although the *N*-glycosylation system is widespread in Archaea, genetic studies of the archaeal *N*-glycosylation pathway, in combination with biochemical and structural approaches, are mainly restricted to three model microorganisms: *Methanococcus maripaludis*, *Sulfolobus acidocaldarius* and *Haloferax volcanii* (Jarrell et al. 2014). Notably, however, structures for AglB are only available from other species including *Archaeoglobus fulgidus* and *Pyrococcus furiosus* (Igura et al. 2007, Matsumoto et al. 2013). Based on the studies in these three microorganisms, an archaeal *N*-glycosylation model has been proposed (Jarrell et al. 2014). On the inner leaflet of the cytoplasmic membrane, various glycosyltransferases (GTs) sequentially add sugar monomers onto a dolichol lipid carrier (either mono- or diphosphorylated; Dol-P or Dol-PP) embedded in the cytoplasmic membrane. This lipid-linked *N*-glycan precursor is then translocated across the cytoplasmic membrane via a flippase, and at the exterior face of the cytoplasmic membrane, the *N*-glycan is transferred *en bloc* by AglB from the lipid carrier onto select asparagine residues in an Asn-Xaa-Ser/Thr (Xaa ≠ Pro) sequon in the acceptor glycoprotein. Additional sugars may still be added to the glycan after its transfer to the acceptor protein (Cohen-Rosenzweig et al. 2012). The OST, as well as a large number of archaeal GTs and enzymes involved in the biosynthesis of the individual sugars comprising the glycan have been identified in the three model microorganisms (Jarrell et al. 2014).

To date, archaeal *N*-glycoproteins are typically external proteins or membrane-bound proteins and include archaeellins (formerly archaeal flagellins (Albers and Jarrell 2015, Jarrell and Albers 2012)), S-layer proteins, pilins, cytochrome *b*<sub>558/556</sub> in *S. acidocaldarius*, and even an archaeal viral protein (Jarrell et al. 2014). Most studies of the archaeal *N*-glycosylation pathway have employed either archaeellins or S-layer proteins as the reporter protein (Jarrell et al. 2010).

In *M. maripaludis*, the swimming organelle archaeellum is composed of three structural proteins: the major archaeellins FlaB1 and FlaB2 and the minor archaeellin FlaB3, the latter forming the curved, hook-like region (Chaban et al. 2007). All three archaeellins are *N*-glycosylated at multiple sites with the tetrasaccharide Sug-4-β-ManNAc3NAcA6Thr-4-β-GlcNAc3NAcA-3-β-GalNAc, where Sug is (5*S*)-2-acetamido-2,4-dideoxy-5-*O*-methyl-L-erythro-hexos-5-ulo-1,5-pyranose, which to date has been found exclusively in this species (Kelly et al. 2009). The Epd pili of *M. maripaludis* (Nair et al. 2014) are also composed of glycoproteins, however the major pilin, EpdE, has an *N*-linked pentasaccharide consisting of the archaeellin tetrasaccharide with an additional hexose added as a branch to the linking sugar GalNAc (Ng et al. 2011).

Using both genetic and biochemical methods, a series of enzymes have been demonstrated in *M. maripaludis* to be involved in the *N*-glycosylation or sugar biosynthesis pathways. The genes for these components have been localised mainly to two regions. The first region extends from *mmp1079* to *mmp1088*, including *aglO* (*mmp1079*), *aglA* (*mmp1080*) and *aglL* (*mmp1088*) encoding the second, third and fourth GTs, respectively (Vandyke et al. 2009); *aglXYZ* (*mmp1081–1083*) encoding enzymes involved in the acetamidino modification of the third sugar (Jones et al. 2012); *aglU* (*mmp1084*) encoding the enzyme that transfers the threonine onto the third sugar; and *aglV* (*mmp1085*) encoding a methyltransferase for the methylation of the unique fourth sugar (Ding et al. 2013). The second region is located from *mmp0350* to *mmp0359*. Several of the genes located in this area are involved in the biosynthesis of the second and third sugars of the glycan, including *aglI7* (*mmp0350*), *aglI8* (*mmp0351*), *aglI9* (*mmp0352*), *aglI20* (*mmp0353*) and *aglI21* (*mmp0357*) (Namboori

and Graham 2008a; Siu et al. 2015; VanDyke et al. 2008). The gene for the OST, *aglB* (*mmp1424*) is found at a separate locus (Vandyke et al. 2009). Besides the work mentioned above, in vitro studies employing heterologously expressed and purified enzymes have identified other genes, including *mmp0705*, *mmp0706*, *mmp1077* and *mmp1680*, that are involved in acetamido sugar biosynthesis (Namboori and Graham 2008b). However, genetic studies are still needed to confirm the involvement of these genes in archaellin *N*-glycan sugar biosynthesis. Key enzymes in the *N*-glycosylation pathway, i.e. the GT responsible for the transfer of the first sugar to the dolichol carrier and the putative flippase, have not yet been identified in *M. maripaludis*, although MMP1423 is a strong candidate for the first glycosyltransferase. MMP1423 is a putative family 4 glycosyltransferase with a GT\_GPT\_archaea domain, found in UDP-GlcNAc:dolichol-P GlcNAc-1-P transferase (GPT)-like proteins in archaea. In Eukaryotes, GPT catalyse the initial step in the *N*-glycosylation pathway, i.e., the transfer of GlcNAc-1-P from UDP-GlcNAc to dolichol-P resulting in the formation of GlcNAc-P-P-dolichol. Previous attempts to delete *mmp1423*, however, have been unsuccessful (Vandyke et al. 2009).

In *M. maripaludis*, *N*-glycosylation plays an important role in archaella formation. Archaella are only assembled using archaellins that are modified with disaccharide or longer *N*-glycans, i.e. in *ΔaglA* and *ΔaglL* mutants, but not with monosaccharide-modified or non-glycosylated archaellins, i.e. *ΔaglO* or *ΔaglB* mutants (Siu et al. 2015; Vandyke et al. 2009). While archaella are assembled with a disaccharide or trisaccharide attached to the archaellins, motility is reduced in these mutants compared to that of wild type (WT) cells as measured in swarm plate assays. The swarming diameters in these assays are directly related to the size of the *N*-linked glycan, with the cells producing a disaccharide glycan swarming the least (Vandyke et al. 2009). Interestingly, this minimum disaccharide *N*-glycan rule does not apply to pili assembly, since pili were observed on the cell surface of a *ΔaglB* mutant where the constituent pilins would be non-glycosylated (Vandyke et al. 2009). On the other hand, not all of the *N*-glycosylation sites normally occupied in the archaellins in WT cells are necessary for archaella

formation (Ding et al. 2015). Archaella could be assembled using FlaB2 in which 3 out of 4 *N*-glycosylation

sites were removed by site-directed mutagenesis (SDM) (leaving only the 1st *N*-glycosylation site intact), but not with non-glycosylated FlaB2 (Ding et al. 2015).

In this work, we studied a six-gene operon *mmp1089–1094*, neighbouring the first identified *N*-glycosylation genetic region, and demonstrated that one gene, *mmp1090*, is involved in the unique fourth sugar biosynthesis pathway. No significant change in the archaellin *N*-glycan was observed in either *Δmmp1091* or *Δmmp1092* mutant cells. Several attempts to delete the other genes in the operon, namely *mmp1093* (annotated as *coaD*, (Sarmiento et al. 2013)), *mmp1094* (annotated as *ppsA* (Sarmiento et al. 2013)) or *mmp1089* (Vandyke et al. 2009) were unsuccessful, suggesting that these genes might be essential for *M. maripaludis*.

## Materials and methods

### Strains and growth conditions

*M. maripaludis* S2 *Δhpt* (Mm900) (Moore and Leigh 2005) and the mutants derived from Mm900 (Table 1) were cultured anaerobically in sealed serum bottles containing 10 mL medium under an atmosphere of CO<sub>2</sub>–H<sub>2</sub> (20:80) at 35 °C with shaking. Unless otherwise specified, cells were cultured in Balch medium III (Balch et al. 1979). McCas medium was used at various steps during the creation of in-frame deletions (Moore and Leigh 2005). A final concentration of 1 mg/mL neomycin was added to McCas medium as needed for selection following transformation. To obtain *M. maripaludis* single colonies, cells were plated onto McCas medium-Noble agar (1.5 %, w/v) plates containing 240 µg/mL 8-azahypoxanthine (Acros Organics, NJ) for selection and incubated inside of an anaerobic canister at 37 °C for 5 days (Moore and Leigh 2005). *M. maripaludis* strains carrying a complementation plasmid were cultured in nitrogen-free medium supplemented with either 10 mM L-alanine or 10 mM NH<sub>4</sub>Cl (Lie et al. 2005) as sole nitrogen source in the presence of puromycin (2.5 µg/mL) for plasmid selection. *Escherichia coli*

**Table 1** Strains and plasmids used in this study

Strains	Description	References
<i>M. maripaludis</i>		
Mm900	<i>M. maripaludis</i> S2 $\Delta hpt$	Moore and Leigh (2005)
$\Delta mmp1090$	Mm900 $\Delta mmp1090$	This study
$\Delta mmp1091$	Mm900 $\Delta mmp1091$	This study
$\Delta mmp1092$	Mm900 $\Delta mmp1092$	This study
<i>E. coli</i>		
TOP10	F- <i>mcrA</i> $\Delta(mrr-hsdRMS-mcrBC)$ $\phi 80lacZAM15$ $\Delta lacX74$ <i>nupG</i> <i>recA1</i> <i>araD139</i> $\Delta(ara-leu)7697$ <i>galE15</i> <i>galK16</i> <i>rpsL</i> (Str <sup>R</sup> ) <i>endA1</i> $\lambda^-$	Invitrogen
Plasmids		
pCRPrTNeo	<i>hmv</i> promoter- <i>hpt</i> fusion plus Neo <sup>r</sup> cassette in pCR2.1Topo; Amp <sup>r</sup>	Moore and Leigh (2005)
pKJ626	pCRPrTNeo harboring the inframe deletion of <i>mmp1089</i>	Vandyke et al. (2009)
pKJ807	pCRPrTNeo harboring the inframe deletion of <i>mmp1090</i>	This study
pKJ806	pCRPrTNeo harboring the inframe deletion of <i>mmp1091</i>	This study
pKJ785	pCRPrTNeo harboring the inframe deletion of <i>mmp1092</i>	This study
pKJ992	pCRPrTNeo harboring the inframe deletion of <i>mmp1093</i>	This study
pKJ877	pCRPrTNeo harboring the inframe deletion of <i>mmp1094</i>	This study
pHW40	<i>nif</i> promoter- <i>lacZ</i> fusion plus Pur <sup>r</sup> cassette; Amp <sup>r</sup>	Lie et al. (2005)
pKJ1189	pHW40 harboring <i>mmp1090</i>	This study
pKJ1115	pHW40 harboring <i>mmp1090</i> Y151A mutant	This study
pKJ1236	pHW40 harboring <i>mmp1090</i> K155A mutant	This study

TOP10 cells used for molecular cloning steps were cultured at 37 °C with shaking in Luria–Bertani broth (LB) or on LB plates with 100 µg/mL ampicillin for selection.

#### Reverse-transcription (RT)-PCR

To determine if the *mmp1089* to *mmp1094* genes were co-transcribed, RT-PCR was performed as per the manufacturer's protocol using a One-Step RT-PCR kit (Qiagen Inc.) with primers (Table 2) which amplify across each of the intergenic regions between neighbouring genes from *mmp1089* to *mmp1094*. The RNA template was extracted from Mm900 cells using a High Pure RNA Isolation Kit (Roche Inc.), followed by an additional DNase treatment (Turbo DNA-free Kit, Ambion) at 37 °C for 30 min. PCR amplifications with the same primer pairs were also performed using either purified RNA not subjected to reverse transcription as template to exclude the possibility of genomic DNA contamination of the RNA, or genomic DNA from Mm900 as template to confirm the amplicon size and primer specificity.

#### Plasmid constructions for in-frame gene deletions

Plasmids used for in-frame deletions of the targeted genes in *M. maripaludis* (Table 1) were constructed as previously described using the primers listed in Table 2 (Chaban et al. 2007, Moore and Leigh 2005). Briefly, ~1 kb of the DNA upstream (gene specific P1 and P2 primers) and downstream (gene specific P3 and P4 primers) of the targeted gene was amplified by PCR and then ligated via the *AscI* restriction sites incorporated into the two interior primers P2 and P3. After ligation, the majority of the target gene is deleted, leaving a short residual 5' and 3' piece which is still in frame. This ligation product was used as template to perform a second PCR with the exterior primers P1 and P4 and the resulting amplification product was cloned into the *BamHI* or *XbaI* restriction site in the vector pCRprTNeo, which carries a neomycin resistant gene but lacks an origin of replication for *M. maripaludis* (Moore and Leigh 2005). DNA sequencing confirmed that all gene deletions in the pCRprTNeo vectors were in frame. These recombinant plasmids were then used for the creation of the specific in-frame gene deletions in *M. maripaludis*.

**Table 2** Primers used in this study

Primers	Sequence	Restriction site <sup>a</sup>
For RT-PCR		
89-90-F	AACTGCCATTATAGTTTGGATGAGT	
89-90-R	ATGAGATCGTAAATATTGGTTCGG	
90-91-F	ATGTTCTAACATTAATCTGTGCTGC	
90-91-R	AGCTAACTGATGCCATGAAGTCTC	
91-92-F	TTCCACAATGCACTGCATCTCCAAGG	
91-92-R	CCACTTGTAGCCTTCCCAATTTCCG	
92-93-F	CCTGTGTTTCCAATTGATGCAACAAG	
92-93-R	GACGTTTATGGCGACTCAATTTCCG	
93-94-F	CTAACATTTAACGGATTTACATCGTG	
93-94-R	GATTGAACTTGTTATAACTTCATGC	
For in-frame deletions		
1090-P1	ATCGGATCCTTGCGAAATGCGGGAATTGAGG	BamHI
1090-P2	GCATGGCGCGCCTGATAAATCCTGCACCGCCAGTTAC	AscI
1090-P3	ATGCGGCGCGCCTGGATGCCTGAAGTTGAATTGG	AscI
1090-P4	ATCGGATCCTCTGCAACTGCGTTAAGTCCTG	BamHI
1091-P1	ATCGGATCCGATGATAATTATGGAATCATGTATG	BamHI
1091-P2	GCATGGCGCGCCTGAGAAGCCTTGTTCCAAATCCTGC	AscI
1091-P3	ATGCGGCGCGCCGGATGGCTTGAAGCAAACGTTG	AscI
1091-P4	ATCGGATCCAATGCAGTATTCGATATCATAAAGC	BamHI
1092-P1	ATGTCTAGATCTAGACTGTGGTCAGGCTGGAAGCAGACCGCAAATCG	XbaI
1092-P2	GCATGGCGCGCCTAAAATGTCTCGAAATATATCCGATC	AscI
1092-P3	ATGCGGCGCGCCGTTGTATCGTTGTTGCGCTACCG	AscI
1092-P4	ATCTCTAGATCTAGATATGGCTTCCGATAAATCCTGCACCGCCAG	XbaI
1093-P1	AAAGGATCCCGACCTGCTCAGGTAGTTTATGCAGG	BamHI
1093-P2	AAGGCGCGCCGTGCAGTATATCAAAGGTACCTCCG	AscI
1093-P3	AAGGCGCGCCAAGAATAAGAAATGGGGAGATAGACAC	AscI
1093-P4	AAAGGATCCTTAACGTGTAAAGTATCGGTAGCGC	BamHI
1094-P1	AAAGGATCCCGTCTCTATTGTAGACTGTAATTGGTGC	BamHI
1094-P2	AAGGCGCGCCGCTAAGTTTCTTTGATAAAGTACCTG	AscI
1094-P3	AAGGCGCGCCTCAATCTGTGGTCAGGCTGGAAGCAG	AscI
1094-P4	AAGGATCCTTGATGCAACAAGTACAAATCCCCCAAGTG	BamHI
Sequencing primers for in-frame deletions		
1090-seq-F	AGGTTGTAGAGAATATATGTGG	
1090-seq-R	CCTTTATCATTCGTTCTTTGTAG	
1091-seq-F	GACTGTTGTATCGTTGTTTGC	
1091-seq-R	CCACATATATTCTCTACAACCTTTG	
1092-seq-F	TATTGGAATTACAAGCGATG	
1092-seq-R	TCCTGCTGCAGGTATGATAGC	
For complementation studies		
1090-comp-F	CCAATGCATGATTAATAAATATACTTACTAC	NsiI
1090-comp-R	GCACGCGTTTAAATCTTTTAAATCCCGTCATCC	MluI
Y151A-SDM-F	GTTATCTCCTGCCGGTTTAAGTA <sup>b</sup>	
Y151A-SDM-R	TACTTAAACC <sup>b</sup> GAGGAGATAAC <sup>b</sup>	
K155A-SDM-F	CTCCTTACGGTTTAAGTGCCTTCTGTGCAGAAGAGTAC <sup>b</sup>	

**Table 2** continued

Primers	Sequence	Restriction site <sup>a</sup>
K155A-SDM-R	GTACTCTTCTGCACAGAA <u><b>GGC</b></u> ACTTAAACCGTAAGGAG <sup>b</sup>	

<sup>a</sup> Restriction sites in the primer sequences are underlined in the table

<sup>b</sup> The codon changes in the SDM primers are bold and underlined in the table

### Generation of an in-frame deletion of targeted genes in *M. maripaludis*

Markerless in-frame deletions of *mmp1090*, *mmp1091* and *mmp1092* were successfully generated using the procedure described previously (Moore and Leigh 2005). Several attempts to create in-frame deletions of *mmp1089*, *mmp1093* and *mmp1094* were also conducted using the same method, but all were unsuccessful. Single transformant colonies picked from the McCas-Noble agar plates containing 8-aza-hypoxanthine were inoculated into Balch medium III and then screened by using washed whole cells in PCR reactions with primers (listed in Table 2) that would amplify across the target gene to identify deletion mutants. The PCR products were examined by agarose gel electrophoresis and the sizes were compared to that predicted for the WT and deletion versions of the gene in order to identify the specific gene deletion mutants. Deletion mutants were re-streaked for purity on Balch medium III agar plates and the presence of the desired gene deletion reconfirmed by PCR.

### Complementation of the *Ammp1090* deletion strain

*mmp1090* was amplified by PCR using Mm900 genomic DNA as template and primers listed in Table 2. The forward primer had an NsiI site incorporated while the reverse primer had a MluI site introduced. The PCR product was digested with NsiI and MluI, and ligated into NsiI/MluI digested vector pHW40, putting transcription of the cloned gene under the control of the regulatable *nif* promoter (Lie et al. 2005). The fidelity of the gene in the complementation plasmid was confirmed by DNA sequencing. This complementation plasmid, designated pKJ1189, was then transformed into *Ammp1090* mutant cells via the PEG-based methodology

(Tumbula et al. 1994). Following recovery overnight, the cells were then sub-cultured in the presence of 2.5 µg/mL puromycin for plasmid selection. Transformant cells were subsequently grown in nitrogen-free medium supplemented with either 10 mM L-alanine (*nif* promoter is induced) or 10 mM NH<sub>4</sub>Cl (*nif* promoter is repressed) in the presence of 2.5 µg/mL puromycin. Complementation was also performed in *Ammp1090* mutant cells with the same vector backbone containing *mmp1090* carrying a mutation at either Y151 (Y155A) or K155 (K155A). These mutations were created using site-directed mutagenesis (SDM) and primers listed in Table 2, essentially as reported previously (Bardy and Jarrell 2003).

### Western blot analysis

Whole cell lysates from the various *M. maripaludis* strains were separated by 15 % SDS-PAGE and transferred onto an Immobilon-P membrane (Millipore Inc.) (Towbin et al. 1979). Chicken anti-FlaB2 specific antibodies were used as primary antibody to recognise the major archaellin FlaB2, which has been routinely used as the reporter protein for N-glycosylation in our previous studies (Ding et al. 2013; Jones et al. 2012). Horseradish peroxidase-conjugated rabbit anti-chicken immunoglobulin Y (Jackson Immuno Research Laboratories, West Grove PA) was used as secondary antibody, and the blot was developed using Immobilon Western Chemiluminescent HRP Substrate (Millipore Canada Inc., Etobicoke ON).

### Archaella purification

Archaella from the Mm900 (WT) strain as well as the *Ammp1090*, *Ammp1091* and *Ammp1092* deletion strains were isolated as previously described (Bardy et al. 2002).



## Mass spectrometry analysis of archaellin *N*-glycan in deletion strains

Each archaella sample (40 µg) was incubated overnight at 37 °C with trypsin (Promega, Madison, WI) at an approximate ratio of 20:1 (protein:enzyme, wt/wt) in 50 mM ammonium bicarbonate. The digests were then analysed by nano-liquid chromatography-tandem mass spectrometry (Nano-LC–MS/MS) using a NanoAquity UPLC system (Waters, Milford, MA) coupled to an Ultima hybrid quadrupole time-of-flight (QTOF) mass spectrometer (Waters). The digests were injected onto an Acclaim PepMax100 C<sub>18</sub> µ-precolumn (5 mm by 300 µm i.d.; Dionex/Thermo Scientific, Sunnyvale CA) and resolved on a 1.7- µm BEH130 C<sub>18</sub> column (100 µm by 100 mm i.d.; Waters, Milford, CA) using the following gradient conditions: 1–45 % ACN, 0.1 % formic acid in 36 min and 45–95 % ACN, 0.1 % formic acid in 2 min. The flow rate was 400 nL/min. MS/MS spectra were acquired on doubly, triply and quadruply charged ions and searched against the NCBI nr database using the Mascot search engine (Matrix Science, Ltd., London, United Kingdom). The spectral datasets were searched for glycopeptide MS/MS spectra which were then interpreted by hand.

## Electron microscopy

*M. maripaludis* cells from an overnight culture were briefly washed with 2 % NaCl and resuspended in phosphate-buffered saline. Cells were then loaded onto carbon-Formvar-coated copper grids and stained with 2 % phosphotungstic acid. Grids were examined in a Hitachi 7000 electron microscope operating at an accelerating voltage of 75 kV.

## Results

Numerous genes immediately adjacent to the gene cluster encompassed by *mmp1089* to *mmp1094* have been previously shown to be involved in the *N*-glycosylation or *N*-glycan sugar biosynthesis pathways in *M. maripaludis* (Ding et al. 2013; Jones et al. 2012, Vandyke et al. 2009). This led us to examine the possible involvement of *mmp1089* to *mmp1094* in the *N*-glycosylation pathway, first by examination of their annotations. The annotation of MMP1089 is as a polysaccharide biosynthesis protein containing RfbX

or MATE (multidrug and toxic compound extrusion) Wzx-like domains often found in polysaccharide export proteins and flippases, MMP1090 as an UDP-glucose 4-epimerase (also named as UDP-galactose 4-epimerase or GalE), MMP1091 as an ADP-glucose pyrophosphorylase, MMP1092 as an auxin efflux carrier, MMP1093 (CoaD) as a 4'-phosphopantetheine adenyltransferase, and MMP1094 (PpsA) as a phosphoenolpyruvate synthase (Table 3). In addition, most of the proteins encoded by *mmp1089*–*mmp1094* have high sequence identity/similarity over almost their entire length to well-studied proteins, as listed in Table 3. While the annotations and bioinformatic analyses of MMP1093 and MMP1094 indicate that these proteins are likely involved in central metabolic pathways, the annotations of MMP1089–MMP1091 suggest that they all may be involved either directly in *N*-glycosylation or in sugar biosynthesis pathways that may be related to *N*-glycan synthesis.

## RT-PCR analysis of the co-transcription of *mmp1094* to *mmp1089*

*mmp1094* to *mmp1089* are orientated in the same direction on the complementary strand in the *M. maripaludis* genome, and opposite from the direction of the neighbouring genes *aglL* (*mmp1088*) and *mmp1095*, as shown in Fig. 1a. This six-gene cluster starts with *mmp1094* and relatively short intergenic regions separate the neighbouring genes: 71 bp between *mmp1094* and *mmp1093*, 73 bp between *mmp1092* and *mmp1091*, 104 bp between *mmp1091* and *mmp1090*, and 25 bp between *mmp1090* and *mmp1089*. Interestingly, *mmp1093* and *mmp1092* are predicted to share a 4 bp overlap. These observations suggest that the six genes might be transcriptionally linked and this was examined by RT-PCR experiments. Using RNA extracted from Mm900 cells as template and subjected to a reverse transcriptase step (RT lanes), amplification products were obtained using primers that could amplify across the intergenic regions linking *mmp1094* to *mmp1093* (predicted size of amplicon is 442 bp), *mmp1093* to *mmp1092* (541 bp), *mmp1092* to *mmp1091* (628 bp), *mmp1091* to *mmp1090* (529 bp) and *mmp1090* to *mmp1089* (533 bp) (Fig. 1b). No amplification products were obtained from PCR reactions using the same purified RNA as template but omitting the reverse transcriptase step (RNA lanes), indicating that the purified RNA

**Table 3** Annotations of MMP1089–1094

	Annotation	Super-families (BLAST)	Conserved domains (BLAST)	Characterized homologues (Protein, species)	Identity and query coverage with characterized homologues (% , %)
MMP1089	Polysaccharide synthesis protein	MATE-like superfamily	RfbX	AgIR, <i>Hfx. volcanii</i>	26, 39
MMP1090	UDP-glucose 4-epimerase	SDR superfamily	WcaG (Nucleoside-diphosphate-sugar epimerases)	GalE, <i>Pyrobaculum calidifontis</i> GalE, <i>Pyrococcus horikoshii</i>	33, 95 32, 95
MMP1091	ADP-glucose pyrophosphorylase (ADPG-PPase) Or UDP-glucose pyrophosphorylase (UDPG-PPase)	Glycosyltransferase family A superfamily	N/A	UDPG-PPase (PFU1356), <i>Pyrococcus furiosus</i> GlmU (MJ1101), <i>Methanocaldococcus jannaschii</i> UDPG-PPase (UGP1), <i>Saccharomyces cerevisiae</i>	47, 98 35, 91 24, 42
MMP1092	auxin efflux carrier	Membrane transport protein superfamily	N/A	MleP, <i>Lactobacillus lactis</i> MleP, <i>Oenococcus oeni</i>	24, 53 22, 97
MMP1093 (CoaD)	4'-phosphopantetheine adenylyltransferase (PPAT)	Nucleotidyl transferase superfamily	N/A	PPAT (PAB0944), <i>Pyrococcus abyssi</i>	38, 95
MMP1094 (PpsA)	Phosphoenol-pyruvate synthase (PEPS)	Pyruvate phosphate dikinase (PPDK) superfamily, Pyruvate-kinase superfamily	PRK06464	PEPS (MTH1118), <i>Methanothermobacter thermautotrophicus</i> PEPS (TK1292), <i>Thermococcus kodakarensis</i> PEPS (PF0043), <i>P. furiosus</i> PEPS (PPS), <i>Thermoproteus tenax</i>	57, 99 55, 99 55, 99 42, 99

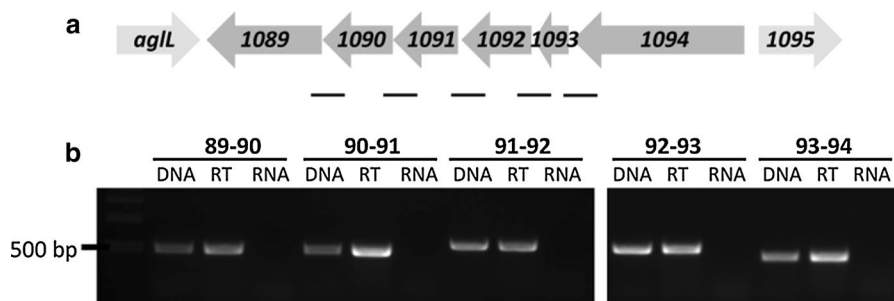
used for RT-PCR was not contaminated with genomic DNA. Using the same primer pairs and genomic DNA as template, PCR products (DNA lanes) were obtained to confirm the amplicon size and primer specificity. All amplicons were of the size predicted from the location of the primers. These results indicate that *mmp1094* to *mmp1089* form an operon.

#### Generation of *Δmmp1090*, *Δmmp1091* and *Δmmp1092* mutants

In order to determine the possible involvement of *mmp1089–1094* in the *N*-glycosylation pathway, each

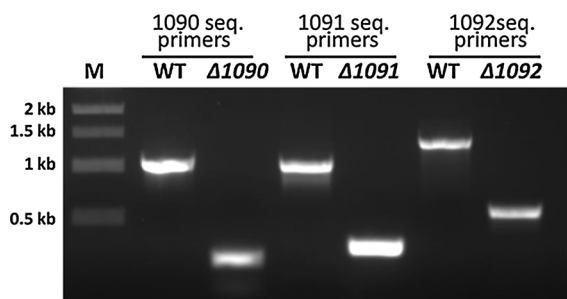
of the six genes was targeted for in-frame deletion. Mutants carrying deletions in each of *mmp1090*, *mmp1091* and *mmp1092* mutants were obtained (Fig. 2). In each case, PCR products, obtained using primers that would amplify across the targeted gene and with mutant cells as template, were smaller than those obtained when Mm900 genomic DNA was used as template, indicating the successful deletion of the corresponding gene. PCR products from each mutant were also sequenced to confirm that each deletion was in-frame. Several attempts to delete *mmp1093* and *mmp1094* were also performed but screening of over 100 transformants in different experiments by PCR





**Fig. 1** *mmp1094* to *mmp1089* are an operon. **a.** Genomic regions between *aglL* and *mmp1095* that were targeted for RT-PCR. Black lines below show the anticipated amplicons obtained from RT-PCR. **b.** RT-PCR confirmation of the co-transcription of *mmp1089–1094*. Amplicons were obtained from RT-PCR using primers amplifying intergenic regions between *mmp1089–1090*, *mmp1090–1091*, *mmp1091–1092*, *mmp1092–1093* and *mmp1093–1094*. No products were obtained from

reactions using the same primer pairs and RNA that had not been subjected to the reverse transcription step as template (RNA lanes), indicating that the RNA template used for RT-PCR was not contaminated with genomic DNA. PCR using the same primer pairs and Mm900 genomic DNA as template (DNA lanes) was conducted to show the amplicon size and primer specificity



**Fig. 2** PCR confirmation of the in-frame deletion of *mmp1090*, *mmp1091* and *mmp1092*. Sequencing primer pairs amplifying across the deletion area of *mmp1090*, *mmp1091* or *mmp1092* were used in PCR with either Mm900 genomic DNA (WT lanes) or corresponding washed deletion mutant cells as template. In all cases, the sizes of the amplicons were as predicted

failed to identify any potential deletion mutants. Attempts to delete *mmp1089*, examined as a possible flippase in previous studies, were also unsuccessful (Vandyke et al. 2009). The failure to obtain mutants in any of these three genes suggests that they might be essential for the survival of *M. maripaludis* under the laboratory culture conditions used.

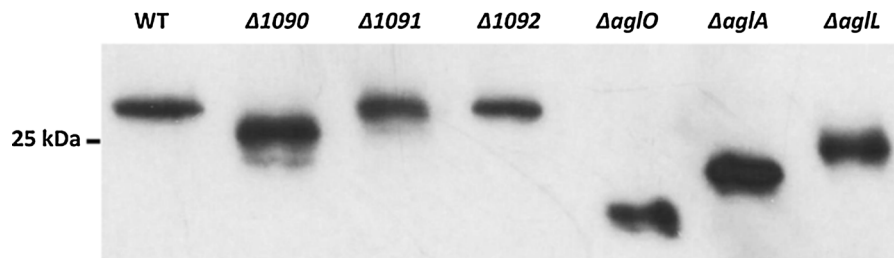
#### Western blot analysis of archaellin FlaB2 from *Δmmp1090*, *Δmmp1091* and *Δmmp1092* mutants

To test if the in-frame deletion of *mmp1090*, *mmp1091* or *mmp1092* resulted in a detectable truncation in the archaellin *N*-glycan, whole cell lysates of

each mutant were first subjected to Western blot analysis, using anti-FlaB2 antibodies, as it has been shown previously that even small truncations of the *N*-glycan results in a faster migration of FlaB2 that is detectable on Western blots (Ding et al. 2013; Jones et al. 2012; Siu et al. 2015; Vandyke et al. 2009). As shown in Fig. 3, FlaB2 from the *Δmmp1090* migrated faster than that from WT cells, indicating a possible truncation in its *N*-glycan. Using whole cell lysates from mutants deleted for *ΔaglO*, *ΔaglA* and *ΔaglL* as an indication of FlaB2 electrophoretic mobility corresponding to attached *N*-glycans missing three, two or one sugar(s), respectively, it was predicted that FlaB2 from *mmp1090* deleted cells would have a glycan lacking the terminal sugar. No reduction of FlaB2 apparent molecular weight was observed in strains deleted for either *mmp1091* or *mmp1092* in Western blots.

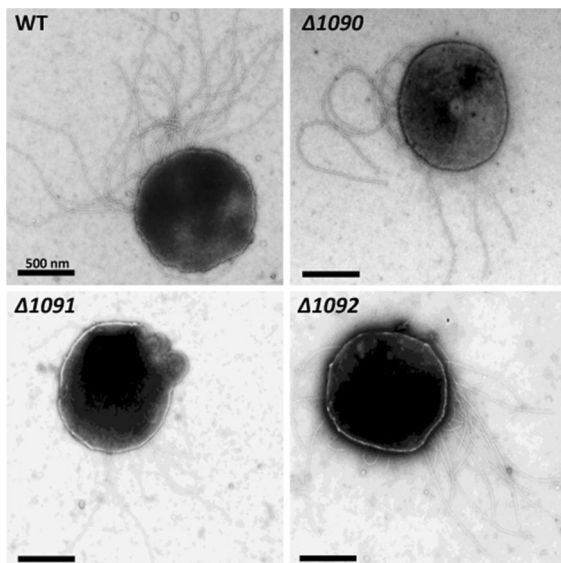
#### In-frame deletion of *mmp1090*, *mmp1091* or *mmp1092* does not interfere with archaella assembly

In *M. maripaludis*, a minimum length disaccharide glycan attached to archaellins is required for archaella formation (Vandyke et al. 2009). When *Δmmp1090*, *Δmmp1091* and *Δmmp1092* mutants were examined for archaella formation by electron microscopy, all mutants were found to be archaellated (Fig. 4). This is in agreement with the Western blot results suggesting a WT-sized glycan in the *Δmmp1091* and *Δmmp1092*



**Fig. 3** Western blot analysis of FlaB2 from *Δmmp1090*, *Δmmp1091* and *Δmmp1092* mutants. Whole cell lysates of Mm900 (WT), as well as the mutants *ΔaglO*, *ΔaglA* and *ΔaglL* (missing the 2nd, 3rd, or 4th GT, respectively) were included for comparison to the *Δmmp1090*, *Δmmp1091* and *Δmmp1092* mutants. The WT, *ΔaglO*, *ΔaglA* and *ΔaglL* strains synthesize FlaB2 with a tetra-, mono-, di-, or tri-saccharide *N*-glycan,

respectively. The electrophoretic mobility of FlaB2 from *Δmmp1090* was the same as that from *ΔaglL*, indicating that FlaB2 from the *Δmmp1090* mutant was likely modified with the truncated trisaccharide reported for the *ΔaglL* mutant. The electrophoretic mobilities of FlaB2 from the *Δmmp1091* and *Δmmp1092* mutants could not be distinguished from that of the WT cells



**Fig. 4** Electron micrographs of WT cells and the *Δmmp1090*, *Δmmp1091* and *Δmmp1092* mutants. Archaeella were observed on the cell surface of the WT (WT) and all three mutants. Bars, 500 nm

mutants and a three-sugar glycan for the *Δmmp1090* mutant.

Mass spectrometry analysis of *N*-glycan structures from *Δmmp1090*, *Δmmp1091* and *Δmmp1092* mutants

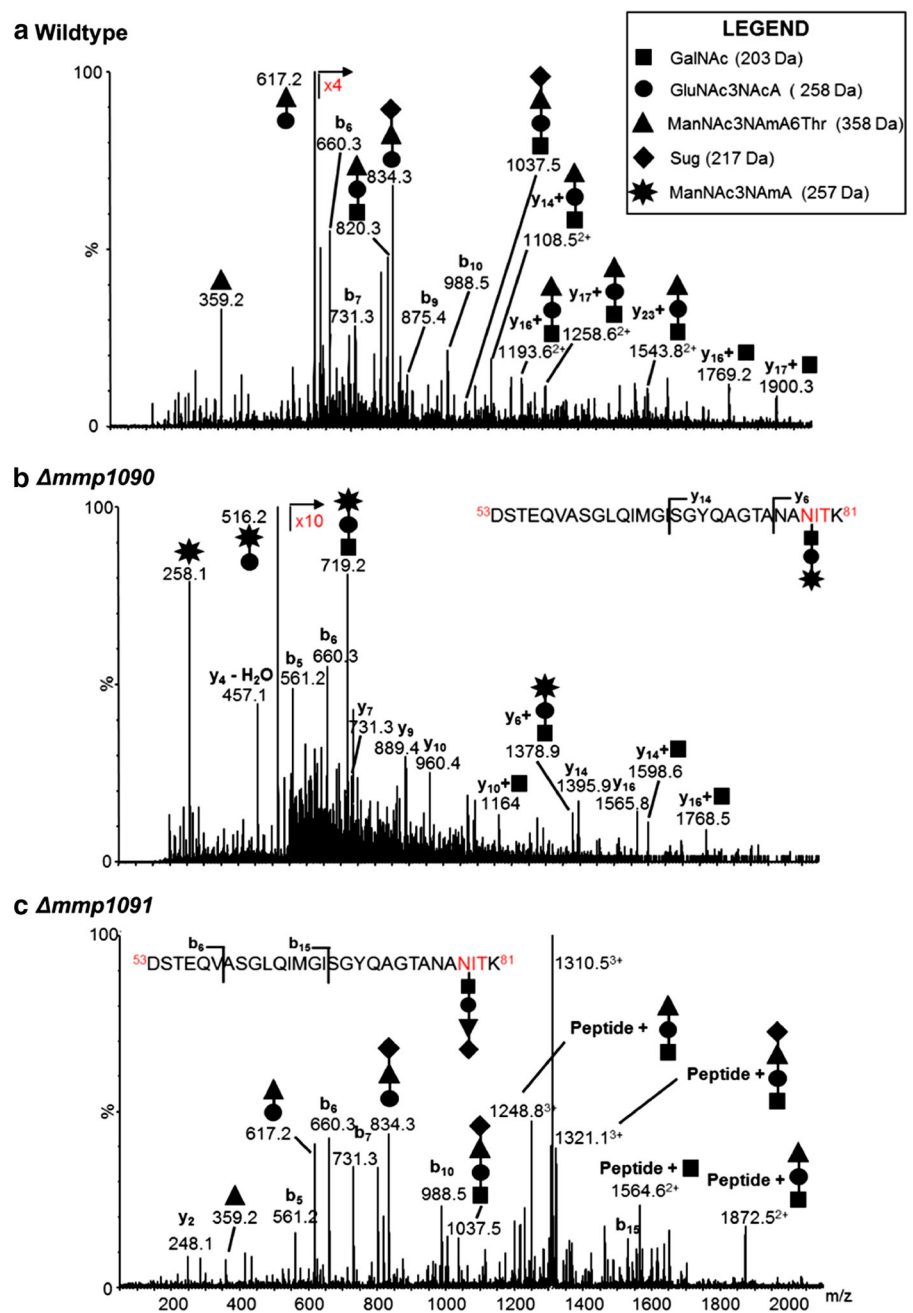
To specifically identify the *N*-glycan structure in each mutant, archaeella were isolated from the *Δmmp1090*, *Δmmp1091* and *Δmmp1092* cells and the attached glycan structure was determined by mass spectrometry

(Fig. 5). The structure of the tetrameric WT glycan (Fig. 5a) has been described previously (Kelly et al. 2009). The glycopeptides from the *Δmmp1090* mutant (Fig. 5b) are modified with a trisaccharide composed of the linking GalNAc, the di-*N*-acetyl glucuronic acid (GlcNAc3NAcA) and a third sugar (ManNAc3-NAmA) that lacks the threonine modification as well as the fourth sugar residue observed on WT glycan. The archaeallin glycopeptides from the *Δmmp1091* mutant are modified almost exclusively (>95 %) with WT glycan (Fig. 5c). The small differences observed (<5 %) could be attributed to side peaks/adducts and these were observed in both parent and mutant samples analysed. This was also true of the glycan structure in the archaeallins of the *Δmmp1092* mutant (data not shown).

Complementation of the *Δmmp1090* strain with WT and mutant versions of *mmp1090*

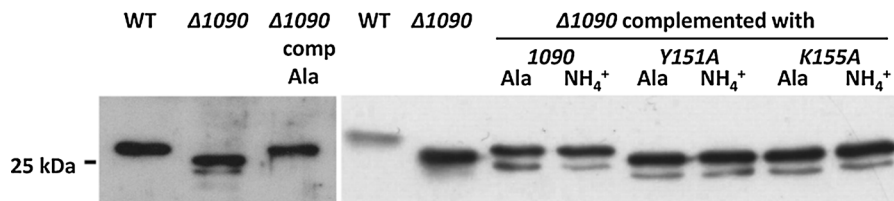
To examine if the in-frame deletion of *mmp1090* was the sole contributor to the defect in *N*-glycan length observed by mass spectrometry, the *Δmmp1090* mutant was complemented with a plasmid bearing a WT copy of *mmp1090* under the control of the *nif* promoter. Complemented cells were cultured in nitrogen-free medium supplemented with 10 mM of either L-alanine (*nif* promoter is induced) or NH<sub>4</sub>Cl (*nif* promoter is repressed) as the sole nitrogen source. FlaB2 from complemented cells cultured in both media was analysed by Western blotting (Fig. 6). Compared to FlaB2 from the *Δmmp1090* mutant cells, the apparent molecular weight of FlaB2 from

**Fig. 5** NanoLC-MS/MS analysis of the FlaB2 tryptic glycopeptide, T<sup>53–81</sup>. The tryptic digests of archaellin isolated from **a** WT as well as **b** *Δmmp1090* and **c** *Δmmp1091* mutant strains of *M. maripaludis* were analyzed by nanoLC-MS/MS on a Nanoaquity UPLC system (Waters) coupled to a Q-TOF Ultima mass spectrometer (Waters). The triply protonated glycopeptide ion (MH<sub>3</sub><sup>3+</sup>) was selected for MS/MS analysis in each case. The FlaB2 tryptic peptide T<sup>53–81</sup> contains one site of *N*-glycosylation. The sequence of this glycopeptide is provided in both panels **b** and **c** in order to illustrate the difference in the nature of the glycan modification observed in the two mutants. The major carbohydrate oxonium ions are identified in the MS/MS spectra using symbols to indicate the sugar residues present. The symbols are identified in the *inset* in panel **a**. The major *b* and *y* ions arising from fragmentation of the peptide bonds are also shown



*Δmmp1090*-complemented cells under alanine growth conditions was restored to that of the WT cells. Under NH<sub>4</sub>Cl growth conditions, the electrophoretic mobility of FlaB2 in the *Δmmp1090*-complemented cells was also restored to that from WT cells. Similar results have been observed in several previous studies using this vector system

(Ding et al. 2013, Jones et al. 2012), a result which we attribute to a small amount of transcription that can occur from the *nif* promoter even in the presence of NH<sub>4</sub>Cl (Lie et al. 2005). This basal expression under NH<sub>4</sub>Cl growth conditions may lead to the synthesis of enough MMP1090 to complement the defect in the deletion strain.



**Fig. 6** Western blot analysis of FlaB2 in lysates of the  $\Delta mmp1090$  mutant complemented *in trans* with WT or mutant versions of *mmp1090*. Both WT *mmp1090* (*1090*) and two *mmp1090* mutants, *Y151A* and *K155A*, were cloned in the shuttle vector pHW40 under an inducible *nif* promoter and transformed into  $\Delta mmp1090$  mutant cells. Complementation cells were cultured in nitrogen-free medium supplemented with either L-alanine (Ala) (promoter on) or  $\text{NH}_4\text{Cl}$  ( $\text{NH}_4^+$ )

conditions (promoter off). The left triplet of lanes shows that the complementation of the  $\Delta mmp1090$  mutant cells with the WT copy of *mmp1090* supplied *in trans* ( $\Delta mmp1090$  comp Ala) restored the FlaB2 apparent molecular weight to the WT size. The right group of lanes shows the apparent molecular weight of FlaB2 in the  $\Delta mmp1090$  cells complemented with the mutant versions of *mmp1090* was not returned to the size of FlaB2 found in WT cells

UDP-glucose 4-epimerase belongs to the Short-chain Dehydrogenases/Reductases (SDR) superfamily whose members typically contain a conserved YxxxK motif important for catalysis (Jörnvall et al. 1995). Two mutant versions of MMP1090, *Y151A* and *K155A*, where the Y151 and K155 of the conserved YxxxK motif were changed to alanine, were also used to complement the *mmp1090* deletion strain but neither could return the FlaB2 to WT size as determined by Western blots (Fig. 6), suggesting that MMP1090 lost its function due to the individual point mutation.

## Discussion

In this study, we have investigated the possible roles of a six-gene operon (*mmp1089–1094*) on the formation of the tetrasaccharide *N*-linked to archaellins. While deletions of *mmp1090*, *mmp1091* and *mmp1092* were obtained and studied, attempts to delete the remaining three genes were unsuccessful, suggesting that these three genes are essential under our normal growth conditions. This is consistent with the results from a recent genome-wide transposon mutagenesis study which also indicated that *mmp1089*, *mmp1093* and *mmp1094* are likely essential (Sarmiento et al. 2013).

We have examined close relatives of *M. maripaludis* S2 for the presence and order of genes homologous to *mmp1089–1094*. The complete genome sequences of several strains of *M. maripaludis* (C5, C6, C7, X1) are available for comparison. In all four of these other strains, genes homologous to *mmp1090–1094* are found adjacent to each other and

in the same order as in the S2 strain. However, only in strains C7 and X1 is the gene homologous to *mmp1089* found adjacent to the *mmp1090* homologue. No *N*-linked glycan structures have been reported in these additional *M. maripaludis* strains for comparison to that from the S2 strain. Other species of *Methanococcus* were also examined and a variety of formats were observed with regards to the *mmp1089–1094* homologues. *Methanococcus vannielii* has a complete set of *mmp1089–1094* homologues in the same order as in *M. maripaludis* S2. *Methanococcus aeolicus* has the *mmp1090–1091* homologues adjacent but the other genes are located around the genome. *Methanococcus voltae* is unusual in not having the *mmp1090–1091* homologues adjacent to each other but the *mmp1092–1094* homologs are clustered.

The annotations of *mmp1093* and *mmp1094* suggest that both gene products are likely to be involved in intermediary metabolism and not *N*-glycosylation (Table 3). MMP1093 is annotated as phosphopantetheine adenylyltransferase (PPAT, encoded by *coaD*), which catalyses the penultimate step of the CoA biosynthesis pathway, i.e. the reversible adenylation of 4'-phosphopantetheine to generate 3'-dephospho-CoA (Geerloff et al. 1999). CoA is an essential cofactor for many enzymatic reactions and genes encoding the last 4 steps of the CoA biosynthesis pathway, sequentially *coaB*, *coaC*, *coaD* and *coaE* (sometimes *coaB* and *coaC* are fused into one gene *coaBC* encoding a bifunctional protein) are found throughout archaeal genomes, indicating the last 4 steps of the CoA biosynthesis pathway are conserved among all the three domains of life (Genschel 2004; Kupke and Schwarz 2006). *coaBC* (*mmp1606*), *coaD* (*mmp1093*)

and *coaE* (*mmp1282*) are not grouped in the *M. maripaludis* S2 genome and all are essential according to a recent transposon mutagenesis study (Sarmiento et al. 2013). MMP1094 is annotated as a phosphoenolpyruvate synthase (PPS) or pyruvate, water dikinase (EC number 2.7.9.2), encoded by *ppsA*, which catalyses the conversion of pyruvate to phosphoenolpyruvate. PPS activity was previously detected in crude cell extracts of *M. maripaludis* S2, although experiments that formally confirm that this activity is due to the product of *mmp1094* are lacking.

MMP1089 is annotated as a polysaccharide synthesis protein. According to BLAST (Basic Local Alignment Search Tool), it belongs to the MATE-like superfamily with a RfbX (Wzx) domain (Table 3). Wzx is considered to be the translocase (flippase) in the Wzy (polymerase)-dependent O-antigen biosynthesis pathway, flipping the isoprenoid lipid-linked O-antigenic unit across the cytoplasmic membrane in Gram-negative bacteria (Islam and Lam 2013; Liu et al. 1996). MMP1089 is an integral membrane protein and predicted to contain 14 transmembrane helices (TMHMM, <http://www.cbs.dtu.dk/services/TMHMM/>; Sonnhammer et al. 1998), sharing similar protein topology with various Wzx proteins from *Salmonella enterica*, *E. coli* and *Pseudomonas aeruginosa* (Cunneen and Reeves 2008; Islam et al. 2010; Marolda et al. 2010). MMP1089 also shares 26 % identity (Table 3) with AglR from *Hfx. volcanii*, the only putative flippase so far identified in archaeal *N*-glycosylation pathways. AglR is thought to be the enzyme that flips Dol-P-mannose across the cytoplasmic membrane; the mannose is subsequently transferred onto the protein-bound *N*-glycan as the terminal sugar of the pentasaccharide that is *N*-linked to the S-layer protein (Eichler 2013; Kaminski et al. 2012). Both its annotation and its genomic location immediately adjacent to sugar biosynthetic and GT genes known to be involved in *N*-glycosylation suggest a role for MMP1089 in *N*-glycosylation. In our previous studies, *mmp1089* had been targeted for deletion in the belief that it may encode the flippase involved in the *N*-linked glycosylation pathway, but none of these attempts were successful (Vandyke et al. 2009). The subsequent transposon mutagenesis study from the Whitman group also suggests the gene is essential (Sarmiento et al. 2013). This is somewhat puzzling since the *N*-glycosylation pathway is not essential in *M. maripaludis* as evidenced by the

deletion of *aglB*, encoding the OST which catalyses the critical terminal step in the pathway (Vandyke et al. 2009). It may be that deletion of the flippase leads to a sequestering of the dolichol carrier that prevents its turnover and use in an essential pathway, as suggested for the putative Wzx-like flippase involved in capsule formation in *Streptococcus pneumoniae* (Xayarath and Yother 2007). Otherwise, MMP1089 may be involved in other, presumably essential, processes in addition to its predicted role as a flippase in *N*-glycosylation.

Successful deletion of the remaining three genes of the operon (*mmp1090*, *mmp1091* and *mmp1092*) was accomplished, although only one had a demonstrable involvement in *N*-glycosylation. Inactivation of *mmp1090* resulted in a truncated archaeellin *N*-glycan missing the terminal sugar residue and the threonine attached to the third sugar residue, the same archaeellin glycan structure found in a mutant deleted for the 4<sup>th</sup> GT, AglL (Vandyke et al. 2009). As it is known that the threonine residue is transferred onto the third sugar by the threonine transferase AglU only after AglL transfers the terminal sugar to the glycan being assembled on the Dol-P carrier (Ding et al. 2013; Jarrell et al. 2014), MMP1090 is most likely to be involved in the biosynthesis of the unique terminal sugar.

MMP1090 is annotated as a UDP-glucose 4-epimerase (Table 3). BLAST searches reveal numerous proteins, described as UDP-glucose-4-epimerases, in both Archaea and Bacteria that have very high sequence identity to MMP1090 over essentially the entire length of the protein. Among these homologues, MMP1090 shares 33 and 32 % amino acid identity with two studied archaeal UDP-glucose 4-epimerases, from the hyperthermophiles *Pyrobaculum calidifontis* (Sakuraba et al. 2011) and *Pyrococcus horikoshii* (Chung et al. 2012). In support of the annotation, several signature motifs or amino acid residues possessed by UDP-glucose 4-epimerases are also found in MMP1090, including an YxxxK motif and a glycine-rich motif (Kallberg et al. 2002; Persson and Kallberg 2013). The tyrosine and lysine in the YxxxK motif, together with a conserved upstream serine residue, establish a catalytic S-Y-K triad (Jörnvall et al. 1995; Oppermann et al. 2003). The glycine rich motif, which is located in the N-terminus of the protein, is important for the binding of NAD<sup>+</sup> as cofactor (Jörnvall et al. 1995). UDP-glucose



4-epimerases are members of the extended family of SDRs and MMP1090 possesses a perfect match (TGGAGFIGSHIVDMLIENGHDV) to the conserved glycine rich motif [ST]G<sub>x2</sub>G[FMQY][DILV]G<sub>x6</sub>[FILMV][ILMV]<sub>x2</sub>G<sub>x2</sub>[ILV] of this family (Kallberg et al. 2002).

Based on their substrate specificity, the UDP-glucose 4-epimerase family has been subdivided into three groups (Ishiyama et al. 2004). In the first group, UDP-glucose 4-epimerase catalyses the interconversion of UDP-glucose and UDP-galactose, e.g. GalE from *E. coli* (eGalE hereafter). A second group of UDP-glucose 4-epimerase, e.g. human UDP-glucose 4-epimerase (hGalE hereafter), catalyses not only the interconversion of UDP-glucose and UDP-galactose, but also the interconversion of UDP-*N*-acetyl-glucosamine (UDP-GlcNAc) and UDP-*N*-acetyl-galactosamine (UDP-GalNAc). The third group, e.g. WbpP from *P. aeruginosa*, preferably catalyses the interconversion of acetylated UDP-hexoses (UDP-GlcNAc and UDP-GalNAc). After comparing protein structures from the three groups of UDP-glucose 4-epimerase, Ishiyama et al. (Ishiyama et al. 2004) proposed that the differences in the catalytic pockets lead to the different substrate specificities. The catalytic pocket from hGalE is ~15 % larger than that from eGalE due to the smaller side chain from N207 and C307 in hGalE compared with that from the corresponding N198 and Y299 in eGalE; thus the steric hindrance prevents the latter enzyme from catalysing the acetylated UDP-hexoses. The two corresponding amino acids from WbpP, A208 and S306, also lead to a larger catalytic pocket. Furthermore, the existence of ordered solvent molecules in the WbpP catalytic pocket results in the preference for acetylated UDP-hexoses as substrates. In MMP1090, BLAST analysis with eGalE, hGalE and WbpP shows the corresponding amino acids are G193 and I281. The short side chains from these two amino acids presumably result in a relatively larger catalytic pocket, indicating that MMP1090 might belong to either group 2 or group 3 UDP-glucose 4-epimerase.

We initially hypothesised that *M. maripaludis* would need a UDP-GlcNAc 4-epimerase to produce UDP-GalNAc from UDP-GlcNAc. UDP-GalNAc is likely the substrate for the first GT to begin the assembly of the tetrasaccharide on the dolichol phosphate carrier, since the linking sugar in the archaeellin *N*-glycan is GalNAc. There is no gene

annotated as a UDP-GlcNAc 4-epimerase in the *M. maripaludis* genome sequence but there are examples of UDP-glucose 4-epimerases also having UDP-GlcNAc 4-epimerase activity, such as the bifunctional GalE involved in the synthesis of the *N*-linked heptasaccharide of *Campylobacter jejuni*, which contains 5 GalNAc residues (Bernatchez et al. 2005). If MMP1090 has this activity and it is necessary for the production of UDP-GalNAc, then its deletion should result in completely non-glycosylated archaeallins, migrating with the same apparent molecular weight as those from the *aglB* (oligosaccharyltransferase) mutant. Clearly this was not the case and the MMP1090 role appears to be in the synthesis of the terminal sugar, even though our examination of the pocket size of MMP1090 suggested it might be able to utilise acetylated UDP-hexoses as substrates.

Complementation studies demonstrated that the electrophoretic mobility of FlaB2 in *Δmmp1090* mutant cells expressing the WT version of *mmp1090 in trans* was indistinguishable from that of FlaB2 from WT cells in Western blots, indicating that MMP1090 is the sole contributor to the loss of the terminal sugar residue of the archaeellin *N*-glycan (the missing threonine is transferred onto the *N*-glycan precursor afterwards (Ding et al. 2013; Vandyke et al. 2009)). In addition, two mutant versions of *mmp1090* were created which resulted in a change of one amino acid, either Y151 or K155, located in the conserved YxxxK motif, to alanine. The corresponding amino acids in hGalE, Y157 and K161, are two of the key amino acids anchoring the cofactor NAD<sup>+</sup> within the enzyme. In addition, Y157 also serves as the active site base by directly interacting with the C4 hydroxyl group in the glucosyl moiety (Thoden et al. 2000). In agreement with our hypothesis that Y151 and K155 are key amino acids in MMP1090, neither of the two mutant versions of MMP1090 could restore the FlaB2 of the *mmp1090* deletion strain to WT size in complementation studies (Fig. 6). **These findings are consistent with MMP1090 being a UDP-glucose 4-epimerase involved in the terminal sugar biosynthesis pathway of the archaeellin *N*-glycan.**

No detectable phenotype related to archaeellin glycosylation was observed in either the *Δmmp1091* or *Δmmp1092* deletion mutants, as the electrophoretic mobility of FlaB2 in Western blots, the archaeellin *N*-glycan structure determined by mass spectrometry and the archaeellation state determined by electron



microscopy all appeared identical to those observed in WT cells. The annotation of MMP1091 indicates that it is an ADP-glucose pyrophosphorylase (ADPG-PPase), while BLAST results show that it is a putative UDP-glucose pyrophosphorylase (UDPG-PPase). ADPG-PPase and UDPG-PPase catalyse similar reactions, i.e. the synthesis of ADP-glucose or UDP-glucose from glucose-1-phosphate and either ATP (ADPG-PPase) or UTP (UDPG-PPase), respectively. ADP-glucose and UDP-glucose can be used as glucosyl donors by glycogen or starch synthases in the biosynthesis of glycogen and starch, which are common energy storage forms in the three domains of life (Henrissat et al. 2002). Archaeal glycogen synthases can use both ADP-glucose and UDP-glucose as donor substrates (Gruyer et al. 2002; Horcajada et al. 2006). In *M. maripaludis*, glycogen was reported to comprise 0.34 % of the cell dry weight in the early stationary phase (Yu et al. 1994). A BLAST search of the *M. maripaludis* S2 genome using MMP1091 as query revealed a second protein, MMP1076, as 37 % identical with a query cover of 91 % and an E value of  $1e-26$ . This protein is annotated as a glucosamine-1-phosphate *N*-acetyltransferase. It is a homologue of the *E. coli* enzyme GlmU, which is a bifunctional enzyme that catalyses the acetylation of glucosamine 1-phosphate and the subsequent transfer of the sugar to UTP. The purified *Methanocaldococcus jannaschii* homologue, MJ1101, also shows both of these activities (Namboori and Graham 2008b) while other archaeal homologues, such as from *Sulfolobus tokodaii*, lack the acetyltransferase activity (Zhang et al. 2005). Interestingly, MJ1101 as well the homologues from *S. tokodaii* and *P. furiosus*, could all utilise glucose-1-phosphate as an additional substrate to generate UDP-glucose (Mizanur et al. 2004; Namboori and Graham 2008b; Zhang et al. 2005). Given that MJ1101 is 68 % identical to MMP1076, it seems likely that the *M. maripaludis* enzyme may also be able to activate glucose-1-phosphate and thus compensate for the loss of MMP1091, although this has not been tested (Namboori and Graham 2008b). Although no archaeellin *N*-glycan deficiency was observed in the *Δmmp1091* mutant, it is premature to conclude that MMP1091 is not involved in the *N*-glycosylation pathway considering its annotation and its genomic position in an operon adjacent to *mmp1090*, as well as the possibility of the lost activity being compensated by MMP1076. In the case of *H. volcanii*, formation of

the hexuronic acid found at position 3 of the pentasaccharide *N*-linked to the S-layer protein does depend on the activity of a glucose-1-phosphate uridylyltransferase (AglF), which acts in concert with the dehydrogenase AglM (Yurist-Doutsch et al. 2008).

Deletion of *mmp1092* also did not have a detectable effect on the archaeellin *N*-glycan. MMP1092 is predicted to be an integral membrane protein containing 10 transmembrane helices, according to bioinformatics tools such as TMHMM and PSORTb (<http://www.psорт.org/psорт/>). While MMP1092 is annotated as an auxin efflux protein, members of this family have also been found in both bacteria and archaea. In bacteria, there are reports of these homologues being malate transporters (TCDB: Transport Classification DataBase; <http://www.tcdb.org/>). Unlike the case with MMP1091, there are no other proteins encoded by *M. maripaludis* S2 with high sequence identity to MMP1092 over a large portion of the protein. Since no detectable phenotype was observed in the *Δmmp1092* mutant, further work is needed to identify the function of MMP1092 but at present no role in *N*-glycosylation can be assigned to it.

The *mmp1089–mmp1094* operon appears to be the end of the large number of genes involved in *N*-glycosylation that are found in this region. Examination of the annotations of genes *mmp1095* and downstream display a number of genes predicted to be involved in ion transport and other functions seemingly unrelated to *N*-glycosylation.

In this study, *mmp1090*, likely encoding a UDP-glucose 4-epimerase, was identified by genetic and mass spectrometry techniques to be involved in the biosynthesis pathway of the terminal sugar residue in the archaeellin *N*-glycan. The terminal sugar in the archaeellin *N*-linked tetrasaccharide, with the structure of (5*S*)-2-acetamido-2,4-dideoxy-5-*O*-methyl- $\alpha$ -L-erythro-hexos-5-ulo-1,5-pyranose, is a unique sugar so far found exclusively in *M. maripaludis* (Kelly et al. 2009). The knowledge of the biosynthesis of this dialdose is very limited and only one protein, AglV, the enzyme responsible for the transfer of the methyl group onto C5 via *O*-linkage, has been genetically identified in its biosynthesis pathway (Ding et al. 2013). This is in contrast to the second and third sugars of the glycan where the biosynthesis pathways are now well known (Ding et al. 2013; Jones et al. 2012; Siu et al. 2015; Vandyke et al. 2009). The involvement of MMP1090 as a UDP-glucose

4-epimerase in the biosynthesis pathway of the terminal sugar provides evidence that this unique sugar is probably synthesised from UDP-glucose or UDP-GlcNAc. We are currently attempting heterologous overexpression and purification of the active enzyme in *E. coli* to identify the substrate specificity of this enzyme to help in the elucidation of its role in the biosynthetic pathway of this unusual sugar. Based on this demonstrated involvement in the *N*-glycosylation pathway, *mmp1090* is here designated as *aglW*, in keeping with the nomenclature scheme for genes involved in archaeal *N*-glycosylation (Chaban et al. 2006; Eichler et al. 2013).

**Acknowledgments** This work was funded by the National Research Council of Canada (SML, JFK) and by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) (to KFJ). Y.D. is sponsored by China Scholarship Council (2010622028).

## References

- Aebi M (2013) N-linked protein glycosylation in the ER. *Biochim Biophys Acta* 1833:2430–2437
- Albers SV, Jarrell KF (2015) The archaeellum: how archaea swim. *Front Microbiol* 6:23
- Balch WE, Fox GE, Magrum LJ, Woese CR, Wolfe RS (1979) Methanogens: reevaluation of a unique biological group. *Microbiol Rev* 43:260–296
- Bardy SL, Jarrell KF (2003) Cleavage of preflagellins by an aspartic acid signal peptidase is essential for flagellation in the archaeon *Methanococcus voltae*. *Mol Microbiol* 50:1339–1347
- Bardy SL, Mori T, Komoriya K, Aizawa S, Jarrell KF (2002) Identification and localization of flagellins FlaA and FlaB3 within flagella of *Methanococcus voltae*. *J Bacteriol* 184:5223–5233
- Bernatchez S, Szymanski CM, Ishiyama N, Li J, Jarrell HC, Lau PC, Berghuis AM, Young NM, Wakarchuk WW (2005) A single bifunctional UDP-GlcNAc/Glc 4-epimerase supports the synthesis of three cell surface glycoconjugates in *Campylobacter jejuni*. *J Biol Chem* 280:4792–4802
- Chaban B, Voisin S, Kelly J, Logan SM, Jarrell KF (2006) Identification of genes involved in the biosynthesis and attachment of *Methanococcus voltae* N-linked glycans: insight into N-linked glycosylation pathways in Archaea. *Mol Microbiol* 61:259–268
- Chaban B, Ng SY, Kanbe M, Saltzman I, Nimmo G, Aizawa SI, Jarrell KF (2007) Systematic deletion analyses of the *fla* genes in the flagella operon identify several genes essential for proper assembly and function of flagella in the archaeon, *Methanococcus maripaludis*. *Mol Microbiol* 66:596–609
- Chung SK, Ryu SI, Lee SB (2012) Characterization of UDP-glucose 4-epimerase from *Pyrococcus horikoshii*: regeneration of UDP to produce UDP-galactose using two-enzyme system with trehalose. *Bioresour Technol* 110:423–429
- Cohen-Rosenzweig C, Yurist-Doutsch S, Eichler J (2012) AglS, a novel component of the *Haloferax volcanii* N-glycosylation pathway, is a dolichol phosphate-mannose mannosyltransferase. *J Bacteriol* 194:6909–6916
- Cunneen MM, Reeves PR (2008) Membrane topology of the *Salmonella enterica* serovar Typhimurium Group B O-antigen translocase Wzx. *FEMS Microbiol Lett* 287:76–84
- Ding Y, Jones GM, Uchida K, Aizawa SI, Robotham A, Logan SM, Kelly J, Jarrell KF (2013) Identification of genes involved in the biosynthesis of the third and fourth sugars of the *Methanococcus maripaludis* archaeallin N-linked tetrasaccharide. *J Bacteriol* 195:4094–4104
- Ding Y, Uchida K, Aizawa SI, Murphy K, Berezuk A, Khursigara CM, Chong JPI, Jarrell KF (2015) Effects of *N*-glycosylation site removal in archaeallins on the assembly and function of archaeella in *Methanococcus maripaludis*. *PLoS One* 10:e0116402
- Eichler J (2013) Extreme sweetness: protein glycosylation in Archaea. *Nature Rev Microbiol* 11:151–156
- Eichler J, Jarrell K, Albers S (2013) A proposal for the naming of N-glycosylation pathway components in Archaea. *Glycobiology* 23:620–621
- Geerloff A, Lewendon A, Shaw WV (1999) Purification and characterization of phosphopantetheine adenylyltransferase from *Escherichia coli*. *J Biol Chem* 274:27105–27111
- Genschel U (2004) Coenzyme A biosynthesis: reconstruction of the pathway in archaea and an evolutionary scenario based on comparative genomics. *Mol Biol Evol* 21:1242–1251
- Gruyer S, Legin E, Bliard C, Ball S, Duchiron F (2002) The endopolysaccharide metabolism of the hyperthermophilic archaeon *Thermococcus hydrothermalis*: polymer structure and biosynthesis. *Curr Microbiol* 44:206–211
- Henrissat B, Deleury E, Coutinho PM (2002) Glycogen metabolism loss: a common marker of parasitic behaviour in bacteria? *Trends Genet* 18:437–440
- Horcajada C, Guinovart JJ, Fita I, Ferrer JC (2006) Crystal structure of an archaeal glycogen synthase: insights into oligomerization and substrate binding of eukaryotic glycogen synthases. *J Biol Chem* 281:2923–2931
- Igura M, Maita N, Obita T, Kamishikiryo J, Maenaka K, Kohda D (2007) Purification, crystallization and preliminary X-ray diffraction studies of the soluble domain of the oligosaccharyltransferase STT3 subunit from the thermophilic archaeon *Pyrococcus furiosus*. *Acta Crystallogr, Sect F: Struct Biol Cryst Commun* 63:798–801
- Ishiyama N, Creuzenet C, Lam JS, Berghuis AM (2004) Crystal structure of WbpP, a genuine UDP-N-acetylglucosamine 4-epimerase from *Pseudomonas aeruginosa*: substrate specificity in udp-hexose 4-epimerases. *J Biol Chem* 279:22635–22642
- Islam ST, Lam JS (2013) Wzx flippase-mediated membrane translocation of sugar polymer precursors in bacteria. *Environ Microbiol* 15:1001–1015
- Islam ST, Taylor VL, Qi M, Lam JS (2010) Membrane topology mapping of the O-antigen flippase (Wzx), polymerase

- (Wzy), and ligase (WaaL) from *Pseudomonas aeruginosa* PAO1 reveals novel domain architectures. *Mbio* 1:3
- Jarrell KF, Albers SV (2012) The archaeallum: an old motility structure with a new name. *Trends Microbiol* 20:307–312
- Jarrell KF, Jones GM, Kandiba L, Nair DB, Eichler J (2010) S-layer glycoproteins and flagellins: reporters of archaeal posttranslational modifications. *Archaea*. doi:10.1155/2010/612948
- Jarrell KF, Ding Y, Meyer BH, Albers SV, Kaminski L, Eichler J (2014) N-Linked glycosylation in Archaea: a structural, functional, and genetic analysis. *Microbiol Mol Biol Rev* 78:304–341
- Jones GM, Wu J, Ding Y, Uchida K, Aizawa S, Robotham A, Logan SM, Kelly J, Jarrell KF (2012) Identification of genes involved in the acetamidino group modification of the flagellin N-linked glycan of *Methanococcus maripaludis*. *J Bacteriol* 194:2693–2702
- Jörnvall H, Persson B, Krook M, Atrian S, González-Duarte R, Jeffery J, Ghosh D (1995) Short-chain dehydrogenases/reductases (SDR). *Biochemistry* 34:6003–6013
- Kallberg Y, Oppermann U, Jörnvall H, Persson B (2002) Short-chain dehydrogenases/reductases (SDRs). *Eur J Biochem* 269:4409–4417
- Kaminski L, Guan Z, Abu-Qarn M, Konrad Z, Eichler J (2012) AgIR is required for addition of the final mannose residue of the N-linked glycan decorating the *Haloferax volcanii* S-layer glycoprotein. *Biochim Biophys Acta* 1820:1664–1670
- Kaminski L, Lurie-Weinberger MN, Allers T, Gophna U, Eichler J (2013) Phylogenetic- and genome-derived insight into the evolution of N-glycosylation in Archaea. *Mol Phylogenet Evol* 68:327–339
- Kelly J, Logan SM, Jarrell KF, Vandyke DJ, Vinogradov E (2009) A novel N-linked flagellar glycan from *Methanococcus maripaludis*. *Carbohydr Res* 344:648–653
- Kupke T, Schwarz W (2006) 4'-phosphopantetheine biosynthesis in Archaea. *J Biol Chem* 281:5435–5444
- Lie TJ, Wood GE, Leigh JA (2005) Regulation of nif expression in *Methanococcus maripaludis*: roles of the euryarchaeal repressor NrpR, 2-oxoglutarate, and two operators. *J Biol Chem* 280:5236–5241
- Liu D, Cole RA, Reeves PR (1996) An O-antigen processing function for Wzx (RfbX): a promising candidate for O-unit flippase. *J Bacteriol* 178:2102–2107
- Marolda CL, Li B, Lung M, Yang M, Hanuszkiewicz A, Rosales AR, Valvano MA (2010) Membrane topology and identification of critical amino acid residues in the Wzx O-antigen translocase from *Escherichia coli* O157:H4. *J Bacteriol* 192:6160–6171
- Matsumoto S, Shimada A, Kohda D (2013) Crystal structure of the C-terminal globular domain of the third paralog of the *Archaeoglobus fulgidus* oligosaccharyltransferases. *BMC Struct Biol* 13:11
- Mizanur RM, Zea CJ, Pohl NL (2004) Unusually broad substrate tolerance of a heat-stable archaeal sugar nucleotidyltransferase for the synthesis of sugar nucleotides. *J Am Chem Soc* 126:15993–15998
- Moore BC, Leigh JA (2005) Markerless mutagenesis in *Methanococcus maripaludis* demonstrates roles for alanine dehydrogenase, alanine racemase, and alanine permease. *J Bacteriol* 187:972–979
- Nair DB, Uchida K, Aizawa SI, Jarrell KF (2014) Genetic analysis of a type IV pili-like locus in the archaeon *Methanococcus maripaludis*. *Arch Microbiol* 196:179–191
- Namboori SC, Graham DE (2008a) Enzymatic analysis of uridine diphosphate N-acetyl-D-glucosamine. *Anal Biochem* 301:94–100
- Namboori SC, Graham DE (2008b) Acetamido sugar biosynthesis in the Euryarchaea. *J Bacteriol* 190:2987–2996
- Ng SYM, Wu J, Nair DB, Logan SM, Robotham A, Tessier L, Kelly JF, Uchida K, Aizawa S, Jarrell KF (2011) Genetic and mass spectrometry analysis of the unusual type IV-like pili of the archaeon *Methanococcus maripaludis*. *J Bacteriol* 193:804–814
- Nothhaft H, Szymanski CM (2010) Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* 8:765–778
- Nothhaft H, Szymanski CM (2013) Bacterial protein N-glycosylation: new perspectives and applications. *J Biol Chem* 288:6912–6920
- Oppermann U, Filling C, Hult M, Shafqat N, Wu X, Lindh M, Shafqat J, Nordling E, Kallberg Y, Persson B, Jörnvall H (2003) Short-chain dehydrogenases/reductases (SDR): the 2002 update. *Chem Biol Interact* 143–144:247–253
- Persson B, Kallberg Y (2013) Classification and nomenclature of the superfamily of short-chain dehydrogenases/reductases (SDRs). *Chem Biol Interact* 202:111–115
- Sakuraba H, Kawai T, Yoneda K, Ohshima T (2011) Crystal structure of UDP-galactose 4-epimerase from the hyperthermophilic archaeon *Pyrobaculum calidifontis*. *Arch Biochem Biophys* 512:126–134
- Sarmiento F, Mrázek J, Whitman WB (2013) Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proc Natl Acad Sci U S A* 110:4726–4731
- Siu S, Robotham A, Logan SM, Kelly JF, Uchida K, Aizawa SI, Jarrell KF (2015) Evidence that biosynthesis of the second and third sugars of the archaeal tetrasaccharide in the archaeon *Methanococcus maripaludis* occurs by the same pathway used by *Pseudomonas aeruginosa* to make a di-N-acetylated sugar. *J Bacteriol* 197:1668–1680
- Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 6:175–182
- Thoden JB, Wohlers TM, Fridovich-Keil JL, Holden HM (2000) Crystallographic evidence for Tyr 157 functioning as the active site base in human UDP-galactose 4-epimerase. *Biochemistry* 39:5691–5701
- Towbin H, Staehelin T, Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A* 76:4350–4354
- Tumbula DL, Makula RA, Whitman WB (1994) Transformation of *Methanococcus maripaludis* and identification of a *Pst*I-like restriction system. *FEMS Microbiol Lett* 121:309–314
- VanDyke DJ, Wu J, Ng SY, Kanbe M, Chaban B, Aizawa SI, Jarrell KF (2008) Identification of putative acetyltransferase gene, MMP0350, which affects proper assembly of both flagella and pili in the archaeon *Methanococcus maripaludis*. *J Bacteriol* 190:5300–5307
- Vandyke DJ, Wu J, Logan SM, Kelly JF, Mizuno S, Aizawa SI, Jarrell KF (2009) Identification of genes involved in the

- assembly and attachment of a novel flagellin N-linked tetrasaccharide important for motility in the archaeon *Methanococcus maripaludis*. *Mol Microbiol* 72:633–644
- Xayarath B, Yother J (2007) Mutations blocking side chain assembly, polymerization, or transport of a Wzy-dependent *Streptococcus pneumoniae* capsule are lethal in the absence of suppressor mutations and can affect polymer transfer to the cell wall. *J Bacteriol* 189:3369–3381
- Yu JP, Ladapo J, Whitman WB (1994) Pathway of glycogen metabolism in *Methanococcus maripaludis*. *J Bacteriol* 176:325–332
- Yurist-Doutsch S, Abu-Qarn M, Battaglia F, Morris HR, Hitchen PG, Dell A, Eichler J (2008) AglF, aglG and aglI, novel members of a gene island involved in the N-glycosylation of the *Haloflex volcanii* S-layer glycoprotein. *Mol Microbiol* 69:1234–1245
- Zhang Z, Tsujimura M, Akutsu J, Sasaki M, Tajima H, Kawarabayashi Y (2005) Identification of an extremely thermostable enzyme with dual sugar-1-phosphate nucleotidyltransferase activities from an acidothermophilic archaeon, *Sulfolobus tokodaii* strain 7. *J Biol Chem* 280:9698–9705