Identification of a novel N-linked glycan on the archaellins and S-layer protein of the thermophilic methanogen, *Methanothermococcus thermolithotrophicus*

John F. Kelly1*, Evgeny Vinogradov1, Jacek Stupak1, Anna C. Robotham1, Susan M. Logan1, Alison Berezuk2*, Cezar M. Khursigara3*, and Ken F. Jarrell1,*

From the 1Human Health Therapeutics Research Centre, National Research Council of Canada, Ottawa, Ontario, Canada, the 2Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario, Canada, and the 3Department of Biomedical and Molecular Sciences, Queen’s University, Kingston, Ontario, Canada

Edited by Gerald W. Hart

Motility in archaea is facilitated by a unique structure termed the archaellum. N-Glycosylation of the major structural proteins (archaellins) is important for their subsequent incorporation into the archaellum filament. The identity of some of these N-glycans has been determined, but archaea exhibit extensive variation in their glycans, meaning that further investigations can shed light not only on the specific details of archaellin structure and function, but also on archaean glycobiology in general. Here we describe the structural characterization of the N-linked glycan modifications on the archaellins and S-layer protein of *Methanothermococcus thermolithotrophicus*, a methanogen that grows optimally at 65 °C. SDS-PAGE and MS analysis revealed that the sheared archaella are composed principally of two of the four predicted archaellins, FlaB1 and FlaB3, which are modified with a branched, heptameric glycan at all N-linked sequons except for the site closest to the N termini of both proteins. NMR analysis of the purified glycan determined the structure to be α-d-glycerod-manno-hep3OMe6OMe-(1–3)-[α-GalNAcA3OMe-(1–2)-]-β-Man-(1–4)-[β-GalA3OMe4OAc6CMe-(1–4)-α-GalA-(1–2)-]-α-GalAN-(1–3)-β-GalNAc-Asn. A detailed investigation by hydrophilic interaction liquid ion chromatography–MS discovered the presence of several, less abundant glycan variants, related to but distinct from the main heptameric glycan. In addition, we confirmed that the S-layer protein is modified with the same heptameric glycan, suggesting a common N-glycosylation pathway. The *M. thermolithotrophicus* archaellin N-linked glycan is larger and more complex than those previously identified on the archaellins of related mesophilic methanogens, *Methanococcus voltae* and *Methanococcus maripaludis*. This could indicate that the nature of the glycan modification may have a role in maintaining stability at elevated temperatures.

Archaea are motile via a domain-specific motility apparatus termed the archaellum (1), which has a number of similarities to type IV pili (2). The major structural subunits of the archaellum, *i.e.* the archaellins, are made as preproteins with type IV pilin-like signal peptides (3, 4) that are removed by a prepilin peptidase-like enzyme, (termed either FlaK or PilB, depending on the species) (5–7). In addition, the archaellins are typically modified with N-linked glycans (8). In contrast to Bacteria, the process of protein N-glycosylation appears to be almost universally found in Archaea (8, 9). The attachment of glycan has been shown to be necessary for proper archaellum assembly and/or motility in several archaea, including *Methanococcus voltae*, *Methanococcus maripaludis*, *Halofexa volcanii*, *Halobacterium salinarum*, and *Sulfobatus acidocaldarius* (10–14). Because archaella are easy to purify, they have often been used to study the unusual, and sometimes unique, N-glycans common in Archaea (15).

The structures of relatively few archaenal N-linked glycans have been determined. Nevertheless, they have exhibited a wide variation in terms of the nature of the monosaccharide linking the glycan to the protein, the degree of glycan branching, the sugar monomers present (including unique sugars), and the modifications of sugar components with amino acids, sulfate, or methyl groups (8). We have previously reported the structures of archaellin N-linked glycans in the mesophilic methanogens, *M. voltae* and *M. maripaludis* (16, 17). The archaellin N-linked glycan in *M. voltae* is a linear trisaccharide (16) (or tetrasaccharide (18)) with GlcNAc as the linking sugar, whereas the equivalent glycan in *M. maripaludis* is a related tetrasaccharide linked via a GalNAc to the protein (17). In addition, the *M. maripaludis* glycan contains the first evidence of a novel monosaccharide: a diglycoside of an aldulose (17).

N-Glycosylation contributes to S-layer formation and stability. Indeed, it has been shown for *S. acidocaldarius* that N-glycosylation of the S-layer is essential for survival (19). It is notable also that *H. volcanii* responds to changes in medium salinity by modulating the attached N-glycans, as well as the sites of glycan attachment to the protein (20). It is presently unclear to what extent the various features of N-glycosylation (*i.e.* the nature and composition of the glycan, the specific sites of attachment, and the extent of N-glycosylation) contribute to the stability of S-layers and archaella under different growth conditions.

The order *Methanococcales* contains archaeallated members that are mesophilic (including *M. voltae*, *M. maripaludis*, and *Methanococcus vannielii*), thermophilic (*Methanothermococcus thermolithotrophicus*), and hyperthermophilic (including...
Methanocaldococcus jannaschii, Methanocaldococcus villosus, and Methanocaldococcus infernus). This affords the opportunity to compare the structure of archaellin N-linked glycans expressed by related organisms growing at very different temperature optima that could contribute to archaellum thermostability.

Here we report the structure of the N-linked glycans attached to archaellins of the thermophilic methanogen, M. thermolithotrophicus. The purified archaellin glycan is more complex than those reported previously for mesophilic Methanococcus species. It is a branched oligosaccharide of seven residues, containing several unusual sugars, and is N-linked to asparagine residues via GalNAc. We were also able to determine that the S-layer protein is modified with the same heptameric glycan.

Results

M. thermolithotrophicus fla operon

The sequence of the fla operon of M. thermolithotrophicus has been previously reported (21, 22), and it is similar to other members of the order Methanococcales (23). In M. thermolithotrophicus, there are four archaellin genes (flaB1, flaB2, flaB3, and flaB4) followed by the fla accessory genes flaC–flaJ. Examination of the predicted amino acid sequence of the four archaellins of M. thermolithotrophicus revealed that they all contain multiple copies of the N-glycosylation sequon Asn-Xaa-(Ser/Thr) (where Xaa is any amino acid except proline), indicating the likelihood of the archaellins being glycoproteins. FlaB1 and FlaB2 have 9 and 16 potential N-glycosylation sites, respectively, whereas FlaB3 possesses five potential sites, and FlaB4 has just three.

Isolation and SDS-PAGE analysis of M. thermolithotrophicus archaella

We first attempted to isolate “intact” archaella (i.e. with attached knob-like anchoring structures at one end) using the detergent OP-10 as reported previously for the archaella of M. voltae and M. maripaludis (31, 33). However, this proved difficult because the procedure led to a massive gray precipitate that hindered further archaella purification. After several unsuccessful attempts to bypass the precipitate problem, the archaella, present as a reported tuft or clump (24, 25), were isolated from M. thermolithotrophicus after brief shearing of the cells instead. Electron microscopic examination of the recovered archaella by negative staining (Fig. 1A) revealed the presence of short filaments, ~12 nm in diameter, as previously reported for many archaella (22). The shortness of the archaella was likely a result of the shearing procedure. Analysis of the sheared, purified material by SDS-PAGE (Fig. 1B) revealed the presence of two or three major protein bands, depending on the length of time the samples were boiled prior to loading on the gel. When the samples were boiled for 1 min prior to loading, three prominent bands with apparent molecular masses of ~44 kDa (band 1), 37 kDa (band 2), and 26 kDa (band 3) were observed, but after a 5-min boiling time, only the 44- and 26-kDa bands were observed. All of the bands stained with a glycoprotein stain (Fig. 1C). The specificity of the staining was shown using whole-cell lysates of M. marisnigri in which only the S-layer glycoprotein was stained. Finally, all three bands reacted with antisera raised against the FlaB2 archaellin of M. voltae (Fig. 1D). It was previously shown that antibodies raised to M. voltae FlaB2 cross-react with archaellins from several different methanogens, presumably because the N termini of archaellins from these methanogens are highly conserved (26, 27).

MS analysis of archaellin tryptic digests

Mascot database searching of the peptide nano-LC–MS/MS spectra identified the upper band as FlaB1 and the lower band as FlaB3. The middle band observed after 1 min of boiling was also identified as FlaB1, suggesting that the shorter boiling time does not fully denature this protein and that archaellin structure significantly influences migration. It should be noted that many of the FlaB1 peptides identified are common to the FlaB2 protein, so the presence of FlaB2 cannot be completely ruled out. However, no peptides unique to FlaB2 were identified in any of the gel band digests. Similarly, no FlaB4 peptides were detected. Peptide coverage of the FlaB1 and FlaB3 amino acid sequences is provided in Fig. S1.
A manual investigation of the unidentified LC–MS/MS spectra confirmed the presence of glycopeptides in the tryptic digests of both gel bands. An example of the MS/MS spectrum obtained for the FlaB1 tryptic glycopeptide, 112SVVLNYSGK120, is shown in Figure 2. The upper half of the spectrum is dominated by the sequential loss of glycan residues from the peptide, whereas the lower m/z region contains glycan oxonium ions and peptide fragment ions confirming the identity of the peptide. The identity (determined by NMR) and mass of each monosaccharide are indicated in the inset. Monosaccharides are symbolized using the Symbol Nomenclature for Glycans format. Glycan A is GalA3O4Me6CMe. Glycan B is GalNAcA3Ome.

FlaB1 and FlaB3 possess nine and five N-linked sequons, respectively, and some of the sequon sites were determined to be glycosylated by manual searching of the nano-LC–MS/MS spectra. A more thorough investigation using HILIC-MS was undertaken to investigate the glycosylation at all sequon sites (see below). We determined that the sequon site closest to the N terminus, which is identical in both archaellins, does not appear to be glycosylated. Unmodified tryptic peptides containing this sequon site were identified by automated Mascot searching (Table S1), and no corresponding glycopeptides were ever identified. In contrast, all of the remaining N-glycosylation sites on both FlaB1 and FlaB3 appear to be heavily occupied. The only FlaB1 or FlaB3 N-sequon–containing peptide that was observed unmodified was a low-abundant FlaB3 peptide containing deamidated Asn172 (Table S1).

Although MS alone could not identify the actual monosaccharides that make up the glycan, some of the residue masses indicated the presence of recognizable sugar classes. For example, the linking monosaccharide had a mass of 203 Da and was likely an N-acetylhexosamine (HexNAc). In fact, most of the archaellin N-glycans identified to date are linked through GlcNAc or GalNAc (8). Similarly, the 162- and 176-Da residues were thought to be a hexose and a hexuronic acid, respectively, whereas the 175-Da sugar was speculated to be a hexuronamide. However, the identity of the three remaining sugars could not be inferred from their masses alone. NMR was required to assign the precise structural configuration of each sugar, as well as the linkage between the sugars.

**NMR structural analysis of the archaellin glycan**

Archaellin protein (8 mg) was extensively digested with proteinase K to enable purification of sufficient quantity of glycan material with minimal peptide backbone for structural analysis. The main component was isolated by anion-exchange chromatography and analyzed by 2D NMR (gCOSY, TOCSY, NOESY, 1H-13C HSQC, 1H-13C HMBC) (Figure 3). Spin systems...
of six monosaccharides were identified in $^1$H NMR spectra. Among them, three unusual monosaccharides were present. Additionally $^{13}$C signals were found for a residue of $\beta$-GalNAc, which had no visible H-1 signal in $^1$H spectrum because of variation of the peptide components to which it was linked.

Monosaccharide A had the COSY-TOCSY-NOESY H-1–H-4 signal pattern of $\beta$-galactopyranose. There was no H-5 and H-6 signals. H-4 was shifted to low field because of O-acetylation. HMBC OMe-H to C-3 correlation showed the presence of OMe at O-3, which also agreed with low field shift of C-3 (Table 1). A singlet methyl signal at 1.57 ppm showed HMBC correlations to C-4 and to two other quaternary carbon signals at 81.8 and 176.9 ppm. This indicated that the C-methyl group is attached to C-5 (signal at 81.8 ppm) and that the 176.9 ppm signal is from C-6 of the monosaccharide A; thus A is uronic acid. Methyl group at C-5 gave NOE correlations to protons A1, A3, and A4, which is expected for $\beta$-galacto-configuration. Thus, it was concluded that A is 3-O-methyl-5-C-methyl-4-O-acetyl-$\beta$-galactopyranosyluronic acid (structure is presented as an inset in Fig. 4).

Monosaccharide B was a heptopyranose with $\alpha$-manno-configuration of the pyranose ring. It had two OMe groups at O-3 and O-6, identified by NOE and HMBC correlations to respective atoms. The configuration of C-6 is proposed to be D because the genes for biosynthesis of d-glycero-d-manno-heptose are present in the sequenced genome (21). Monosaccharide C was identified as $\alpha$-galactosaminuronic acid with OMe at O-3, identified by NOE and HMBC correlations. Monosaccharide E was identified by NMR as $\alpha$-galactopyranosyluronic acid. MS data indicate that it had a residue mass of 175 Da, 1 Da less than expected for an uronic acid. This indicates amidation of the carboxyl group (indicated as N in $\alpha$-GalAN). Monosaccharides D and F were $\alpha$-GalA and $\beta$-Man, respectively, identified using vicinal proton coupling constants and $^{13}$C shifts. Monosaccharide G was identified by MS as a HexNAc. In NMR spectra its signals were fuzzy and weak because of the presence of various peptides at the reducing end. Nevertheless, it was possible to identify this residue as $\beta$-GalNAc.

The sequence of monosaccharides was deduced from NOE correlations A1:D4; B1:F3; C1:F2; D1:E1,2,3; and E1: G3,4 and HMBC correlations (not all transglycosidic HMBC signals were obtained) A1:D4; D1:E2; and F1:E4. Residue G was substituted to O-3 as confirmed by $^{13}$C NMR shifts. All
linkages agreed with $^{13}$C signals positions. A schematic of the archaellin glycan is presented as an inset in Fig. 3, and the structure of the glycan in chair projections is shown in Fig. 4.

**The archaellins of *M. thermolithotrophicus* are modified with a group of related N-glycans**

Although the heptasaccharide identified by NMR is the predominant N-glycan on these archaellins, reverse-phase LC–MS/MS analysis detected the presence of less abundant but related glycans. To investigate this observation more thoroughly, the tryptic digests of the FlaB1 and FlaB3 gel bands were analyzed by HILIC–MS/MS. HILIC resolves molecules based on their hydrophilicity, allowing for the separation of glycosylated and nonglycosylated peptides. The HILIC–MS/MS total ion chromatograms for the FlaB1 and FlaB3 tryptic digests are presented in Fig. 5. In both cases, a relatively complex set of glycopeptide peaks can be observed eluting later in the chromatograms, well-separated from the nonglycosylated peptides. Table 2 lists the identity of the more abundant glycopeptides identified in each analysis, and their MS/MS spectra are presented in Fig. S2. All N-sequons in FlaB1 and FlaB3 were shown to be glycosylated except for the sequon closest to the N terminus.

To date, we have identified seven unique but related N-glycans on the archaellins of *M. thermolithotrophicus*. The extracted ion chromatograms for the glycopeptides (MH$_2^{2+}$ ion) derived from $^{112}$SVVLNYSGK$^{120}$ (FlaB1) and $^{156}$NTTPVINK$^{163}$ (FlaB3) are presented in Fig. 6. The corresponding MS/MS spectra for the $^{112}$SVVLNYSGK$^{120}$ glycopeptides together with the proposed glycan compositions are presented in Fig. S3. The heptasaccharide identified by NMR is clearly the most abundant glycan present, but when taken together, the other glycans make up ~30% of the total glycan modifying these archaellins. Some of the minor

---

**Table 1**

NMR chemical shifts of the heptameric archaellin glycan

<table>
<thead>
<tr>
<th>Monosaccharides</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$-GalA3OMe5Me4OAc</td>
<td>4.79</td>
<td>3.61</td>
<td>3.79</td>
<td>5.67</td>
<td>1.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>102.2</td>
<td>71.9</td>
<td>79.1</td>
<td>72.4</td>
<td>81.8</td>
<td>176.9</td>
<td>19.1</td>
</tr>
<tr>
<td>$\alpha$-DDHep3,6OMe$_2$</td>
<td>5.56</td>
<td>4.16</td>
<td>3.80</td>
<td>3.53</td>
<td>3.94</td>
<td>3.79</td>
<td>3.71; 3.85</td>
</tr>
<tr>
<td>H</td>
<td>99.7</td>
<td>66.8</td>
<td>81.6</td>
<td>67.4</td>
<td>70.6</td>
<td>81.5</td>
<td>60.9</td>
</tr>
<tr>
<td>C $\alpha$-GalNAcA3OMe</td>
<td>5.49</td>
<td>4.21</td>
<td>4.02</td>
<td>4.63</td>
<td>4.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>99.1</td>
<td>49.8</td>
<td>76.0</td>
<td>66.4</td>
<td>74.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>5.46</td>
<td>4.12</td>
<td>3.96</td>
<td>4.57</td>
<td>4.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D $\alpha$-GalA</td>
<td>96.8</td>
<td>58.7</td>
<td>71.4</td>
<td>82.1</td>
<td>73.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>5.33</td>
<td>3.99</td>
<td>4.03</td>
<td>4.47</td>
<td>4.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>95.1</td>
<td>72.1</td>
<td>68.2</td>
<td>79.9</td>
<td>72.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E $\alpha$-GalAN</td>
<td>4.99</td>
<td>4.31</td>
<td>3.73</td>
<td>3.91</td>
<td>3.34</td>
<td>3.77; 3.92</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>102.6</td>
<td>77.0</td>
<td>74.2</td>
<td>70.1</td>
<td>78.3</td>
<td>61.8</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>5.12</td>
<td>4.08</td>
<td>3.93</td>
<td>4.30</td>
<td>3.78</td>
<td>3.78; 3.78</td>
<td></td>
</tr>
<tr>
<td>G $\beta$-GalNAc</td>
<td>79.4</td>
<td>50.7</td>
<td>78.9</td>
<td>65.5</td>
<td>77.6</td>
<td>62.3</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.** Structure of the archaellin N-linked glycan of *M. thermolithotrophicus* as determined by NMR and MS. Each residue has been assigned a letter to aid in the description of the NMR spectra (please refer to the text).
Structures of N-linked glycans in \textit{M. thermolithotrophicus}

![Figure 5. HILIC-MS total ion chromatograms obtained for the in-gel tryptic digests of FlaB1 (a) and FlaB3 (b), respectively. The principal glycopeptides in the labeled peaks are listed in Table 2, and their corresponding MS/MS spectra are presented in Fig. S2. Rel. Int., relative intensity.](image)

Table 2

<table>
<thead>
<tr>
<th>Peak no.</th>
<th>Mass (Da)</th>
<th>Amino acid sequence</th>
<th>Glycan species</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3991.8</td>
<td>241NGTTVWGADDSSLVVALQDADK265</td>
<td>1413.4</td>
</tr>
<tr>
<td>2</td>
<td>2379.0</td>
<td>1131SVLVNYSKG120</td>
<td>1413.4</td>
</tr>
<tr>
<td>3</td>
<td>3954.6</td>
<td>209SYDGTTFTIVMNTSTNIVTADK231</td>
<td>1413.4</td>
</tr>
<tr>
<td>4</td>
<td>1956.8</td>
<td>189VTVLSNNTITSENITTVYSR208</td>
<td>175155</td>
</tr>
<tr>
<td>5</td>
<td>6717.2</td>
<td>133LTNAFAINYSNVSHPVTTTVIK154</td>
<td>1413.4</td>
</tr>
<tr>
<td>6</td>
<td>5216.1</td>
<td>132LTNAFAINYSVSHPVTTTVIK154</td>
<td>174</td>
</tr>
<tr>
<td>7</td>
<td>4717.2</td>
<td>185VTSLSNHTSEETTVYSR208</td>
<td>1413.4</td>
</tr>
<tr>
<td>8</td>
<td>1623.3</td>
<td>153DNTTNGSTD5TSTNIVTADK175</td>
<td>1413.4</td>
</tr>
<tr>
<td>9</td>
<td>3623.4</td>
<td>163GDIVAIALNTGTSMDPR181</td>
<td>1413.4</td>
</tr>
<tr>
<td>10</td>
<td>4397.7</td>
<td>1131VTDNTSDYDSLTSGGSFVESDVWNSSVK138</td>
<td>1413.4</td>
</tr>
</tbody>
</table>

The S-layer protein of \textit{M. thermolithotrophicus} is modified with the same N-linked heptameric glycan

Akca et al. (28) purified the S-layer protein from \textit{M. thermolithotrophicus} and, using N-terminal amino acid sequence information, were able to clone the S-layer gene. The identified \textit{S}-layer protein of \textit{Mtc thermolithotrophicus} (Swiss-Prot accession no. Q8X235.1) is a 559-amino acid protein with a 28-amino acid signal peptide and is predicted to have a mass of 59.1 kDa (56.54 kDa after signal peptide removal). On SDS-PAGE, however, the protein migrated with an apparent molecular mass of 82 kDa, which may be due to glycosylation because the protein contains six potential sites for N-glycosylation. Similarly, SDS-PAGE examination of our membrane preparations of \textit{M. thermolithotrophicus} also revealed that the major band in the sample migrated at ~82 kDa, and LC–MS/MS analysis of the excised band identified it as the same S-layer protein identified by Akca et al. (28). Manual analysis confirmed that four linkage sites were occupied with the same heptameric glycan as observed on the archaellins (Fig. 7). Two sites, Asn155 and Asn515, appear not to be glycosylated. A detailed analysis of glycan heterogeneity at each site was not performed, although glycans lacking an acetyl or methyl group were detected.

Discussion

We isolated archaella from the thermophilic methanogen \textit{M. thermolithotrophicus} by shearing archaellum from the cells. Purified archaella sample, when boiled in Laemmli buffer for 1 min, yielded three major bands by SDS-PAGE with apparent molecular masses of ~44, 35, and 26 kDa, respectively. However, the 37-kDa band disappeared when the archaellins were boiled for 5 min prior to loading on the gel. The bands stained with glycoprotein stain and cross-reacted in Western blots with antibodies raised against FlaB2 of the mesophilic relative, \textit{M. voltae}. As was mentioned earlier, antibodies raised to \textit{M. voltae} FlaB2 were shown to cross-react with archaellins from several different methanogens (26, 27).

\textit{M. thermolithotrophicus} possesses four archaellin genes encoding FlaB1, FlaB2, FlaB3, and FlaB4. The first two archaellin genes encode proteins of 331 and 435 amino acids with predicted molecular masses of 34.93 and 42.26 kDa, respectively. These are much larger than the typical size of archaellins found in other members of the \textit{Methanococcales} (i.e. \textit{M. voltae}, \textit{M. maripaludis}, \textit{M. vannielii}, and \textit{M. janaschii}), which are usually ~200–220 amino acids in length (4, 22, 29). The remaining two archaellin genes of \textit{M. thermolithotrophicus} encode proteins of smaller length: FlaB3 is 216 amino acids long (predicted molecular mass, 22.77 kDa), whereas FlaB4 is 217 amino acids long (predicted molecular mass, 22.89 kDa). The 44- and 37-kDa protein bands were identified by MS as FlaB1, and the 26-kDa protein was identified as FlaB3. A previous study (30) of the archaella of \textit{M. thermolithotrophicus} also observed three major protein components upon SDS-PAGE analysis with apparent masses of 62, 44, and 26 kDa. In this early study, the bands shifted in mass depending on the length of time the samples were boiled prior to gel loading, although none were assigned to a gene product. We hypothesize that this same phenomenon is occurring here and is related to incomplete denaturation of the archaellin.

Although we only identified two archaellin proteins in this analysis, all four archaellins may be present in the assembled structure. In other archaea with multiple archaellin genes, all archaellins were detected in purified archaella samples, although often at varying abundances (16, 17, 31, 32). One of the unidentified archaellins, FlaB2, has 16 potential sites for N-glycosylation, and if the majority of these are occupied with
glycan, this protein could conceivably run at a significantly higher apparent molecular mass near 60 kDa, and could be the 62-kDa band previously observed by Kostyukova et al. (30).

The fact that we did not detect FlaB2 or FlaB4 may be either due to their low abundance in archaella or because they occupy a cell-proximal location and remain attached to the cells following shearing.

The archaella of some archaea, including M. voltae (31), M. maripaludis (33), and H. salinarum (34), have a curved hook-like region of varying length at the cell-proximal end. There is evidence that this curved region is composed primarily of one type of archaellin. In the case of M. voltae, archaella were first removed by a brief shearing of the cells. The sheared cells were then detergent-treated to extract the cell-proximal archaella stubs (31). These archaella pieces were much shorter, and many had curved hook-like ends when examined by EM. Comparison of the SDS-PAGE protein profiles of the sheared preparation with the archaella stubs revealed the presence of FlaB3 only in the stub fraction, suggesting that it likely formed the hook region. Because the archaella of M. thermolithotrophicus also have a curved cell-proximal region (24), it seems likely that this portion of the archaellum may also be composed of one of the undetected archaellins. M. voltae also has a fourth archaellin, FlaA, which is not observed in Coomassie-stained gels after SDS-PAGE of sheared samples or archaella stubs, but its presence can be shown in both samples immunologically (31). This suggests FlaA is not observed in Coomassie-stained gels because it is present in only small amounts, and this could be true of either FlaB2 or FlaB4 in M. thermolithotrophicus archaella. A previous study reported on the transcription of the fla operon of M. thermolithotrophicus (22). The longest transcript detected was ~3 kb extending from flaB1 through flaB2 and flaB3. Attempts to detect transcripts for flaB4 or downstream fla accessory genes (flaC or flaH) were unsuccessful. However, because the fla accessory genes are required for archaellation, they must be transcribed in archaellated cells but likely at low levels. Because flaB4 is located between flaB1–flaB3 (for which a transcript was detected) and flaC–flaH (which must be transcribed for archaellation to occur), we believe it is most likely that flaB4 is also transcribed. Specific antibodies, which could allow for immunological detection of minor amounts of FlaB2 and FlaB4, are not available.

Using MS, we determined that FlaB1 and FlaB3 are glycosylated at multiple sites with a branched heptameric glycan with a mass of 1413.4 Da (Fig. 2). The full structure of the glycan was determined by NMR (Fig. 4). This glycan is more complex than...
the linear tri- and tetrasaccharide glycans found on the archaellins of M. voltae (16) and M. maripaludis (17). One of the sugars, 3-O-methyl-5-C-methyl-4-O-acetyl-β-D-galactopyranosyluronic acid, has never been reported previously. The sugar linking the glycan to the asparagine is GalNAc as it is for M. maripaludis, whereas this position is occupied by GlcNAc in M. voltae.

The observed mass of FlaB1 by SDS-PAGE (44 kDa) was approximately 9 kDa greater than the predicted mass from amino acid sequence only. This suggests that six or seven of the nine possible FlaB1 N-linked glycosylation sites are likely modified with glycan. In fact, using reverse-phase LC-MS and HILIC-MS, we determined that eight of the nine sequon sites are glycosylated and that they appear to be fully occupied. Similarly, the FlaB3 protein migrated at a mass ~3 kDa greater than the predicted mass, suggesting that two of the five potential sites are likely glycosylated. Here again, HILIC-MS analysis confirmed glycosylation at four of the five sequon sites. These results suggest that the mass shifts observed by SDS-PAGE are underestimating the degree of glycosylation on both proteins.

Analysis of the annotated genome sequence of M. thermolithotrophicus revealed one large region encompassing WP_018154793.1 to WP_018154819.1 plus WP_083876343.1 to WP_083876346.1 that contains genes annotated as potentially involved in glycan synthesis and assembly. Some of these genes are annotated as being involved in the synthesis and transfer of heptose, which is consistent with the identification of the dimethyl D,D-heptose residue in the glycan (e.g. WP_018154816.1, D-sedoheptulose 7-phosphate isomerase; WP_018154815.1, D-glycero-β-D-manno-heptose 1-phosphate adenylyltransferase; WP_018154814.1, bifunctional heptose 7-phosphate kinase/heptose 1-phosphate adenylyltransferase; WP_083876343.1, lipopolysaccharide heptosyltransferase family protein). Interestingly, homologues to both WP_018154816.1 and WP_018154815.1 are found in M. voltae and M. maripaludis, whereas M. maripaludis but not M. voltae has a homologue to WP_018154814.1. Neither M. maripaludis nor M. voltae has a homologue of WP_083876343.1, the glycosyltransferase family 9 protein. Neither M. maripaludis nor M. voltae have a heptose derivative in their N-linked glycans. All four of the M. thermolithotrophicus proteins have homologues in Methanococcus aeolicus, but glycan structures have not been reported for this organism. It is currently not possible to test the role of these potential agl genes in M. thermolithotrophicus N-glycan formation because there are no genetic tools available for this thermophile.

We have also demonstrated that the M. thermolithotrophicus S-layer protein is modified at multiple sites with the same N-linked glycan as found on the archaellins. It is commonly found in archaea that S-layer proteins, archaellins, and type IV pilins are all glycoproteins. Few studies have examined the nature of the N-glycans attached to the different proteins in the same cell, but when it has been reported, the same or very similar N-glycans have been found. An identical trisaccharide was found attached to both archaellins and the S-layer protein of M. voltae (16). In H. volcanii, the pentasaccharide initially reported attached to the S-layer (35, 36) was later found on the archaellins (12) and the type IV pilins (37). In S. acidocaldarius, a complex tribranched hexasaccharide first reported linked to cytochrome b_{558/566} was later shown to be attached to numerous surface proteins including the S-layer protein and archaellin (11, 38). The situation is a little different for M. maripaludis where a tetrasaccharide was found attached to the archaellins, whereas the major type IV pilin had the same tetrasaccharide but modified with the addition of a hexose extending as a branch from the linking galactosamine residue (39). The most complicated situation is present in H. salinarum, in which the S-layer protein is modified with two different N-linked glycans: a repeating unit glycan found at a single site and a sulfated linear oligosaccharide found at 10 sites. The latter, but not the repeating unit glycan, is also found on the archaellins (40).

We observed a significant amount of heterogeneity in the N-linked glycans. Approximately 70% of the glycan signal was from the heptasaccharide identified by NMR with the remaining 30% comprised of related glycans lacking monosaccharides or functional groups or having alternative monosaccharides.
Heterogeneity of archaeal N-glycans has been previously reported, especially in extreme halophiles. In *H. volcanii*, analysis of N-linked glycans attached to the S-layer protein and to type IV pilins revealed the presence of glycopeptides modified with mono-, di-, tri-, tetra-, and pentasaccharides (36). Interestingly, the third sugar was different in the trisaccharide compared with the tetrasaccharide and pentasaccharide. In an analysis of type IV pilins in *H. volcanii*, the proteins were also modified by the same pentasaccharide as the S-layer as well as shortened glycans corresponding to 1–4 sugars; the tetrasaccharide and pentasaccharide were more abundant than the shorter glycans (37).

The *M. thermolithotrophicus* archaellin glycan heterogeneity is enabled by the relaxed specificity of the oligosaccharyltransferase, AglB. AglBs are known to have the ability to transfer a variety of related and even unrelated glycans to their target proteins. AglBs from a variety of related methanogens, including from *M. thermolithotrophicus*, can complement an aglB deletion in *M. maripaludis* (41). This means that *M. thermolithotrophicus* AglB, which normally transfers a multibranched heptasaccharide, can also efficiently transfer the linear tetrasaccharide of *M. maripaludis* to its appropriate target protein. Similarly, relaxed substrate specificity was shown for the AglBs of extreme halophiles in which the focus was on N-glycosylation of S-layer proteins. An aglB mutant of *H. volcanii* could be complemented by AglBs from several different extreme halophiles, which typically transfer very different glycans (42). Furthermore, in a study of various glycosyltransferase gene mutants in *M. maripaludis*, AglB was shown to efficiently transfer mutant shortened forms of the glycan to the target proteins (13).

Thus, it seems that at least some of the heterogeneity of archaellin N-linked glycans observed in *M. thermolithotrophicus* may be attributed to AglB transferring truncated, incomplete glycans from the lipid carrier to the archaellins.

The increased complexity of the glycan in this thermophilic member of Methanococcales compared with the simpler structures synthesized by the mesophilic members of the same order may indicate that more complex glycans play a role in archaellin stability at elevated temperatures. This hypothesis is difficult to test because of the lack of a manipulatable genetic system for *M. thermolithotrophicus*, making it impossible to generate truncated glycans in vivo. However, there are well-documented examples where glycosylation of proteins results in increased thermostability (43). In archaea, it was recently shown that the type IV pili of *Sulfolobus islandicus* can survive extreme conditions including boiling in 1% SDS (44). This remarkable stability has been attributed to extensive O-glycosylation of the component pilins. Mutant studies in various archaea have revealed that interference with the N-glycosylation pathway can negatively impact archaellin assembly and function, pilus clumping, S-layer stability, and the subsequent ability of cells to adapt to medium or low osmolarity (8, 37). Recently, it was shown that the psychrophilic methanogen *Methanolobus psychrophilus* increases the extent of N-glycosylation on its S-layer proteins in response to growth above optimal temperatures (45).

Many archaea thrive in more extreme habitats where cell surface structures, mainly S-layers, archaella, and pili, are directly exposed to harsh environmental condition with very high temperatures and/or low pH levels. The extensive and diverse glycosylation of S-layer proteins, archaellins, and pilins across species may play an important role in the stability of these structures under adverse and diverse extreme conditions. Because the N-linked glycans of the thermophilic *M. thermolithotrophicus* are more complex than those of its mesophilic relatives, it would be of great interest to investigate the nature of N-linked glycans in archaellated hyperthermophilic members of the order Methanococcales, such as *M. jannaschii* (46) and *M. villosus* (47), that are heavily archaellated, to see whether this is a general trend or simply a unique feature of *M. thermolithotrophicus*. Increased complexity in the glycans in methanogens growing at >80°C would support a role for the nature of the glycan in archaellin thermostability.

**Experimental procedures**

**Strains and growth conditions**

*M. thermolithotrophicus* DSM 2095 (type strain) (25) was obtained from Richard Sparling (University of Manitoba, Winnipeg, Manitoba, Canada). It was grown in Balch medium III (48) in 1-liter stoppered bottles containing 120 ml of medium under a headspace of CO₂/H₂ (20:80) at 65°C with shaking. *Methanoculleus marisnigri* JR1 (type strain) (49) was grown in Balch medium III under a headspace of CO₂/H₂ (20:80) at 30°C with shaking.

**Isolation of archaella**

The cells (6 liters) were harvested by centrifugation (8,000 × g for 15 min), resuspended in Balch medium III, and sheared for 40 s in a Waring blender (50). Intact cells were removed by centrifugation (8,000 × g for 15 min), and the supernatant was centrifuged for 45 min at 20,000 × g to remove membrane fragments. A crude archaella preparation was obtained from that supernatant by a further centrifugation (80,000 × g/90 min). That pellet was resuspended in a minimum volume and loaded onto a KBr gradient (0.5 g KBr/ml 25 mM HEPES buffer, pH 7.5) and centrifuged for 18 h at 35,000 rpm in an SW41 rotor. The white band of archaella was removed, diluted in 2% NaCl, and centrifuged (80,000 × g/90 min) to obtain the final archaella sample free of KBr.

**SDS-PAGE, Western blotting, and staining procedures**

Archaella samples were boiled in electrophoresis sample buffer for either 1 or 5 min prior to loading onto 10% acrylamide gels. The samples were electrophoresed using the Laemmli buffer system (51). Coomassie Blue staining was performed as described (52). Glycoprotein staining was carried out using a glycoprotein detection kit (Sigma–Aldrich), which is a modification of the periodic acid–Schiff methodology. For this staining, whole cells and the purified S-layer of *M. marisnigri* were used as controls because it is known that the S-layer of this organism is a glycoprotein (molecular weight, 138,000) that is stained with the periodic acid–Schiff reagent (53). The S-layer of *M. marisnigri* was isolated using Triton X-100 at 80°C as described by Bayley and Koval (54). For Western
blotting, the samples were electrophoresed and transferred to Immobilon-P membrane (Millipore) using a semidy filter apparatus. The blots were probed with antibodies raised in chickens to FlaB2 of *M. voltae* (31) followed by a secondary horseradish peroxidase–conjugated rabbit anti-chicken IgY (Jackson ImmunoResearch Laboratories). Blots were developed with a chemiluminescent detection kit (Roche Molecular Biochemicals).

**EM**

Archaella isolated from *M. thermolithotrophicus* were placed on 200-mesh carbon-coated copper grids and allowed to adhere for 1 min. The grids were then washed with 2% (w/v) NaCl and stained with 2% (w/v) phosphotungstic acid, pH 7.0. The samples were imaged under standard operating conditions using an FEI Tecnai G2 F20 transmission electron microscope operating at 200 kV with a bottom-mount Gatan 4k CCD camera.

**MS analysis of archaella**

Archaella samples were prepared for MS analysis by boiling for either 1 or 5 min in Laemmli buffer, prior to separation by SDS-PAGE and staining with Coomassie Blue. The gel bands were cut out, placed in clean Eppendorf tubes, and destained by shaking in 100 mM ammonium bicarbonate, 30% acetonitrile. The gel bands were dehydrated with acetonitrile and then rehydrated in 50 mM ammonium bicarbonate containing 250 ng of sequencing grade trypsin and MS grade LysC (Promega, Madison, WI, USA) and incubated overnight in a 37 °C oven. The digest solutions were then transferred to clean Eppendorf tubes and stored at 4 °C until analyzed.

The archaella tryptic digests were analyzed by nano-LC–MS/MS by using an M-Class nano-UPLC system (Waters, Milford, MA, USA) coupled to a Synapt G2Si quadrupole TOF mass spectrometer (Waters). The peptides were injected onto an Acclaim PepMax100 C18 μ-precolumn (5 mm × 300 μm inner diameter; Dionex/Thermo Scientific, Sunnyvale, CA, USA) and resolved on a 1.7 μm BEH130 C18 column (100 × 100 μm inner diameter; Waters) using the following gradient conditions: 1 to 40% organic mobile phase (acetonitrile, 0.1% formic acid) over 14 min followed by an increase to 60% in 3 min. The flow rate was 500 nl/min. MS/MS spectra were acquired on doubly, triply, and quadruply charged ions. The nano-LC–MS/MS files were converted to Mascot Generic Files (.mgf) using autoPLGS 1.0.0 (Waters) and searched against a custom database using the Mascot search engine (Matrix Science Ltd., London, UK). The in-house database was small, containing ~100 sequences of prokaryotic proteins including the suspected archaella and S-layer protein sequences. The search conditions were as follows: 1) the protease was trypsin, and up to one missed cleavage was allowed; 2) methionine oxidation and asparagine deamidation were included as variable modifications; 3) there were no fixed modifications; 4) the mass tolerance for both the precursor and fragment ions was set at 0.2 Da; and 5) the threshold peptide score was set at 25. The unidentified MS/MS spectra were searched manually for evidence of glycopeptides.

**MS analysis of the S-layer protein**

To characterize the N-glycosylation on the S-layer protein, a *M. thermolithotrophicus* membrane protein preparation was first prepared. The cells were harvested at 13,000 × g/2 min in a bench top centrifuge. The pellet was resuspended in 1% NaCl and diluted 20-fold with ice-cold ddH₂O. Viscosity was reduced upon the addition of DNase/RNase. The membranes were obtained by centrifugation of the lysed cell suspension at 13,000 × g/10 min. The membrane pellet was resuspended in ddH₂O, and proteins were resolved by SDS-PAGE. A large protein band at ~80 kDa was excised, processed, and analyzed in the same manner as described above for the archaella bands except that the digest was analyzed on a Q-TOF2 hybrid quadrupole TOF mass spectrometer (Waters).

**HILIC–MS/MS analysis of archaella in-gel tryptic digests**

The gel band tryptic digests were analyzed using HILIC–MS/MS using a HP1260 HPLC system (Agilent) coupled to the Synapt G2Si quadrupole TOF mass spectrometer. The digests were evaporated to dryness and reconstituted in 0.1% TFA (Sigma–Aldrich) in 80% aqueous acetonitrile. Approximately, 2 μg of each digest was injected onto an ACQUITY UPLC BEH amide column (130 Å, 1.7 μm, 150 × 1.0 mm; Waters). Mobile phase A and B were 0.1% TFA in ddH₂O and 0.1% TFA in acetonitrile, respectively. The digest peptides and glycopeptides were resolved using the following gradient: 85% mobile phase B (0–2 min), decrease to 50% B (2–50 min), hold at 50% B (50–55 min), and then decrease to 20% B (55–60 min). The flow rate was 50 μl/min. MS and MS/MS conditions were as described above for the reverse-phase nano-LC–MS/MS analysis with the electrospray ionization source being adapted for capillary flow. The resulting spectra were examined manually for evidence of glycopeptides.

**Archaella glycan purification for NMR analysis**

To obtain glycan material devoid of archaella protein backbone for structural analysis, 8 mg of purified archaella was digested with proteinase K at a ratio of 1:1 (Sigma–Aldrich) in 20 mM Tris, pH 8.0, for 48 h. The proteinase K–digested material was lyophilized and resuspended in distilled H₂O. The sample was then fractionated on a Biogel P10 column in a 0.1% TFA, 80% acetonitrile gradient with UV detection at 220 nm. The fractions were collected and re-examined by 1H NMR for presence of glycan.

**NMR spectroscopy**

NMR experiments were carried out on a Bruker AVANCE III 600 MHz and 900 MHz spectrometers with 5-mm Z-gradi ent probes with acetone internal reference (2.225 ppm for 1H and 31.45 ppm for 13C) using standard pulse sequences cosygpprqf (gCOSY), mlevphpr (TOCSY; mixing time, 120 ms), roesyphpr (Rotating frame overhauser enhancement spectroscopy (ROESY); mixing time, 500 ms), hsqcedetgp (gHSQC),

**Structures of N-linked glycans in M. thermolithotrophicus**

The samples were electrophoresed and transferred to Immobilon-P membrane (Millipore) using a semidy filter apparatus. The blots were probed with antibodies raised in chickens to FlaB2 of *M. voltae* (31) followed by a secondary horseradish peroxidase–conjugated rabbit anti-chicken IgY (Jackson ImmunoResearch Laboratories). Blots were developed with a chemiluminescent detection kit (Roche Molecular Biochemicals).

A large protein band at ~80 kDa was excised, processed, and analyzed in the same manner as described above for the archaella bands except that the digest was analyzed on a Q-TOF2 hybrid quadrupole TOF mass spectrometer (Waters).

The gel band tryptic digests were analyzed using HILIC–MS/MS using a HP1260 HPLC system (Agilent) coupled to the Synapt G2Si quadrupole TOF mass spectrometer. The digests were evaporated to dryness and reconstituted in 0.1% TFA (Sigma–Aldrich) in 80% aqueous acetonitrile. Approximately, 2 μg of each digest was injected onto an ACQUITY UPLC BEH amide column (130 Å, 1.7 μm, 150 × 1.0 mm; Waters). Mobile phase A and B were 0.1% TFA in ddH₂O and 0.1% TFA in acetonitrile, respectively. The digest peptides and glycopeptides were resolved using the following gradient: 85% mobile phase B (0–2 min), decrease to 50% B (2–50 min), hold at 50% B (50–55 min), and then decrease to 20% B (55–60 min). The flow rate was 50 μl/min. MS and MS/MS conditions were as described above for the reverse-phase nano-LC–MS/MS analysis with the electrospray ionization source being adapted for capillary flow. The resulting spectra were examined manually for evidence of glycopeptides.

To obtain glycan material devoid of archaella protein backbone for structural analysis, 8 mg of purified archaella was digested with proteinase K at a ratio of 1:1 (Sigma–Aldrich) in 20 mM Tris, pH 8.0, for 48 h. The proteinase K–digested material was lyophilized and resuspended in distilled H₂O. The sample was then fractionated on a Biogel P10 column in a 0.1% TFA, 80% acetonitrile gradient with UV detection at 220 nm. The fractions were collected and re-examined by 1H NMR for presence of glycan.

NMR experiments were carried out on a Bruker AVANCE III 600 MHz and 900 MHz spectrometers with 5-mm Z-gradient probes with acetone internal reference (2.225 ppm for 1H and 31.45 ppm for 13C) using standard pulse sequences cosygpprqf (gCOSY), mlevphpr (TOCSY; mixing time, 120 ms), roesyphpr (Rotating frame overhauser enhancement spectroscopy (ROESY); mixing time, 500 ms), hsqcedetgp (gHSQC),
Structures of N-linked glycans in M. thermolithotrophicus

hsqcetgpm (gHSQC-TOCSY; 80-ms TOCSY delay), and hmbcgplpndqf (HMBC; 70- or 100-ms long-range transfer delay). Spectral widths were 10 ppm for proton and 200 ppm for carbon observations, respectively. Edited $^1$H–$^{13}$C HSQC were recorded with 1500 data points in F2 and 256 in F1 with heteronuclear $^1$J_C,H constant of 142.8 Hz (CH and CH$_3$ signals were positive, and CH$_2$ signals were negative). Resolution was kept at $<$3 Hz/pt in F2 in proton–proton correlations and 5 Hz/pt in F2 of H–C correlations. The spectra were processed and analyzed using the Bruker Topspin version 2.1 program.

Data availability

The MS data have been deposited to the ProteomeXchange Consortium via the PRIDE [1] partner repository with the data set identifier PXD019548. The NMR spectral files have been deposited in the Biological Magnetic Resonance Bank under accession number 50186.

Acknowledgments—We thank Sam Williamson for preparing the nano-LC–MS/MS platform for the experiments described here and Dr. Janelle Sauvageau for submitting the NMR spectral data to the BMR database.


Funding and additional information—This work was supported in part by Discovery Grants from the Natural Sciences and Engineering Research Council of Canada (to K. F. J. and C. M. K.).

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

Abbreviations—The abbreviations used are: HILIC, hydrophilic interaction liquid ion chromatography; HSQC, heteronuclear single quantum coherence spectroscopy; gCOSY, gradient correlated spectroscopy; TOCSY, total correlation spectroscopy; HMBC, heteronuclear multiple bond correlation spectroscopy; ROESY, rotating frame overhauser enhancement spectroscopy; ddH$_2$O, double-distilled H$_2$O.

References


