Protein glycosylation in Archaea: Sweet and extreme

Doron Calo², Lina Kaminski², and Jerry Eichler¹,²
¹Department of Life Sciences, Ben Gurion University of the Negev, Beersheva 84105, Israel
²Received on March 7, 2010; revised on March 31, 2010; accepted on March 31, 2010

While each of the three domains of life on Earth possesses unique traits and relies on characteristic biological strategies, some processes are common to Eukarya, Bacteria and Archaea. Once believed to be restricted to Eukarya, it is now clear that Bacteria and Archaea are also capable of performing N-glycosylation. However, in contrast to Bacteria, where this posttranslational modification is still considered a rare event, numerous species of Archaea, isolated from a wide range of environments, have been reported to contain proteins bearing Asn-linked glycans moieties. Analysis of the chemical composition of the Asn-linked polysaccharides decorating archaeal proteins has, moreover, revealed the use of a wider variety of sugar subunits than seen in either eukaryal or bacterial glycoproteins. Still, although first reported some 30 years ago, little had been known of the steps or components involved in the archaeal version of this universal posttranslational modification.

Now, with the availability of sufficient numbers of genome sequences and the development of appropriate experimental tools, molecular analysis of archaeal N-glycosylation pathways has become possible. Accordingly using halophilic, methanogenic and thermophilic model species, insight into the biosynthesis and attachment of N-linked glycans decorating archaeal glycoproteins is starting to amass. In this review, current understanding of N-glycosylation in Archaea is described.

Keywords: Archaea/extremophiles/N-glycosylation/posttranslational modification

Archaea: the third domain of life

The Archaea were first recognized as a distinct domain of life, unrelated to either Bacteria or Eukarya, in 1977, as a result of Carl Woese’s pioneering use of 16S ribosomal (r)RNA analysis (Fox et al. 1977; Woese and Fox 1977). At the time, those microorganisms assigned to the archaeal domain were all extremophiles, i.e. species able to thrive in the face of some of the most physically challenging conditions on the planet. Archaea were shown to reside in seemingly ‘harsh’ surroundings, such as those characterized by extremes in salinity, pH or temperature, to live in sulfur-based environments or to produce methane as a by-product of their anaerobic respiration (cf. Rothschild and Mancinelli 2001). However, with the subsequent widespread application of 16S rRNA analysis, it became clear that Archaea are major denizens of ‘normal’ biological niches, including seawater, soil and even our own intestinal flora (DeLong 1998; Chaban, Ng, et al. 2006). Indeed, it is now recognized that Archaea are major players in the ecosystem, important for processes as diverse as the Earth’s nitrogen cycle and global warming (Francis et al. 2007; Galperin 2007).

As befitting a distinct form of life, Archaea possess traits that distinguish them from either Bacteria or Eukarya, in addition to their characteristic rRNA. For example, archaeal membranes are composed of phospholipids of strikingly different composition than those used elsewhere. Whereas bacterial and eukaryal phospholipids comprise fatty acyl groups ester-linked to the sn-1,2 positions of glycerol, in Archaea, membrane lipids comprise polyisoprenyl groups ether-linked to the sn-2,3 positions of a glycerol backbone (Sprott 1992). On the other hand, Archaea share much in common with both Bacteria and Eukarya. Like Bacteria, Archaea are single-celled organisms surrounded by a single membrane and lack internal organelles. Moreover, in both cases, the genome is organized as a single, circular chromosome, often divided into operons (Koonin and Wolf 2008). When, however, replication, DNA packaging, transcription and other aspects of information processing are considered, Archaea are more reminiscent of Eukarya (Sandman and Reeve 2000; Bell and Jackson 2001). Thus, archaeal biology can be considered a mosaic of archaeal-specific, bacterial-like and eukaryal-like traits. As such, the study of Archaea has not only revealed biological strategies not seen elsewhere but has also served to reveal links between bacterial and eukaryal processes previously thought to be unrelated, as well as providing first examples of universal biological phenomena. With respect to protein glycosylation, Archaea have provided examples of each of these three scenarios.

Archaeal N-linked glycoproteins: here, there and everywhere

Since Neuberger reported that a carbohydrate group was an integral part of ovalbumin some 75 years ago (Neuberger 1938), it was believed that N-glycosylation was a trait restricted to Eukarya. This belief was, however, challenged in 1976 with the demonstration that the surface (S)-layer glycoprotein of the extreme halophile, Halobacterium salinarum, experiences this same processing event (Mescher and Strominger 1976a). De-
tailed analysis showed that the S-layer glycoprotein is modi-
ified by two different Asn-linked oligosaccharides, with Asn-2 being
decorated with a repeating sulfated pentasaccharide, bound via
an N-glycosylamine bond, and 10 other Asn residues being
modified with a sulfated glycan, linked through a glucose res-
idue (Wieland et al. 1980; Wieland et al. 1983; Lechner et al.
1985a; Lechner and Wieland 1989) (Figure 1A). At the time of
first being shown to contain N-linked glycans,
*Hbt. salinarum* was de-

**Fig. 1.** The composition of representative N-linked glycans decorating archaeal glycoproteins. N-Linked glycans from (A) the *Hbt. salinarum* S-layer glycoprotein, (B) the *M. fervidus* S-layer glycoprotein and (C) *S. acidocaldarius* cytochrome *b*$_{558/566}$ are presented. In (A), the upper glycan is found at *Hbt. salinarum* S-layer glycoprotein Asn-2, while the lower glycan is found at 10 other sequons. The abbreviations used here but not in the text are: fGal, galactofuranose; GalUA, galacturonic acid; GlcUA, glucuronic acid; Man, mannose; Me, methyl; Qui, quinovose.

Methanogenic Archaea, possessing the ability to generate
CH$_4$ from CO$_2$ and H$_2$ or other carbon sources and detected
in environments spanning a broad spectrum of temperature, sa-
linity, pressure and pH (Thauer et al. 2008), also express
glycoproteins bearing N-linked glycans. For example, the
*Methanothermus fervidus* S-layer glycoprotein is modi-
ified by a hexasaccharide composed of a *N*-acetylgalactosamine (Gal-
NAc) subunit linked to the target asparagine residue, three
mannose residues and finally, two methylated hexoses (man-

Not long after the *Hbt. salinarum* S-layer glycoprotein was
identified as the first nonenukaryal N-modified glycoprotein,
other similarly modified archaeal polypeptides were described.
These included the *Hbt. salinarum* flagellin, shown to bear the
same glycan-linked sulfated polysaccharide as the S-layer gly-
copein from this species (Wieland et al. 1985), and the S-
layer glycoprotein of *Halofexx volcanii*, a halophilic species
first isolated from the Dead Sea (Mullakanbhai and Larsen
1975; Sump et al. 1990). Since, N-linked glycans presenting
eous diversity in their sugar composition have been iden-
tified in numerous archaeal proteins derived from a variety of
species isolated from a wide range of environmental niches.
voltae" glycan at this position. The *M. maripaludis* glycan is, however, capped by the unique terminal sugar, 2-acetamido-2,4-dideoxy-5-O-methyl-hexos-5-ulo-1,5-pyranose. This apparently represents the first example of a naturally occurring diglycoside of an aldulose (Kelly et al. 2009).

Thermophilic and hyperthermophilic Archaea, the latter growing optimally at temperatures above 80°C, also contain experimentally confirmed N-modified glycoproteins. *Thermoplasma acidophilum*, growing optimally at 56°C and pH 2, lacks a cell wall but is surrounded by a highly glycosylated plasma membrane reported to contain a branched glycan largely based on mannose subunits yet also containing glucose and galactose moieties, asparagine-linked through GlcNAc (Yang and Haug 1979a). Cytochrome b558/566 from *Sulfolobus acidocaldarius*, originally isolated from solfatara regions in Yellowstone National Park and shown to grow optimally at 75 to 80°C and pH 2 to 3 (Brock et al. 1972), is modified by a hexasaccharide that includes a glucose subunit, two mannose residues, a 6-sulfoquinovose subunit and two GlcNAc groups, one of which serves as the linking sugar (Figure 1C; Hettmann et al. 1998; Zähringer et al. 2000). Although considered common in glycolipids of chloroplasts and photosynthetic Bacteria (Imhoff et al. 1982), 6-sulfoquinovose (or 6-deoxy-6-sulfoglucose) had not been previously detected in a glycoprotein.

For a listing of additional confirmed or proposed N-modified archaeal glycoproteins, along with the evidence supporting the claim for glycan modification, the reader is directed to the review by Eichler and Adams (2005).

**Not like everybody else**

Today, it is clear that organisms from all three domains are capable of performing N-glycosylation (Weerapana and Imperiali 2006; Abu-Qarn and Eichler, et al. 2008). Still, archaeal N-glycosylated proteins can be distinguished from their eukaryal and bacterial counterparts not only by the extremophilic nature of their source species or the unusual sugars employed but also in terms of the amino acid sequence surrounding modified sequons and with respect to how oligosaccharides are N-linked to target proteins (Table 1).

<table>
<thead>
<tr>
<th>Table 1. N-glycosylation in the three domains of life</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Eukarya</strong></td>
</tr>
<tr>
<td>Essential property? Yes</td>
</tr>
<tr>
<td>Sugar donor Nucleotide-activated sugars</td>
</tr>
<tr>
<td>Dolichol-phosphate-bound glycans</td>
</tr>
<tr>
<td>Flippase Process ATP independent</td>
</tr>
<tr>
<td>Identity Rf1 may be involved in flipping</td>
</tr>
<tr>
<td>OST Oligomeric status Multimeric</td>
</tr>
<tr>
<td>Catalytic subunit St3</td>
</tr>
<tr>
<td>Localization Lumenal face of the ER membrane</td>
</tr>
<tr>
<td>Sequon recognized N-X-S/T; X≠P</td>
</tr>
<tr>
<td>Type E</td>
</tr>
<tr>
<td>Modification of N-linked glycan? Trimming in ER, elaboration in Golgi</td>
</tr>
</tbody>
</table>

Based motif reportedly employed in bacterial (i.e. *Campylobacter jejuni*) N-glycosylation where X and Z are any residue but proline (Kowarik et al. 2006; Abu-Qarn and Eichler 2007). Still, although relatively few sequons in archaeal glycoproteins have been experimentally verified as being modified, the amino acid composition within and surrounding these sequons can be distinguished from the comparable positions in eukaryal glycoproteins (Abu-Qarn and Eichler 2007). For instance, if the modified Asn is considered as position 0, there is a high probability of finding aromatic residues at positions −2 and −1, a small hydrophobic residue (Gly or Val) at the X position and a larger hydrophobic residue (Ile, Leu, Met, Phe, Trp or Tyr) at the +1 position in a eukaryal glycoprotein (Ben-Dor et al. 2004; Petrescu et al. 2004). In contrast, in only few archaeal sequons confirmed as being modified are Phe, Trp or Tyr detected at positions −2 and −1, while Ser or Thr are readily found at the X position and Ala and Gly predominate at the +1 position. No example in Archaea of an aromatic residues being present at the +1 position has been reported. Indeed, the only time a Tyr was detected at this site was in a sequon experimentally shown to be nonmodified (Lechner and Sumper 1987). In vitro analysis of oligosaccharyltransferase (OST) activity in the hyperthermophile, *Pyrococcus furiosus*, revealed that modification of an Asn-X-Thr-bearing reporter peptide was more efficient than was processing of the same peptide presenting an Asn-X-Ser sequon, with either the sequon Thr or Ser being essential for glycosylation and the presence of Pro at the X position preventing sequon modification (Igura et al. 2008). On the other hand, as elaborated below in the section dealing with the archaeal OST, modifications in archaeal sequon composition are tolerated in *Hbt. salinarum*, suggesting that predictive algorithms may have overlooked archaeal glycoproteins bearing similarly modified or even novel N-glycosylation sites (Zeitler et al. 1998).

The linking sugar serving to attach an oligosaccharide to a modified Asn residue also differs in glycoproteins across the three domains of life. In bacterial N-glycoproteins characterized thus far, bacillosamine (2,4-diamino-2,4,6-trideoxyglycopyranose) serves as the linking sugar (Young et al. 2002), whereas in the vast majority of eukaryal N-glycoproteins, GlcNAc serves this role (Spiro 1973). Rare instances of eukaryal reliance on other linking sugars have been reported, such as glucose in the case of laminin (Schreiner et al. 1994). In Archaea, a va-
riety of saccharides have been shown to serve as linking sugars, including glucose, GlcNAc and GalNAc. Indeed, as noted above, in the Section Archeal N-linked glycoproteins: here, there and everywhere, two different linking sugars can be employed within the same Archaeal glycoprotein, as exemplified by the Hbt. salinarum S-layer glycoprotein. The ability to use different linking sugars in the same glycoprotein is especially intriguing when one considers that Hbt. salinarum seemingly only encodes a single OST (Magidovich and Eichler 2009; see below).

**Early insight into the archaean N-glycosylation pathway**

Encouraged by the conclusion made at the time that N-glycosylation was restricted to Eukarya and Archaia, initial efforts aimed at defining the pathway responsible for this posttranslational modification in Archaea sought further similarities between the archaean pathway and its eukaryal counterpart.

**The sugar carrier**

As in Eukarya (Burda and Aebi 1999), the archaean N-linked oligosaccharide is apparently assembled on a dolichol carrier, rather than the undecaprenol carrier used in bacterial N-glycosylation (Szymanski and Wren 2005). Accordingly, Archaea contain both dolichol phosphate and dolichol pyrophosphate bearing either mono- or polysaccharides. Specifically, *Hbt. salinarum* has been reported as containing C60-dolichols carrying glucose, mannose and GlcNAc units, as well as a sulfated tetrasaccharide (Mescher et al. 1976; Lechner et al. 1985a). *Hfx. volcanii* contains mannosyl-(1-4)-galactosyl phosphodolichol, lesser quantities of sulfated or phosphorylated dihexosyl phosphodolichol and dolichol phosphate bearing a tetrasaccharide comprising mannose, galactose and rhamnose subunits. In this haloarchaeon, the glycans were all linked to C55- and/or C60-dolichol moieties (Kuntz et al. 1997). The saccharide-charged dolichol species in *Hfx. volcanii* are unusual in that the α-terminal isoprene unit is saturated and since only monophosphorylated dolichol is observed.

Proof for the involvement of such glycan-charged dolichols in archaean protein N-glycosylation has also been provided. The transfer of radiolabeled glucose from uridine diphosphate (UDP)-[3H]glucose to glycoproteins was shown to proceed through a glucose-containing phosphopolyisoprenol intermediate in *Hfx. volcanii* (Zhu et al. 1995). Treatment with bacitracin, a compound that interferes with recycling of pyrophosphodolichol polysaccharide carriers following release of their bound oligosaccharides (Stone and Strominger 1971), was able to prevent modification of the *Hbt. salinarum* S-layer glycoprotein by the glucose-linked sulfated glycan otherwise added to 10 sequons of the protein (Mescher and Strominger 1978). In contrast, bacitracin treatment failed to prevent addition of the GlcNAc-linked repeating sulfated pentasaccharide at Hbt. salinarum S-layer glycoprotein Asn-2, suggesting the use of dolichol phosphate rather than dolichol pyrophosphate as the carrier for this glycan (Wieland et al. 1980). Indeed, the reportedly exclusive use of phosphodolichol sugar carriers in *Hfx. volcanii* N-glycosylation is reflected in the inability of bacitracin to prevent such protein modification in this species (Kuntz et al. 1997; Eichler 2001). Further evidence linking dolichol-charged glycans to N-glycosylation comes with the observation that in *Hbt. salinarum*, the glycan moiety of the pyrophosphodolichol-bound sulfated polysaccharide is also detected on the S-layer glycoprotein and flagellins in this species (Lechner et al. 1985a; Wieland et al. 1985). The sulfated polysaccharide is methylated in the pyrophosphodolichol-linked form but not when protein-bound (Lechner et al. 1985b). In contrast, the hexasaccharide moiety attached to the *M. fervidus* S-layer glycoprotein retains the methylation introduced at the dolicholpyrophosphate carrier level (Hartmann and König 1989; Kächer et al. 1993), as also seems to be the case in *Hfx. volcanii*, where methylation of the S-layer glycoprotein-bound pentasaccharide is observed (Magidovich et al. 2010).

**Facing the outside**

When the topology of N-glycosylation was considered, further parallels between the eukaryal and archaean pathways were shown, with evidence obtained pointing to archaean N-glycosylation as occurring on the outer surface of the cell, the topological equivalent of the luminal-facing leaflet of the endoplasmic reticulum (ER) membrane bilayer, the site of N-glycosylation in Eukarya. Localization of archaean N-glycosylation to the external cell surface is supported by studies showing the ability of *Hbt. salinarum* cells to modify cell-impermeable, sequon-bearing hexapeptides with sulfated oligosaccharides (Lechner et al. 1985a). Although unable to cross the halaarchaeal plasma membrane (Mescher and Strominger 1978), bacitracin is nonetheless able to interfere with N-glycosylation in *Hbt. salinarum*, ultimately preventing the transfer of sulfated oligosaccharides to the S-layer glycoprotein (Mescher and Strominger 1978; Wieland et al. 1980). Additional support for archaean N-glycosylation occurring on the outer cell surface came with the description of membrane-bound pyrophosphatases that orient their active site to the exterior (Meyer and Schäfer 1992; Amano et al. 1993). It was proposed that such enzymes could participate in the archaean N-glycosylation process by dephosphorylating dolichol pyrophosphate, presumably following transfer of oligosaccharides from the lipid carrier to protein targets (Meyer and Schäfer 1992). Accordingly, an externally oriented membrane-bound pyrophosphatase from the thermoacidophile *Sulfolobus tokodaii* has been recently shown to be able to hydrolyze isopentenylpyrophosphate and geranylpyrophosphate (Manabe et al. 2009). Finally, studies supporting a cotranslational mode of membrane protein insertion in Archaea also lend support to protein glycosylation transpiring on the exterior surface of the cell (Gropp et al. 1992; Dale and Krebs 1999; Ring and Eichler 2004).

**Enzymes of glycosylation**

In addition to these early efforts aimed at delineating steps and components of the archaean N-glycosylation pathway, several groups also addressed enzymes putatively involved in this posttranslational modification, including enzymes proposed to participate in the assembly of the oligosaccharide-charged phosphodolichol carrier. GlcNAc transferase activity was partially characterized from *Hbt. salinarum* membranes (Mescher et al. 1976). Photo-affinity experiments using 5-azido-[32P] UDP-glucose identified a putative phosphodolichol glucose synthase in *Hfx. volcanii* homogenates (Zhu et al. 1995), while phosphodolichol mannose synthase, susceptible to inhibition...
by amphotericin, was purified from *T. acidophilum* (Zhu and Laine 1996).

Thus, the first two decades of research into archaeal *N*-glycosylation provided glimpses of selected aspects of the process, in a variety of model systems. Detailed understanding of the archaeal version of this posttranslational modification would have to, however, wait until the dawn of the genome era.

**Homology shows the way**

The examination of Archaea at the genome level began in 1996, when the first complete sequence of an archaeon (corresponding to the fourth genome sequenced overall), namely that of the methanogen *Methanocaldococcus* (then *Methanococcus*) *jannaschii*, was published (Bult et al. 1996).

With the additional archaeal genome sequences now available (today, over 140 archaeal genomes are at various stages of completion (http://genomesonline.org/index2.htm; Oct., 2009)), bioinformatics approaches were enlisted in the search for components of the archaeal *N*-glycosylation pathway. Analysis of the genome of *Archaeoglobus fulgidus*, as a sulfur-metabolizing organism that grows optimally at 83°C (Klenk et al. 1997), first revealed the presence of gene clusters containing sequences predicted to encode elements of a *N*-glycosylation pathway (Burda and Aebi 1999). The major breakthrough in identifying components of archaeal *N*-glycosylation pathways soon followed, coming from efforts combining the identification of archaeal homologues of known elements of the parallel eukaryal and bacterial pathways, together with gene deletion and subsequent analysis of the effects of such deletion on the *N*-glycosylation of reporter glycoproteins. In this manner, Abu-Qarn and Eichler (2006) and Chaban, Voisin, et al. (2006), respectively, studying *N*-glycosylation in *Hfx. volcanii* and *M. voltae*, simultaneously identified gene products experimentally verified as participating in this posttranslational modification.

**Sweet and salty: *N*-glycosylation in *Haloferax volcanii***

As the major polypeptide in the species and previously shown to be amenable to the study of other posttranslational modifications (e.g. signal peptide cleavage (Fine et al. 2006) and protein lipid modification (Konrad and Eichler 2002)), the *Hfx. volcanii* S-layer glycoprotein represents an excellent model for addressing archaeal *N*-glycosylation. At the time of its description (Sumper et al. 1990; Mengele and Sumper 1992), it was reported that of the seven putative *N*-glycosylation sites present in the S-layer glycoprotein, Asn-13 and Asn-505 were each modified by a linear chain of β-1-4 linked glucose resi-

---

**Fig. 2.** Working models of archaeal *N*-glycosylation pathways. Currently identified components of the *N*-glycosylation pathways of (A) *Hfx. volcanii*, (B) *M. voltae* and (C) *M. maripaludis*. The abbreviations used are: dolP, dolichol phosphate; dolPP, dolichol pyrophosphate; NDP, nucleoside diphosphate; 1P, 1 phosphate. Dolichol, embedded in the schematically-depicted plasma membrane (drawn as the two parallel horizontal lines), is represented by a vertical line. The identities of the sugars represented by each shape are listed next to Asn-linked glycan in each panel. In (B), the optional introduction of a 220 or 260 Da sugar seen in some *M. voltae* species (Chaban et al. 2009) is reflected by the use of parentheses.
dues, while Asn-274 and/or Asn-279 were described as bearing a glucose-, idose- and galactose-containing polysaccharide. Later efforts relying on more sophisticated mass spectrometry tools revealed the *Hfx. volcanii* S-layer glycoprotein to instead be modified at Asn-13 and Asn-83 by a pentasaccharide comprising two hexoses, two hexuronic acids and a 190 Da species (Abu-Qarn et al. 2007), subsequently shown to correspond to a methyl ester of hexuronic acid (Magidovich et al. 2010). It was also shown that the sequon at Asn-370 is not modified (Abu-Qarn et al. 2007).

In addition to redefining the N-linked glycan profile of the S-layer glycoprotein, these more recent studies also served to define components of the *Hfx. volcanii* N-glycosylation pathway. The initial use of bioinformatics tools to identity *Hfx. volcanii* homologues of eukaryal and bacterial N-glycosylation genes (Abu-Qarn and Eichler 2006) led to the identification of open reading frames, subsequently remained *agl* (archaeal glycosylation) genes following the nomenclature of Chaban, Voisin, et al. (2006) and shown to encode for glycosyltransferases (GTs) (i.e. *AglD, AglE, AglI* and *AglJ*), an OST (i.e. *AglB*) and other sugar-processing enzymes (i.e. *AglF*) involved in the assembly and attachment of the pentasaccharide decorating the *Hfx. volcanii* S-layer glycoprotein (Abu-Qarn et al. 2007; Abu-Qarn, Giordano, et al. 2008; Yurist-Doutsch et al. 2008). *aglG*, encoding a putative GT, was next identified through it being positioned between the *aglB* and *aglI* sequences (Yurist-Doutsch et al. 2008). Indeed, when the various *agl* sequences were mapped to the *Hfx. volcanii* genome, it was noted that apart from *aglD*, all were sequestered to a gene island stretching from *aglJ* to *aglB* (Yurist-Doutsch and Eichler 2006). However, as the current annotation of the *Hfx. volcanii* genome did not recognize the *aglE* sequence as an open reading frame (Abu-Qarn, Giordano, et al. 2008), the *agl* gene island was subjected to manual reannotation. In this manner, additional sequences, i.e. *aglP, aglQ* and *aglR*, were identified (Yurist-Doutsch and Eichler 2009), while *aglM* was shown to lie beyond the original gene island borders (Yurist-Doutsch et al. 2010).

Although each of the identified *Hfx. volcanii* Agl proteins has been assigned a specific role through a combination of gene deletion and mass spectrometry approaches, only AglF, AglM and AglP have been characterized biochemically. AglF was shown to be a glucose-1-phosphate uridyltransferase involved in the biosynthesis of the hexuronic acid found at position three of the S-layer glycoprotein-linked pentasaccharide (Yurist-Doutsch et al. 2010). AglM, a UDP-glucose dehydrogenase, was shown to participate in the biosynthesis of the hexuronic acid found at pentasaccharide position two and likely of the hexuronic acids found at positions three and four, as well (Yurist-Doutsch et al. 2010). In a combined in vitro reconstitution experiment, AglF and AglM were shown to work in a coordinated manner to generate UDP-glucuronic acid from glucose-1-phosphate and uridine triphosphate (UTP) in a NAD⁺-dependent manner (Yurist-Doutsch et al. 2010). Finally, AglP was confirmed to be a S-adenosyl-L-methionine-dependent methyltransferase responsible for the formation of the methyl ester of hexuronic acid found at position four of the pentasaccharide (Magidovich et al. 2010). The current working model of the *Hfx. volcanii* N-glycosylation pathway is presented in Figure 2A.

**N-glycosylation in the methanogens**

Following the discovery that the same novel N-linked trisaccharide is attached to both the S-layer glycoprotein and flagellins in *M. voltae* (Voisin et al. 2005), efforts focused on defining the pathway of N-glycosylation in this obligate anaerobic methanogen. The fact that the same N-linked glycan decorates both the S-layer glycoprotein and flagellins in *M. voltae* (as is also the case in *Hbt. salinarum* (Wieland et al. 1985)) points to a common pathway for N-glycosylation in a given species. As with *Hfx. volcanii*, studies of the N-glycosylation pathway of *M. voltae* began with the search for homologues of known eukaryal or bacterial N-glycosylation genes (Chaban, Voisin, et al. 2006). Accordingly, this strategy led to the identification of *aglA*, shown to participate in the addition of the terminal sugar residue of the trisaccharide, namely N-acetyl mannnuronic acid linked to threonine, as well as of the OST, *aglB*. Through its ability to complement a conditionally lethal *alg7* mutation in yeast, it was concluded that AglH is a GlcNAc-1-phosphate transferase, catalyzing the transfer of UDP-activated GlcNAc to dolichol pyrophosphate and responsible for the first step in the *M. voltae* N-glycosylation process (Shams-Eldin et al. 2008). Finally, the second sugar residue of the trisaccharide, a diacetylated glucuronic acid, is added by the combined actions of two enzymes, AglC and AglK (Chaban et al. 2009), as reflected in Figure 2B.

The N-linked glycan decorating the flagellins of *M. maripaludis* corresponds to a tetrasaccharide in which the sugar residues found at positions two and three are essentially identical to the two outer sugar subunits of the *M. voltae* N-linked trisaccharide (Kelly et al. 2009). As such, *M. maripaludis* AglA was identified as participating in the addition of the sugar subunit at position three of the tetrasaccharide, an acetylated and acetamidino-modified mannuronic acid linked to threonine (VanDyke et al. 2009). However, given that the linking sugar of the *M. maripaludis* glycan is GalNAc, rather than the GlcNAc employed in *M. voltae* glycoproteins, the *M. maripaludis* enzyme involved in adding the diacetylated glucuronic acid to position two of the tetrasaccharide, AglO, is not homologous to either *M. voltae* AglC or AglK (VanDyke et al. 2009). Finally, AglL was assigned as participating in either the attachment of the threonine to the third sugar residue or the addition of the terminal sugar, the novel 2-acetamido-2,4-dideoxy-5-O-methyl-hexos-5-ulo-1,5-pyranose (VanDyke et al. 2009) (Figure 2C).

**Pass the sugar: AglB, the archaeal OST**

OSTs serve to transfer lipid-linked polysaccharides to select Asn residues of target proteins. While the OST of higher Eu- karya exists as a multimeric complex based on the Stt3 subunit (Zufferey et al. 1995), the OST of Archaea, like that of Bacteria, comprises only a single component, i.e. AglB (Abu-Qarn and Eichler 2006; Chaban, Voisin, et al. 2006; Igura et al. 2008). AglB was first identified in *Hfx. volcanii* and *M. voltae*, when it was shown that deletion of the encoding gene led to a loss of N-glycosylation of reporter glycoproteins (Abu-Qarn and Eichler 2006; Chaban, Voisin, et al. 2006; Abu-Qarn et al. 2007). Like its eukaryal (i.e. Stt3) and bacterial (i.e. PglB) counterparts, AglB comprises a multiple membrane-spanning
Insight into OST action in Archaea, and indeed, across evolution, has been provided by solution of the three-dimensional structure of the C-terminal soluble domain of *P. furiosus* AglB (*Igura et al. 2008*). Providing the first portrayal of a catalytic β-helix-based central core and the DxxK motif to DxxKxxx[MI], now termed the MI motif in the bacterial enzyme. Based on this observation, a reassessment of sequence alignment data extended the *P. furiosus* AglB DxxK motif is replaced by a MxxI motif in the bacterial enzyme. Based on this observation, a reassessment of sequence alignment data extended the *P. furiosus* AglB DxxK motif to DxxKxxx[MI], now termed the MI motif. The recent solution of the *C. jejuni* PglB structure (*Maita et al. 2010*) revealed an impressive degree of folding similarity to *P. furiosus* AglB. However, such comparison also showed that rather than being universal in Stt3/PglB/AglB family members, the *P. furiosus* AglB DxxK motif is replaced by a MxxI motif in the bacterial enzyme. Based on this observation, a reassessment of sequence alignment data extended the *P. furiosus* AglB DxxK motif to DxxKxxx[MI], now termed the MI motif. The detailed analysis of these motifs across Stt3/PglB/AglB family members in the three domains of life was addressed, the differential distribution of the MI and DK motifs, as well as of a variant of the DK motif comprising a DxxMxxx[KI] signature (the DM motif), allowed evolutionary patterns to be drawn. It was shown that together
with the WWDYG motif, eukaryal Stt3 proteins and archaeal AglB proteins from members of the archaeal phyla Crenarchaeota (including *P. furiosus*) and some members of the phylum Euryarchaeota employ the DK motif to present an E-type catalytic center, as do the sole sequenced members of the archaeal phyla, Korarchaeota, Nanoarchaeota and Thaumarchaeota. Along with bacterial PglB proteins, other members of Euryarchaeota, including *M. voltae* and *M. maripaludis*, employ the WWDYG and MI motifs to form the B-type catalytic center. Yet other Euryarchaeota, such as *Hfx. volcanii*, combine the WWDYG and DM motifs to create the A-type catalytic center. Archaeal AglB proteins thus contain catalytic centers of all three types (Table II). Moreover, in those archaeal species encoding multiple AglB sequences (cf. Magidovich and Eichler 2009), only a single type of catalytic center is found in all of the predicted AglB proteins of that species.

Bioinformatics-based assignment of multiple versions of AglB being present within a given archaeal species still requires experimental confirmation that each sequence is in fact expressed and capable of OST activity. Indeed, justification for the existence of multiple versions of AglB in a single species is not obvious if one considers the semipropriety of *Hbt. salinarum* AglB. In this species, it was shown that replacement of the Ser-4 residue of the S-layer glycoprotein 7Asn-7Ala-7Ser sequn with Val, Leu or Asn did not prevent N-glycosylation at the Asn-2 position (Zeitler et al. 1998). Moreover, the same OST is apparently responsible for adding both the repeating sulfated pentasaccharide moiety through a GalNAc link to Asn-2 as well as the sulfated polysaccharide unit attached via a glucose subunit at 10 other N-glycosylation sites of the S-layer glycoprotein (Lechner and Wieland 1989).

Putative N-glycosylation pathway components in other Archaea

Although pathways of N-glycosylation have only been outlined in *Hfx. volcanii*, *M. voltae* and *M. maripaludis*, N-glycosylated proteins have been identified in numerous archaeal species found in a variety of environments (cf. Eichler and Adams 2005). In the vast majority of these species, nothing is known of the N-glycosylation process.

As a first step to redressing this situation, the Carbohydrate-Active Enzymes (CAZy) database (http://www.cazy.org/fam/acc_GT.html; Cantarel et al. 2009) was consulted to address the presence and distribution of predicted GTs in 56 archaeal genomes (Magidovich and Eichler 2009). In addition to identifying AglB in all but two species considered (i.e. *Aeropyrum pernix* and *Methanopyrus kandleri*), it was also shown that GTs assigned to the AglB family and the GT-B fold-containing retaining GT4 family as being the prototypes that eventually gave rise to the array of glycosyltransferases seen today (Lairson et al. 2008). Furthermore, unlike what is observed in Eukarya and Bacteria, where known and predicted GTs are respectively assigned to 67 and 54 different GT families, proteins identified as archaeal GTs are distributed among only 13 GT families.

In *Hfx. volcanii*, aglB is found within a gene cluster that includes the agl GTs of this species, with the exception of aglD, as well as non-GT-encoding agl sequences serving other N-glycosylation-related roles (Yurist-Doutos and Eichler 2009; Magidovich and Eichler 2010; Yurist-Doutos et al. 2010). In contrast, the agl genes of *M. maripaludis S2* are more widely distributed in the genome, with only aglC being found in proximity to aglB (VanDyke et al. 2009). In other Archaea, varying degrees of such aglB-centered clustering of GT-encoding genes are seen (Magidovich and Eichler 2009). While substantial gene clustering is observed in halophiles and many methanogens, this arrangement is seen 3-fold less often in hyperthermophilic species. The evolutionary significance of this observation remains to be considered.

In addition to the different verified and predicted N-glycosylation pathway components described in halophilic and methanogenic archaea, additional archaeal sugar-modifying enzymes have been reported. For instance, the pathway of UDP-acetamido sugar biosynthesis has been described for two members of the class Methanococcales, namely *M. maripaludis* and *M. jannaschii*. In each case, the pathway enzymes having been purified following expression in *Escherichia coli* and biochemically characterized (Namboori and Graham 2008). Specifically, *M. maripaludis* MMP1680 (or *M. jannaschii* MJ1420) catalyzes the isomerization and transamination of fructose-6-P to create α-d-glucosamine-1-phosphate, which is converted to α-d-glucosamine-1-phosphate by the *M. maripaludis* MMP1077 (or *M. jannaschii* MJ1100) phosphomutase. *M. maripaludis* MMP1076 (or *M. jannaschii* MJ1101) then catalyzes both the acetylation of glucosamine-1-phosphate and the transfer of the resulting GlcNAc-1-phosphate to UTP to generate UDP-GlcNAc and pyrophosphate. *M. maripaludis* MMP0705 (or *M. jannaschii* MJ1504) can isomerize UDP-GlcNAc to create UDP-N-acetyl-α-d-mannosamine (UDP-ManNAc), which is oxidized by *M. maripaludis* MMP0706 (or *M. jannaschii* MJ0048) to produce UDP-N-acetylmannosaminuronate (UDP-ManNAcA). Given the homologies of *M. maripaludis* MMP1680 (or MJ1420) to bacterial GlmS, of MMP1077 (or MJ1100) to GlmM and of MMP1076 (or MJ1101) to GlmU, these studies point to methanoarchaea as producing acetamido sugars, e.g. GlcNAc, using the bacterial pathway (Mengin-Lecreulx and van Heijenoort 1994), rather than the Leloir pathway used by Eukarya (Milowski et al. 2006). Still, despite the fact that acetamido sugars have been detected in the N-linked glycans of methanoarchaeal glycoproteins (see section on N-glycosylation in methanogens and Figure 1), the involvement of these enzymes in N-glycosylation remains to be demonstrated. The same is true of a *P. furiosus* sugar nucleotidytransferase with extremely broad sugar and nucleotide substrate specificity (Mizanur et al. 2004) as well as of phosphohexomutases from *Pyrococcus horikoshii* (Akutsu et al. 2005), *Sulfolobus solfataricus* (Ray et al. 2005) and *Thermococcus kodakaraensis* (Rashid et al. 2004), reported to possess...
both phosphoglucomutase and phosphomannomutase activities. In contrast, deletion of the gene encoding the putative *M. maripaludis* acetyltransferase MMP0350 impaired flagellin N-glycosylation (VanDyke et al. 2008), a process known to involve acetylated sugars (Kelly et al. 2009).

Finally, in *Pyrolobus fumarrii*, growing optimally at 106°C (Blöchl et al. 1997), novel UDP-sugars, including UDP-β-GlcNAc-3-NAc and UDP-β-GlcNAc3NAc-(4-1)-β-GlcNAc3NAc, have been reported (Gonçalves et al. 2008). Nothing, however, is known of their biosynthesis, whether they are employed in N-glycosylation or, indeed, whether *P. fumarrii* even performs this post-translational modification.

**Roles of N-glycosylation in Archaea**

Whereas numerous N-glycosylated proteins have been described in a variety of Archaea living across a wide range of environmental niches, the roles served by this post-translational modification remain, for the most part, unexplored. Indeed, as *aglB* can be deleted from both *Hfx. volcanii* and *M. voltae* (Abu-Qarn and Eichler 2006; Chaban, Voisin, et al. 2006), it would seem that N-glycosylation is not essential for the survival of these species, at least under the conditions tested. Nonetheless, N-glycosylation may contribute to the ability of Archaea and their proteins to survive or adapt to the harsh environments in which these organisms can thrive.

In comparing the N-linked glycan profiles of two haloharchaeal S-layer glycoproteins (Mengele and Sumper 1992), it was noted that the *Hbt. salinarum* protein not only experiences a higher degree of N-glycosylation than does the same protein in *Hfx. volcanii* but that the glycans of former are enriched in sulfated glucuronic acids, in contrast to the neutral sugars found in the latter. Thus, relative to its *Hfx. volcanii* counterpart, the *Hbt. salinarum* S-layer glycoprotein presents a drastically increased surface charge density, a property thought to contribute to the stability of haloharchaeal proteins in the face of molar salt concentrations (Madern et al. 2000). Accordingly, the *Hbt. salinarum* S-layer glycoprotein also contains 20% more acidic amino acid residues than does the *Hfx. volcanii* S-layer glycoprotein (Lechner and Sumper 1987; Sumper et al. 1990). As a result of these considerations, *Hbt. salinarum* is able to grow in higher salt concentrations than does *Hfx. volcanii*. In *Hfx. volcanii*, however, absent or defective N-glycosylation greatly reduced the ability of cells to grow at increasing salt concentrations (Abu-Qarn et al. 2007). The protection that an enhanced negative surface charge could also afford in the face of acidic conditions has been suggested as the reason for N-glycosylation of *S. acidocaldarius* cytochrome b₅₅₈/b₅₆₆ (Hettmann et al. 1998; Zähringer et al. 2000). Moreover, it was proposed that much of the protein surface is shielded from the pH 2 environment in which these cells exist by the high degree of N-glycosylation it experiences (Zähringer et al. 2000). It has also been postulated that the N-linked glycan of the *M. fervidus* S-layer glycoprotein is involved in the stabilization of this surface protein at high temperatures (Kärcher et al. 1993). Available evidence also points to Archaea as being able to modulate the N-glycosylation profile of target proteins. For instance, glycosylation of *Methanospirillum hungatei* flagellins was reported to only occur in low phosphate media (Southam et al. 1990). In *Hfx. volcanii*, the transcription of the *agl* genes involved in N-glycosylation occurs in a coordinated yet differential manner in the face of different growth paradigms, pointing to this post-translational modification as serving an adaptive role (Yurist-Doutsch et al. 2008; Yurist-Doutsch et al. 2010).

In addition to possibility affording advantages in the face of environmental challenges, archaeal N-glycosylation has been cited as providing structural support. This is best exemplified in *Hbt. salinarum*, where bacitracin treatment transformed rod-shaped cells into spheres (Mescher and Strominger 1976b). In cell wall-lacking *T. acidophilum*, the glycan moieties of the major membrane glycoprotein coating the cell surface have been suggested to either trap water molecules or encourage interaction between cell surface proteins. In either case, protein glycosylation would contribute to cell surface rigidity (Yang and Haug 1979b). N-glycosylation of archaeal proteins has also been implicated in protein assembly and function. Defective *Hfx. volcanii* N-glycosylation resulted in an unstructured S-layer, comprised solely of the S-layer glycoprotein (Sumper et al. 1990), while the absence of N-glycosylation compromised S-layer stability (Abu-Qarn et al. 2007). Missing or partial *M. voltae* flagellin N-glycosylation led to a lack or reduction in flagella numbers, concomitant with motility defects (Chaban, Voisin, et al. 2006). Likewise, bacitracin-mediated interference with *Methanococcus deltae* flagellin glycosylation resulted in a loss of flagellation (Kalmokoff et al. 1992). In mutant *Hbt. salinarum* cells over-producing under-glycosylated flagellins, increased levels of flagella were detected in the growth medium, implying that proper flagellin glycosylation is necessary for effective flagellar incorporation into the plasma membrane (Wieland et al. 1985). In addition, glycosylation appears to play a role in stabilization against proteolysis, as reflected by the loss of *Hfx. volcanii* S-layer glycoprotein resistance to added protease in mutants lacking many of the Agl proteins involved in assembling and attaching the pentasaccharide N-linked to the protein (Yurist-Doutsch et al. 2008; Yurist-Doutsch et al. 2010). Finally, N-glycosylation has been proposed to modulate the interaction of binding proteins with the cell membrane or envelope of *S. acidocaldarius* (Albers et al. 2004), while in *Methanothermus sociabilis*, the concept of N-glycosylation playing a role in cell aggregation has been raised (Kärcher et al. 1993).

**Archaeal O-glycosylation**

In addition to N-glycosylation, archaeal proteins can also experience O-glycosylation. In the cases of both the *Hbt. salinarum* and the *Hfx. volcanii* S-layer glycoproteins, Thr-rich regions adjacent to the predicted membrane-spanning domain of the protein are modified with galactose–glucose disaccharides (Mescher and Strominger 1976a; Sumper et al. 1990). Unfortunately, virtually nothing is known of the archaeal O-glycosylation pathway at present.

**Concluding remarks**

Although initially described over 30 years, only recently have major advances in delineating the archaeal pathway
of N-glycosylation pathway been made. What has been learned thus far hints that continued examination of the archaean version of this universal posttranslational modification will provide insight into the process not obtained studying either the parallel and better-characterized eukaryal or bacterial pathways. Moreover, as is so often the case when a given biological process is considered from the archaean perspective, analysis of N-glycosylation in a limited number of model systems has revealed facets of the process seemingly unique to members of this domain of life. On the other hand, particularly from the perspective of structural biology, the study of N-glycosylation in Archaea has served to elucidate aspects of the process common to all three domains of life. The on-going development of improved molecular tools for working with a variety of archaean strains growing across a range of environmental conditions ensures that these unusual microorganisms will continue to expand our understanding of N-glycosylation and other protein processing events.

Funding
Research in the Eichler laboratory is supported by the Israel Science Foundation (grant 30/07) and the US Army Research Office (grant W911NF-07-1-0260). L.K. is the recipient of a Negev-Zin Associates Scholarship.

Conflict of interest statement
None declared.

Abbreviations
ER, endoplasmic reticulum; GalNAc, N-acetylgalactosamine; GlcNAc, N-acetylglucosamine; GT, glycosyltransferase; OST, oligosaccharidyltransferase; rRNA, ribosomal RNA; S-layer, surface layer; UDP, uridine diphosphate; UTP, uridine triphosphate.

References


