

## An analysis of amino acid sequences surrounding archaeal glycoprotein sequons

MEHTAP ABU-QARN<sup>1</sup> AND JERRY EICHLER<sup>1,2</sup>

<sup>1</sup> Department of Life Sciences, Ben Gurion University, Beersheva 84105, Israel

<sup>2</sup> Corresponding author (jeichler@bgu.ac.il)

Received June 20, 2006; accepted July 26, 2006; published online...

**Summary** Despite having provided the first example of a prokaryal glycoprotein, little is known of the rules governing the *N*-glycosylation process in Archaea. As in Eukarya and Bacteria, archaeal *N*-glycosylation takes place at the Asn residues of Asn-X-Ser/Thr sequons. Since not all sequons are utilized, it is clear that other factors, including the context in which a sequon exists, affect glycosylation efficiency. As yet, the contribution to *N*-glycosylation made by sequon-bordering residues and other related factors in Archaea remains unaddressed. In the following, the surroundings of Asn residues confirmed by experiment as modified were analyzed in an attempt to define sequence rules and requirements for archaeal *N*-glycosylation.

**Keywords:** Archaea, *N*-glycosylation, oligosaccharide transferase, post-translational modification.

### Introduction

The covalent attachment of polysaccharides represents one of the most prevalent post-translational modifications proteins experience. In *N*-glycosylation, glycan moieties are linked to Asn residues found as part of NXS or NXT motifs referred to as sequons, where X is any residue apart from proline, although rare exceptions to these rules exist (Gavel and von Heijne 1990, Kaji et al. 2003). Although essential for *N*-glycosylation, the presence of a sequon is not sufficient for modification, given the large number of unmodified sequons in both glycosylated and non-glycosylated proteins (Ben-Dor et al. 2004, Petrescu et al. 2004).

To delineate the rules determining whether modification of a given sequon occurs, researchers have considered numerous and varied properties of the target protein. These include the amino acid composition and secondary structure of the sequon and its surroundings, the number of sequons in the protein, their position(s) relative to the protein termini or to other sequons in cases where multiple potential *N*-glycosylation sites exist. Accordingly, studies on eukaryal glycoproteins have pointed to the importance of the character of the amino acids occupying several positions both upstream and downstream of Asn residues experimentally verified to be glycan-

bearing. Based on consideration of the residues neighboring those eukaryal glycoprotein sequons experimentally shown to be modified, Ben-Dor et al. (2004) proposed a series of rules designed to predict whether a given sequon is modified. Similarly, a survey of some 500 nonredundant glycoproteins listed in the PDB database revealed an increased occurrence of aromatic residues at the +2 position following sequon Ser/Thr residues and that *N*-glycosylation often transpires at sites where changes in secondary structure occur (Petrescu et al. 2004). Other studies reported that the presence of a Pro residue immediately following a sequon most often hinders glycosylation (Mellquist et al. 1998). In addition, distinctions can be made between the rules governing the glycosylation of NXS and NXT sequons. For example, glycosylation occurs less frequently at NXS than at NXT sequons (Bause and Legler 1981, Kasturi et al. 1995). Furthermore, NXS glycosylation efficiency is affected by the character of the residues at the X position and at the +1 positions following the sequon serine (Shakin-Eshleman et al. 1996, Mellquist et al. 1998).

Once thought to be exclusively restricted to Eukarya, it is now clear that prokaryotes are also capable of *N*-glycosylation (Lechner and Wieland 1989, Szymanski and Wren 2005). In addressing *N*-glycosylated proteins in *Campylobacter jejuni*, the only bacterium for which a full *N*-glycosylation pathway has been described (Szymanski et al. 1999), it was observed that modification occurs at NXS/T sequons, where X is any residue apart from proline, as in Eukarya (Nita-Lazar et al. 2005). Although essential for bacterial *N*-glycosylation, the presence of a sequon is not sufficient for Asn modification. In a recent detailed analysis of *C. jejuni* glycoproteins (Kowarik et al. 2006), it was noted that only those Asn residues found as part of a D/EYNXS/T motif, where X and Y are not proline, experience glycosylation. By contrast, glycosylation of a given sequon in yeast did not require a negatively charged residue at the -2 position. Moreover, earlier analysis of eukaryal *N*-glycosylation sites revealed that Asp and Glu are disfavored at this position (Ben-Dor et al. 2004). Such differences in the sequence requirements of eukaryal and bacterial *N*-glycosylation likely reflect differences between eukaryal and bacterial oligosaccharide transferases (OST), responsible for transferring the lipid-linked glycan moieties to sequons in the target

protein (Kelleher and Gilmore 2006, Weerapana and Imperiali 2006). In Eukarya, OST comprises a multisubunit complex at the heart of which lies the Stt3 protein in yeast and its homologues in other species. In Bacteria, the Stt3 homologue PglB is thought to be solely responsible for OST activity (Wacker et al. 2002).

Although prokaryal *N*-glycosylation was first described in Archaea (Mescher and Strominger 1976, Lechner and Sumper 1987), still little is known of either the extent of such modification or the rules governing the process in this form of life. Indeed, while the ability to glycosylate proteins has been demonstrated in numerous species by a variety of experimental methods, the modification of glycoprotein Asn residues has been directly demonstrated in only a limited number of cases (Eichler and Adams 2005).

In this study, archaeal glycoprotein sequences were considered in an attempt to unmask any putative contribution to Asn glycosylation made by residues found in the positions bordering modified sequons. Defining such *N*-glycosylation sequence requirements in Archaea carries significant implications for understanding how the enzymes involved in the process (in particular the archaeal Stt3 homologue) operate. In addition, selected archaeal genomes were analyzed in an effort to predict the extent of *N*-glycosylation in Archaea because such information could provide insight into the relationship between post-translational modification and the ability of archaeal polypeptides to remain properly folded in the face of extreme physical conditions that would normally lead to protein denaturation, loss of solubility and aggregation.

## Materials and methods

### *Predicted N-glycosylated archaeal secretory proteins*

A database of proteins secreted by ten archaeal species (*Aeropyrum pernix*, *Archaeoglobus fulgidus*, *Halobacterium* sp. NRC-1, *Methanococcus jannaschii*, *Methanothermobacter thermoautotrophicus*, *Pyrococcus abyssi*, *Pyrococcus horikoshii*, *Sulfolobus solfataricus*, *Thermoplasma acidophilum* and *Thermoplasma volcanium*) was created as described previously (Bardy et al. 2003). Briefly, predicted protein-encoding genes from each organism were analysed using SignalP v.2.0 ([www.cbs.dtu.dk/services/SignalP-2.0](http://www.cbs.dtu.dk/services/SignalP-2.0)), designed to detect the presence of signal peptides. In these analyses, protein sequences were examined using the hidden Markov model with truncation set to 70 amino acids, trained on the different data sets available (Eukarya, Gram-positive or Gram-negative Bacteria). Proteins predicted as containing signal peptides by at least one data set were then subcellularly localized by Psort (Nakai and Horton 1999; [www.psort.nibb.ac.jp](http://www.psort.nibb.ac.jp)). Only those proteins considered to be secreted in Gram-positive organisms or localized to the periplasm or outer membrane of Gram-negative organisms (listed in Bardy et al. 2003) were scanned for the presence of sequons.

## Results and discussion

### *Sequence profile of regions bordering modified sequons in archaeal glycoproteins*

Amino acid residues near sequons could influence *N*-glycosylation by modulating access to the oligosaccharide transferase active site or the affinity of such interactions, by affecting sequon conformation or by providing sequon residues with partners for hydrogen bonding or other associations that could interfere with *N*-glycosylation. Toward determining whether any of these potential scenarios are applicable to archaeal *N*-glycosylation, the ten amino acids both preceding and following archaeal sequons experimentally verified as glycan-bearing (Table 1) were considered, as were the amino acids surrounding unmodified and proposed, but as yet uncharacterized, sequons in the same proteins. At each amino acid position, the relative frequency of a representative of a particular amino acid group was noted. In this study, amino acids were first clustered based on their being non-polar (AGVLIPF-MWC), uncharged and polar (QNSTY), positively charged (RKH) or negatively charged (DE). Where warranted, the amino acid clusterings adopted by Ben-Dor et al. (2004), designed to further divide residues along chemical property lines, were also considered. In these instances, the groups of the first clustering set were MLIV, TSC, FYW, AG, KR, DE, QN, P and H, whereas the groups comprising the second clustering set were DE, KRH, QNST, AMLIVFWY, C, G and P. In the latter sets of clusterings, the nonpolar and the uncharged, polar amino acids were further subdivided according to their chemical properties (e.g., hydrophobic, aromatic, small in size).

A search of the literature (June, 2006) uncovered 27 experimentally verified archaeal sequons distributed between ten proteins from five species, including halophiles (*Halobacterium salinarum*, *Haloferax volcanii*), methanogens (*Methanothermobacter fervidus*, *Methanococcus voltae*) and a thermoacidophile (*Sulfolobus acidocaldarius*). The evidence for Asn glycosylation in these cases came either from amino acid sequencing (Lechner and Sumper 1987, Lin and Tang 1990, Brockl et al. 1991, Mengele and Sumper 1992) or from mass spectrometry approaches (Paul et al. 1986, Zahringer et al. 2000, Voisin et al. 2005). Examination of the sequons themselves revealed that, as in Eukarya and Bacteria (Gavel and von Heijne 1990, Nita-Lazar et al. 2005), the X position immediately downstream of every glycosylated Asn residue was never a Pro. In one third of the sequons, however, Ser or Thr were found at the X position. By contrast, the same residues account for only 17% of the amino acid content of these glycoproteins, apart from the sequon and bordering 10 residues. When the residues bordering glycosylated sequons were considered, several general traits were observed. In contrast to what has been reported for *C. jejuni*, where the sequon motif was recently expanded to include a necessary Asp or Glu at the -2 position upstream of the sequon Asn residue (Kowarik et al. 2006), negatively charged residues were detected only at the -2 position in the archaeal sequences in one of the 27 exam-

Table 1. Neighboring residues of experimentally verified *N*-glycosylated sequons in Archaea. References: 1 = Paul et al. 1986; 2 = Lechner and Sumper 1987; 3 = Sumper et al. 1990; 4 = Mengele and Sumper 1992; 5 = Brockl et al. 1991; 6. Voisin et al. 2005; 7 = Zahringer et al. 2000; and 8 = Lin and Tang 1990. Abbreviation: SLG = S-layer glycoprotein. *N*-glycosylated sequons are shown in bold.

Species	Protein	No. of <i>N</i> -glycosylation sites <sup>1</sup>	Sequence surrounding verified <i>N</i> -glycosylated sequons	Ref.
<i>H. salinarum</i>	SLG <sup>a</sup>	12	VAFTGGAAAANASDLNDYQRFNE LGIIAKSDAVNSSGGVKDNIDTS	1
<i>H. volcanii</i>	SLG	7	SISDLEDEVGN <sup>17</sup> YTS <sup>19</sup> SGPPTGDQIR	2
			GNLDADESSEFNKTIQSGDRVFLG	3
			IGDTSEVGIANSSATNTSGSSTG	4
			EVGIANSSATNTSGSSTGPTVET PVVETGTVDINGTASGANSVLVI	
<i>M. fervidus</i>	SLG	20	IAGAAEVKNLNTSKTFTLIQEAI	5
	FlaA	2	GKGDMAVITINCTNLDLAPRTRL	6
<i>M. voltae</i>	FlaB1	5	VAAVAASVLINTSGFLQQKAST GLQVTGVTGVNSTDSENITHIVA VTGVNSTDSENITHIVAYITPKA YNGVSAVLKANESAIDGTS <sup>19</sup> GTPD PDVLSATLLSNTSATEYQVVSLQ	6
	FlaB2	6	VAAVAASVLINTSGFLQQKAST LQISQVMGMHNSNINKTAIYIS VMGMHNSNINKTAIYISP <sup>19</sup> NAGS DGSNKR <sup>19</sup> VYKYNESSYKDLTNGGD NKGDLVAITLNTTSFSTTPRTSI ISFTTPATYLNDSKVVQLQ	6
	FlaB3	2	RKMSLLYDSSNASNGGK <sup>19</sup> VRLSTA NGGK <sup>19</sup> VRLSTANGTSDIFKYDDVY	6
	SLG	2	DDANLKSLLTNGTND <sup>19</sup> FTELDAGK	6
	CbsA	16	TPGTTYSLERNY <sup>19</sup> TNYVSIINGKE	7
<i>S. acidocaldarius</i>	Thermopsin	11	PNNTEITVMSNGTILLWHSRPV PAGIASFGLYNYSGNVTPYVITT ASFGLYNYSGNVTPYVIT <sup>19</sup> NEML	8

<sup>1</sup> Both putative and verified *N*-glycosylation sites.

ples considered. By contrast, nonpolar residues were found in this position 48% of the time, whereas uncharged polar residues were detected in 37% of the cases. The remaining examples presented positively charged residues at this position. At the +1 position following the sequon, either Ala or Gly were detected in 37% of the instances, although the same residues account for only 18% of the amino acid content outside the sequon and bordering residues, thus suggesting a preference for a small amino acid at the +1 site. The presence of Pro at the +1 position did not prevent glycosylation, as demonstrated by thermopsin peptide 2, although it should be noted that this is the only example where Pro is detected at this position. Further from the sequon, possible amino acid preferences were noted at the -4 position, where charged residues were detected only once but Ser and Thr were detected 41% of the time, and at the -8, -9 and -10 positions, where nonpolar residues were noted in 81%, 56% and 67% of the sequences, respectively. Specifically, Ala and Gly were observed at position -8 in 48% of the cases and at position -9 in 33% of the cases. Again, since Ser and Thr together and Ala and Gly together account for 17% and 18% of the amino acid content outside the sequon and its up- and downstream flanking 10 residues, respectively, the character of the -4, -8, -9 and -10 positions appears to be an

enrichment in a particular type of residue.

Insight into archaeal *N*-glycosylation can be derived not only from analysis of the environment of modified sequons but also by addressing the surroundings of sequons shown to contain unmodified Asn residues. In the set of sequences examined in the present study, only Asn-17 of the *H. salinarum* S-layer glycoprotein was verified as unmodified despite being part of an NXT sequon (Lechner and Sumper 1987). In contrast to the 27 glycosylated archaeal sequons, where aromatic residues were never observed at the +1 position, the position immediately after <sup>17</sup>NYT<sup>19</sup> is occupied by a Tyr residue. Given the apparent preference of the glycosylation machinery for smaller residues at this site (see above), it is thus possible that the presence of a bulky Tyr at this site hinders glycan attachment. When the amino acids proceeding (NDYQRFNENT) and following (YSTASEDGKT) this *H. salinarum* S-layer glycoprotein sequon were considered, no marked differences from the corresponding residues surrounding modified sequons were evident. Similarly, no obvious patterns were apparent when the amino acid composition of the upstream and downstream regions surrounding the remaining 56 proposed, but as yet unverified, sequons in the same proteins was considered (Table A1). Nonetheless, no Pro residues were detected at

the X position in any of these sequons, whereas at the +1 position, only nine sequences presented an aromatic residue and just one sequon offered a Pro at this site. Thus, it is likely that many of the additional sequons are glycosylated.

In addition to addressing questions of amino acid sequence, examining the distribution of modified sequons may also provide details on the *N*-glycosylation process in Archaea. Such information could offer insight into the ability of the archaeal oligosaccharide transferase to process sequons lying in close proximity to each other, thereby reflecting spatial considerations of the enzyme's activity. In the present study, such analysis revealed that, despite their proximity, both Asn residues within the <sup>308</sup>NSSATNTS<sup>315</sup> section of the *H. volcanii* S-layer glycoprotein are modified (Mengele and Sumper 1992). Similarly, Asn-72 and Asn-77 of *M. voltae* FlaB2, again separated by four residues, are both modified (Voisin et al. 2005). In Eukarya, the closest glycosylated pair of naturally occurring Asn residues, found in haptoglobin-1, are three residues apart (Kurosky et al. 1980, Gavel and von Heijne 1990). At present, no information on the acceptable proximity of Asn for dual glycosylation tolerated by the *C. jejuni* *N*-glycosylation machinery is available.

In terms of the distribution of modified sequons within the glycoprotein, only the *M. voltae* flagellins were considered, since 14 of the 15 putative *N*-glycosylation sites have been characterized (Voisin et al. 2005). In the four proteins considered, i.e., FlaA, FlaB1, FlaB2 and FlaB3, *N*-glycosylation sites were distributed throughout the proteins and showed no preference for any particular secondary structural region.

#### *The environments of occupied archaeal NXS and NXT sequons are distinct*

Several studies have pointed to the differential processing of NXS and NXT sequons in eukaryal glycoproteins (Kasturi et al. 1995, Shakin-Eshleman et al. 1996, Mellquist et al. 1998, Ben-Dor et al. 2004). When the two classes of sequons found in the archaeal glycoprotein sequences considered in the present study were examined, it appeared that the residues surrounding experimentally verified NXS sequons ( $n = 14$ ) revealed traits distinct from the residues surrounding experimentally verified NXT sequons ( $n = 13$ ). As reflected in Figure 1, the two sequon classes could be distinguished at the X position, with Ser or Thr being detected in 50% of the NXS sequons, but in only 15% of the NXT sequons. By contrast, the X position of NXT sequons corresponded to Gly in 31% of the sequences, but in only 14% of the NXS sequons. At the -2 position, Ser or Thr were detected in 46% of the NXT sequons, but in only 7% of the NXS sequons. At the +1 position, NXS sequons incorporated Gly at the X position 36% of the time, whereas Gly was not detected in the same position of the NXT sequons. Distinctions between NXS and NXT were also noted at positions further upstream or downstream of the sequon. Acidic residues were observed in 39% of the -7 positions of those sequences including the NXS sequon, but were absent from the same position in NXT-containing sequences. At the +8, +9 and +10 positions, NXS-containing sequences pre-

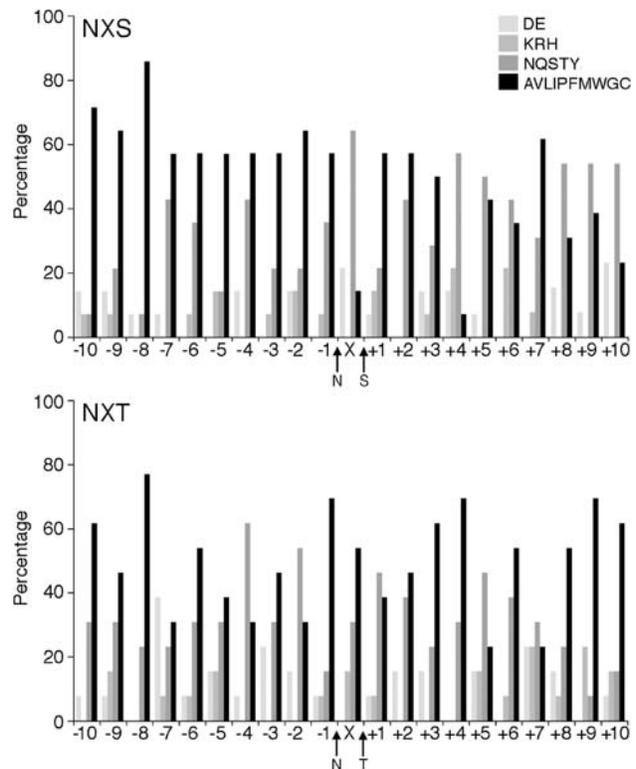


Figure 1. Analysis of the characteristics of amino acids surrounding modified Asn residues in archaeal glycoproteins. The relative proportions of acidic (DE), basic (KRH), polar uncharged (NQSTY) and polar (AVLIPFMWGC) residues were considered at 10 positions up- and downstream of modified archaeal NXS (upper panel, based on 14 sequences) and NXT (lower panel, based on 13 sequences) sequons. In each case, the character of the residue at the X position was also considered. In each panel, the positions of the modified Asn as well as the sequon Ser (upper panel) or Thr (lower panel) are indicated.

sented Ser or Thr 46% of the time, whereas the same residues were only employed at these positions in 8% of the NXT-containing sequences. On the other hand, the hydrophobic residues Ile, Leu, Met and Val were observed in positions +9 and +10 in 46% and 39% of those sequences including NXT, respectively, but only 15% and 8% of the time at the same positions in NXS-containing sequences.

#### *Putative N-glycosylation of predicted secretory proteins in Archaea*

To get a general sense of the extent of *N*-glycosylation in Archaea, a previously assembled database of predicted secretory proteins from 10 archaeal genomes (Bardy et al. 2003) was scanned for the presence of NXS and NXT sequons (where X was not a proline), assuming these to be potential *N*-glycosylation sites (Table A2). Such analysis predicted major differences in the levels of putative *N*-glycosylated secreted proteins expressed by the various genomes examined (Table 2). Whereas only 39% of the 121 proteins supposedly secreted by *A. pernix* are predicted to be *N*-glycosylated, 30 of the 32 predicted *M. jannaschii* secretory proteins (i.e., 94%)

Table 2. Tally of predicted N-glycosylated secretory proteins in selected archaeal species.

Species	No. of predicted secretory proteins	No. of predicted N-glycosylated secretory proteins <sup>1</sup>				No. of sequons per predicted secreted glycoprotein			
		NXS + NXT	NXS alone	NXT alone	Total	1	2–4	5–6	> 6
<i>A. pernix</i>	121	10	23	14	47	26	18	3	0
<i>A. fulgidus</i>	79	28	9	13	50	17	25	3	5
<i>Halobacterium</i> sp. NRC-1	68	10	11	13	34	19	12	3	0
<i>M. jannaschii</i>	32	22	3	5	30	2	13	4	11
<i>M. thermoautotrophicum</i>	68	30	2	13	45	7	17	9	12
<i>P. abyssi</i>	28	13	4	7	24	5	13	3	3
<i>P. horikoshii</i>	63	21	10	11	42	12	18	4	8
<i>S. solfataricus</i>	46	25	6	5	36	3	18	4	11
<i>T. acidophilum</i>	26	17	5	1	23	3	5	5	10
<i>T. volcanium</i>	24	12	2	3	17	5	5	0	7
Total	555	188	75	85	348	99	144	38	67

<sup>1</sup> The motifs NPS (40 instances) and NPT (20 instances) were omitted from the tally.

contain NXS or NXT sequons. The other genomes considered are predicted to N-glycosylate 50–88% of their putatively secreted proteins. Moreover, not only do the different species potentially N-glycosylate their predicted secretomes to differing extents, but they can also often be distinguished by the unbalanced presence of either the NXS or NXT motif. In *A. pernix*, a total of 37 proteins contain either the NXS or NXT sequon alone yet only 10 proteins contain both. In *Halobacterium* sp. NRC-1, 11 proteins contain the NXS sequon, 13 contain the NXT sequon, whereas 10 proteins contain both. By contrast, the NXS and NXT sequons are found in only three and five *M. jannaschii* proteins, respectively, whereas 22 polypeptides present both motifs. In *M. thermoautotrophicum*, most of the 45 predicted N-glycosylated secretory proteins contain comparable amounts of the NXS and NXT sequons. This number includes, however, 13 proteins containing the NXT motif alone and only two sequences that include just the NXS motif. In *A. pernix*, the NXS motif is found in 23 proteins, whereas only 14 proteins present the NXT sequence.

The putative glycoprotein populations in the different species also differed in the number of proposed N-glycosylation sites per protein. In *A. fulgidus*, 42 proteins contained 1–4 putative N-glycosylation sites, whereas only eight proteins contained five or more such sites. In *Halobacterium* sp. NRC-1, 31 proteins are predicted to contain 1–4 N-glycosylation motifs, but only three proteins are thought to contain five or six such sites. No proteins are predicted to contain more than six sequons in this species. Likewise, *A. pernix* is thought to include 44 predicted glycoproteins that contain four or less possible N-glycosylation motifs, but only three proteins presenting more. By contrast, 43% of the putative secreted proteins that may be N-glycosylated in *S. solfataricus* contain more than six sequons. In *M. jannaschii*, only two proposed secretory proteins contain single N-glycosylation sites, whereas 28 sequences present multiple sequons for possible modification.

#### *Comparing sequences bordering modified archaeal, bacterial and eukaryal sequons*

Analysis of experimentally confirmed glycan-modified sequons and their surrounding residues in archaeal glycoprotein, albeit limited in number, reveals that although the archaeal sequences share much in common with their eukaryal and bacterial counterparts, differences from the rules and tendencies governing eukaryal and bacterial sequon glycosylation also exist.

Based on available data, the most obvious difference between sequon glycosylation in Archaea and Bacteria (as exemplified by *C. jejuni*) is the necessity for an acidic residue in the –2 position in the latter but not the former. Analysis of the putative N-glycosylation sites of the proposed secretory protein populations of 10 archaeal genomes failed to detect an enhanced presence of Asp or Glu at this position. Although further characterization of N-glycosylated proteins from other Bacteria will confirm whether such a requirement for negatively charged residues at the –2 position is unique to *C. jejuni*, the independence of archaeal sequon modification from this rule points to the archaeal Stt3, believed to be responsible for the transfer of a lipid-linked glycan moiety to target sequons (Abu-Qarn and Eichler 2006, Chaban et al. 2006), being closer to its eukaryal than to its *C. jejuni* homologue, PglB (Wacker et al. 2002). This relationship was previously suggested on the basis of sequence comparison, even though the archaeal OST appears to be composed of a single Stt3-like component as in Bacteria (Abu-Qarn and Eichler 2006, Chaban et al. 2006). However, the amino acid composition surrounding modified sequons in archaeal and eukaryal glycoproteins also show differences. In eukaryal glycoproteins, 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 (Ben-Dor et al. 2004, Petrescu et al. 2004). By contrast, Phe, Trp and Tyr were detected at archaeal positions -2 and -1 in only 4% and 11% of the archaeal sequences, respectively, Ser or Thr were readily found at the X position and Ala and Gly predominated at the archaeal +1 position. Aromatic residues were never detected at this site. The only time a Tyr was detected at the +1 position was in the only sequon experimentally shown as being unmodified (see above). Nonetheless, as is the case in archaeal glycoproteins, Asp and Glu are also rarely found at the -2 and -1 positions adjacent to modified sequons in eukaryal *N*-glycosylated proteins (Petrescu et al. 2004). Moreover, in both archaeal and eukaryal glycoproteins, basic residues are poorly represented at position +1.

#### *Insights into the archaeal N-glycosylation process from sequence analysis*

The present study not only offers insight into the relationship between the environment in which archaeal sequons lie and their likelihood of being modified but also sheds light onto the interplay between protein translation, protein translocation and *N*-glycosylation in Archaea. Earlier results have shown that the Asn residues of sequons found < 12–14 positions downstream from a transmembrane domain are not modified by the vertebrate oligosaccharide transferase, suggesting the active site of the complex lies 30–40 Å from the ER membrane (Nilsson and von Heijne 1993). If the active site of archaeal Stt3 is similarly positioned, then the modification of the Asn residue 9 positions from the C-terminus of *M. voltae* FlaB2 (Voisin et al. 2005) would require glycosylation at this position to transpire post-translationally, following translocation of the protein across the membrane. However, topological analysis of the yeast and mouse Stt3 proteins (Kim et al. 2005) predicts that the WWDYG motif, thought to be responsible for the oligosaccharide transferase activity of the protein (Yan and Lennarz 2002, Wacker et al. 2002), is located 42 residues from the last transmembrane domain of the polypeptide. By contrast, predictive software (HMMTOP; Tusnady and Simon 2001) puts the same motif only 25 residues from the final transmembrane domain of the *H. volcanii* protein (Abu-Qarn and Eichler 2006). Thus, the spatial dimensions of co-translational/translocational modification by archaeal Stt3 may differ from those employed by its eukaryal homologue. Similar spatial considerations also apply when considering that Asn residues in the *H. salinarum*, *M. fervidus* and *H. volcanii* S-layer glycoproteins found one, six and 12 positions from the N-terminus, respectively, are glycan-modified. These observations suggest that glycosylation occurs following signal peptide cleavage since these residues would otherwise lie too close to the plasma membrane to be accessible to the archaeal Stt3 active site. A similar scenario has been proposed for the processing of eukaryal signal peptide-bearing glycoprotein precursors (Chen et al. 2001).

#### Conclusions

The finding that the characteristics of the amino acids sur-

rounding glycosylated sequons in Archaea differ from those in Bacteria or Eukarya is in accord with earlier observations pointing to unique aspects of the archaeal process. Although NXS sequons are clearly modified in archaeal glycoproteins, exchanging the Ser residue of a particular Asn-Ala-Ser sequon within the *H. salinarum* S-layer glycoprotein sequence with a Val, Leu or Asn did not prevent *N*-glycosylation (Zeitler et al. 1998). Moreover, whereas glycans are linked to Asn residues of eukaryal glycoproteins almost exclusively through *N*-acetylglucosamine, Archaea can use other linking sugars (Lechner and Wieland 1989, Voisin et al. 2005). Together, these observations suggest that the archaeal OST works differently from its eukaryal and bacterial counterparts or that it is less stringent in its target selection criteria. Alternatively, Archaea could express a second version of the OST that functions in a novel manner. To date, no candidates for this putative second OST have been proposed, although *A. fulgidus* is thought to encode two Stt3 proteins (Burda and Aebi 1999) that could be imagined to function distinctly. Accordingly, different catalytic activities have been reported for the two Stt3 homologues detected in vertebrates (Kelleher et al. 2003).

Analysis of the regions bordering the glycan-bearing Asn residues of archaeal glycoproteins may provide insight into various aspects of the *N*-glycosylation process in Archaea. With the recent development of experimental systems for deciphering the steps involved in the archaeal version of this post-translational modification (Abu-Qarn and Eichler 2006, Chaban et al. 2006), the rules governing sequon occupation, including those proposed here, can be tested.

#### Acknowledgments

This work was supported by the Israel Science Foundation–The Charles H. Revson Foundation (Grant 433/03).

#### References

- Abu-Qarn, M. and J. Eichler. 2006. Protein *N*-glycosylation in Archaea: defining *Haloferax volcanii* genes involved in S-layer glycoprotein glycosylation. *Mol. Microbiol.* 61:511–525.
- Bardy, S.L., J. Eichler and K.F. Jarrell. 2003. Archaeal signal peptides—a comparative survey at the genome level. *Protein Sci.* 12: 1833–1843.
- Bause, E. and G. Legler. 1981. The role of the hydroxy amino acid in the triplet sequence Asn-Xaa-Thr(Ser) for the *N*-glycosylation step during glycoprotein biosynthesis. *Biochem. J.* 195:639–644.
- Ben-Dor, S., N. Esterman, E. Rubin and N. Sharon. 2004. Biases and complex patterns in the residues flanking protein *N*-glycosylation sites. *Glycobiology* 14:95–101.
- Brockl, G., M. Behr, S. Fabry, R. Hensel, H. Kaudewitz, E. Biendl and H. König. 1991. Analysis and nucleotide sequence of the genes encoding the surface-layer glycoproteins of the hyperthermophilic methanogens *Methanothermobacter fervidus* and *Methanothermobacter sociabilis*. *Eur J. Biochem.* 199:147–152.
- Burda, P. and M. Aebi. 1999. The dolichol pathway of *N*-linked glycosylation. *Biochim. Biophys. Acta* 1426:239–257.
- Chaban, B., S. Voisin, J. Kelly, S.M. Logan and K.F. Jarrell. 2006. Identification of genes involved in the biosynthesis and attachment of *Methanococcus voltae* *N*-linked glycans: insight into *N*-linked glycosylation pathways in Archaea. *Mol. Microbiol.* 61:259–268.

- Chen, X., C. VanValkenburgh, H. Liang, H. Fang and N. Green. 2001. Signal peptidase and oligosaccharyltransferase interact in a sequential and dependent manner within the endoplasmic reticulum. *J. Biol. Chem.* 276:2411–2416.
- Eichler, J. and M.W. Adams. 2005. Posttranslational protein modification in Archaea. *Microbiol. Mol. Biol. Rev.* 69:393–425.
- Gavel, Y. and G. von Heijne. 1990. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng.* 3:433–442.
- Kaji, H., H. Saito, Y. Yamauchi, T. Shinkawa, M. Taoka, J. Hirabayashi, K. Kasai, N. Takahashi and T. Isobe. 2003. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify *N*-linked glycoproteins. *Nat. Biotechnol.* 21:667–672.
- Kasturi, L., J.R. Eshleman, W.H. Wunner and S.H. Shakin-Eshleman. 1995. The hydroxy amino acid in an Asn-X-Ser/Thr sequon can influence *N*-linked core glycosylation efficiency and the level of expression of a cell surface glycoprotein. *J. Biol. Chem.* 270:14,756–14,761.
- Kelleher, D.J. and R. Gilmore. 2006. An evolving view of the eukaryotic oligosaccharyl transferase. *Glycobiology* 16:47R-62R.
- Kelleher, D.J., D. Karaoglu, E.C. Mandon and R. Gilmore. 2003. Oligosaccharyltransferase isoforms that contain different catalytic STT3 subunits have distinct enzymatic properties. *Mol. Cell* 12:101–111.
- Kim, H., G. von Heijne and I.M. Nilsson. 2005. Membrane topology of the STT3 subunit of the oligosaccharyl transferase complex. *J. Biol. Chem.* 280:20,261–20,267.
- Kowarik, M., N.M. Young, S. Numao et al. 2006. Definition of the bacterial *N*-glycosylation site consensus sequence. *EMBO J.* 25:1957–1966.
- Kurosky, A., D.R. Barnett, T.H. Lee, B. Touchstone, R.E. Hay, M.S. Arnott, B.H. Bowman and W.M. Fitch. 1980. Covalent structure of human haptoglobin: a serine protease homolog. *Proc. Natl. Acad. Sci. USA* 77:3388–3392.
- Lechner, J. and M. Sumper. 1987. The primary structure of a prokaryotic glycoprotein. Cloning and sequencing of the cell surface glycoprotein gene of halobacteria. *J. Biol. Chem.* 262:9724–9729.
- Lechner, J. and F. Wieland. 1989. Structure and biosynthesis of prokaryotic glycoproteins. *Annu. Rev. Biochem.* 58:173–194.
- Lin, X. and J. Tang. 1990. Purification, characterization, and gene cloning of thermopsin, a thermostable acid protease from *Sulfolobus acidocaldarius*. *J. Biol. Chem.* 265:1490–1495.
- Mellquist, J.L., L. Kasturi, S.L. Spitalnik and S.H. Shakin-Eshleman. 1998. The amino acid following an asn-X-Ser/Thr sequon is an important determinant of *N*-linked core glycosylation efficiency. *Biochemistry* 37:6833–6837.
- Mengele, R. and M. Sumper. 1992. Drastic differences in glycosylation of related S-layer glycoproteins from moderate and extreme halophiles. *J. Biol. Chem.* 267:8182–8185.
- Mescher, M.F. and J.L. Strominger. 1976. Purification and characterization of a prokaryotic glycoprotein from the cell envelope of *Halobacterium salinarium*. *J. Biol. Chem.* 251:2005–2014.
- Nakai, K. and P. Horton. 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* 24:34–36.
- Nilsson, I.M. and G. von Heijne. 1993. Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J. Biol. Chem.* 268:5798–5801.
- Nita-Lazar, M., M. Wacker, B. Schegg, S. Amber and M. Aebi. 2005. The N-X-S/T consensus sequence is required but not sufficient for bacterial *N*-linked protein glycosylation. *Glycobiology* 15:361–367.
- Paul, G., F. Lottspeich and F. Wieland. 1986. Asparaginyln-*N*-acetylgalactosamine. Linkage unit of halobacterial glycosaminoglycan. *J. Biol. Chem.* 261:1020–1024.
- Petrescu, A.J., A.L. Milac, S.M. Petrescu, R.A. Dwek and M.R. Wormald. 2004. Statistical analysis of the protein environment of *N*-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* 14:103–114.
- Shakin-Eshleman, S.H., S.L. Spitalnik and L. Kasturi. 1996. The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of *N*-linked core-glycosylation efficiency. *J. Biol. Chem.* 271:6363–6366.
- Szymanski, C.M. and B.W. Wren. 2005. Protein glycosylation in bacterial mucosal pathogens. *Nat. Rev. Microbiol.* 3:225–237.
- Szymanski, C.M., R. Yao, C.P. Ewing, T.J. Trust and P. Guerry. 1999. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol. Microbiol.* 32:1022–1030.
- Tusnady, G.E. and I. Simon. 2001. The HMMTOP transmembrane topology prediction server. *Bioinformatics* 17:849–850.
- Voisin, S., R.S. Houlston, J. Kelly, J.R. Brisson, D. Watson, S.L. Bardy, K.F. Jarrell and S.M. Logan. 2005. Identification and characterization of the unique *N*-linked glycan common to the flagellins and S-layer glycoprotein of *Methanococcus voltae*. *J. Biol. Chem.* 280:16,586–16,593.
- Wacker, M., D. Linton, P.G. Hitchen et al. 2002. *N*-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* 298:1790–1793.
- Weerapana, E. and B. Imperiali. 2006. Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* 16:91R-101R.
- Yan, Q. and W.J. Lennarz. 2002. Studies on the function of oligosaccharyl transferase subunits. Stt3p is directly involved in the glycosylation process. *J. Biol. Chem.* 277:47,692–47,700.
- Zahringer, U., H. Moll, T. Hettmann, Y.A. Knirel and G. Schafer. 2000. Cytochrome b558/566 from the archaeon *Sulfolobus acidocaldarius* has a unique Asn-linked highly branched hexasaccharide chain containing 6-sulfoquinovose. *Eur. J. Biochem.* 267:4144–4149.
- Zeitler R., E. Hochmuth, R. Deutzmann and M. Sumper. 1998. Exchange of Ser-4 for Val, Leu or Asn in the sequon Asn-Ala-Ser does not prevent *N*-glycosylation of the cell surface glycoprotein from *Halobacterium halobium*. *Glycobiology* 8:1157–1164.

## Appendix

Table A1. Neighboring residues of putative sequons in archaeal glycoproteins containing experimentally verified N-glycosylated sequons. References: 1 = Lechner and Sumper 1987; 2 = Sumper et al. 1990; 3 = Brockl et al. 1991; 4 = Voisin et al. 2005; 5 = Zahringer et al. 2000; and 6 = Lin and Tang 1990. Abbreviation: SLG = S-layer glycoprotein. Putative sequons are shown in bold.

Species	Protein	Sequence surrounding putative sequons	Ref.
<i>H. salinarum</i>	SLG	DTVDTGLVVD <b>ND</b> STTEIVDDYEN STTEIVDDYENT <b>SIS</b> DVDYAYAI DYVNSNEEL <b>TNG</b> SALDGVSTDDD AYVVGSEVD <b>ING</b> TANEGTDDVVL LLQGDASIE <b>IN</b> STDEWNSDGQWS DVPLSNVEPG <b>NY</b> TVEADDGDNTD DQTTVDQPEN <b>Q</b> TMTTTMTETTT ETTT <b>EM</b> TT <b>Q</b> ENT <b>T</b> ENGSEGTSDG MTTT <b>Q</b> ENT <b>T</b> ENGSEGTSDGESGG DVD <b>S</b> G <b>S</b> AT <b>P</b> N <b>V</b> TLLAPRITDSE	1
<i>H. volcanii</i>	SLG	QYVVGSEVD <b>ING</b> TATSSDSVAIY EREDTTTSSD <b>N</b> ATDTTTTTDGPT	2
<i>M. fervidus</i>	SLG	DGN <b>I</b> IV <b>G</b> PG <b>NY</b> TENILVNKSLT PG <b>NY</b> TENILVN <b>K</b> SLTLLK <b>S</b> NGSA SLTLK <b>S</b> TLK <b>S</b> NG <b>S</b> AIINAVSSEK IGGK <b>NG</b> I <b>M</b> EN <b>V</b> TGCTITNNTIQ MEN <b>V</b> TGCTITNNTIQNAFVSGWE YYGG <b>NG</b> ICLV <b>N</b> STNNTITN <b>N</b> IIR G <b>NG</b> ICLV <b>N</b> STNNTITN <b>N</b> IIRNNT NNTITN <b>N</b> IIRNNTW <b>NG</b> IN <b>V</b> CE <b>S</b> K V <b>CE</b> SK <b>GN</b> I <b>K</b> NNTIMYSGGIGIY N <b>NI</b> EN <b>N</b> R <b>I</b> IN <b>A</b> TYGGIYLF <b>R</b> PS N <b>N</b> K <b>I</b> CR <b>NY</b> IAN <b>V</b> SSGGGG <b>M</b> SGAI PK <b>G</b> ELVD <b>N</b> I <b>W</b> NT <b>T</b> EGG <b>NY</b> WS <b>NY</b> T NT <b>T</b> EGG <b>NY</b> WS <b>NY</b> TG <b>N</b> NTGD <b>G</b> T <b>G</b> N GG <b>NY</b> WS <b>NY</b> TG <b>N</b> NTGD <b>G</b> T <b>G</b> NIP <b>Y</b> Y VR <b>W</b> DK <b>I</b> V <b>P</b> EG <b>N</b> H <b>T</b> IK <b>A</b> EIP <b>S</b> AE FS <b>K</b> IV <b>Q</b> GF <b>V</b> Q <b>N</b> K <b>T</b> FTITLT <b>N</b> L <b>G</b> K YY <b>I</b> SI <b>Y</b> T <b>N</b> P <b>V</b> NG <b>T</b> K <b>V</b> S <b>Y</b> RE <b>L</b> T <b>I</b> T Y <b>R</b> E <b>L</b> T <b>I</b> T <b>L</b> K <b>P</b> NE <b>T</b> K <b>I</b> E <b>L</b> G <b>K</b> Y <b>P</b> F VA <b>A</b> V <b>A</b> AS <b>V</b> L <b>I</b> NT <b>S</b> G <b>F</b> L <b>Q</b> Q <b>K</b> AS <b>T</b> AD <b>S</b> D <b>Y</b> SD <b>A</b> L <b>I</b> N <b>A</b> TT <b>G</b> F <b>A</b> N <b>I</b> AD <b>N</b> V	3
<i>M. voltae</i>	FlaA	VA <b>A</b> V <b>A</b> AS <b>V</b> L <b>I</b> NT <b>S</b> G <b>F</b> L <b>Q</b> Q <b>K</b> AS <b>T</b>	4
	SLG	AD <b>S</b> D <b>Y</b> SD <b>A</b> L <b>I</b> N <b>A</b> TT <b>G</b> F <b>A</b> N <b>I</b> AD <b>N</b> V	4
<i>S. acidocaldarius</i>	CbsA	FL <b>S</b> SI <b>F</b> T <b>L</b> EN <b>V</b> SL <b>A</b> Q <b>T</b> SP <b>Q</b> IS <b>V</b> AG <b>Y</b> WS <b>Q</b> IP <b>W</b> T <b>N</b> IS <b>L</b> TAN <b>I</b> PM <b>A</b> PT TH <b>Y</b> LL <b>V</b> KA <b>A</b> W <b>NG</b> SW <b>I</b> F <b>I</b> EE <b>W</b> QA W <b>S</b> T <b>A</b> V <b>A</b> GI <b>Y</b> P <b>N</b> AS <b>G</b> P <b>L</b> FR <b>M</b> IE <b>L</b> G <b>K</b> E <b>E</b> T <b>G</b> R <b>I</b> V <b>F</b> NY <b>S</b> G <b>I</b> T <b>L</b> P <b>A</b> P <b>N</b> NT N <b>Y</b> SG <b>I</b> T <b>L</b> P <b>A</b> P <b>N</b> TE <b>I</b> T <b>V</b> M <b>S</b> NG <b>T</b> I ND <b>G</b> M <b>F</b> Y <b>G</b> Y <b>V</b> N <b>S</b> T <b>W</b> Y <b>Y</b> PD <b>R</b> A <b>A</b> IM I <b>W</b> M <b>W</b> V <b>S</b> G <b>A</b> T <b>W</b> N <b>N</b> ST <b>Y</b> DP <b>A</b> FK <b>S</b> N <b>I</b> W <b>M</b> W <b>V</b> S <b>G</b> A <b>T</b> W <b>N</b> N <b>S</b> T <b>Y</b> DP <b>A</b> FK <b>S</b> N <b>I</b> W D <b>P</b> A <b>F</b> K <b>S</b> N <b>I</b> W <b>Q</b> NE <b>S</b> L <b>T</b> G <b>L</b> S <b>Y</b> VD <b>S</b> G N <b>H</b> G <b>F</b> AV <b>P</b> LY <b>T</b> N <b>N</b> T <b>M</b> Y <b>E</b> VD <b>T</b> AG <b>I</b> W W <b>Y</b> TP <b>V</b> ASE <b>G</b> L <b>NG</b> SL <b>F</b> FI <b>W</b> T <b>G</b> A <b>K</b> Y F <b>F</b> I <b>W</b> T <b>G</b> A <b>K</b> Y <b>E</b> NG <b>S</b> W <b>V</b> VE <b>F</b> AR <b>P</b> LS V <b>P</b> LD <b>Y</b> Q <b>P</b> F <b>M</b> P <b>N</b> IT <b>V</b> G <b>K</b> T <b>Y</b> Y <b>V</b> A <b>F</b> A P <b>Q</b> NI <b>Y</b> FF <b>P</b> HR <b>N</b> T <b>T</b> G <b>A</b> T <b>I</b> SS <b>G</b> LY <b>V</b> I <b>T</b> T <b>N</b> E <b>M</b> L <b>G</b> Y <b>V</b> N <b>I</b> T <b>S</b> L <b>L</b> A <b>Y</b> N <b>R</b> E <b>A</b> L T <b>L</b> Q <b>F</b> N <b>I</b> V <b>L</b> S <b>V</b> N <b>T</b> S <b>NG</b> V <b>Y</b> AY <b>W</b> L <b>Q</b> D N <b>S</b> L <b>T</b> F <b>I</b> D <b>N</b> V <b>W</b> N <b>L</b> T <b>G</b> S <b>L</b> ST <b>L</b> SS <b>S</b> A S <b>F</b> PL <b>S</b> Y <b>I</b> T <b>I</b> N <b>M</b> S <b>Y</b> T <b>S</b> NA <b>V</b> Y <b>V</b> W <b>I</b> I <b>S</b> AS <b>L</b> M <b>I</b> NG <b>N</b> NY <b>T</b> P <b>NG</b> LY <b>Y</b> DA <b>E</b> L N <b>G</b> A <b>P</b> T <b>S</b> F <b>N</b> SL <b>N</b> CT <b>L</b> G <b>L</b> Y <b>I</b> S <b>NG</b> S N <b>C</b> T <b>L</b> G <b>L</b> Y <b>I</b> S <b>NG</b> S <b>I</b> TP <b>V</b> PS <b>L</b> Y <b>T</b> F G <b>V</b> PI <b>A</b> Y <b>NG</b> I <b>E</b> N <b>L</b> T <b>I</b> L <b>T</b> N <b>N</b> F <b>S</b> V <b>I</b> L G <b>I</b> EN <b>L</b> T <b>I</b> L <b>T</b> N <b>N</b> F <b>S</b> V <b>I</b> L	5
	Thermopsin	P <b>Q</b> NI <b>Y</b> FF <b>P</b> HR <b>N</b> T <b>T</b> G <b>A</b> T <b>I</b> SS <b>G</b> LY <b>V</b> I <b>T</b> T <b>N</b> E <b>M</b> L <b>G</b> Y <b>V</b> N <b>I</b> T <b>S</b> L <b>L</b> A <b>Y</b> N <b>R</b> E <b>A</b> L T <b>L</b> Q <b>F</b> N <b>I</b> V <b>L</b> S <b>V</b> N <b>T</b> S <b>NG</b> V <b>Y</b> AY <b>W</b> L <b>Q</b> D N <b>S</b> L <b>T</b> F <b>I</b> D <b>N</b> V <b>W</b> N <b>L</b> T <b>G</b> S <b>L</b> ST <b>L</b> SS <b>S</b> A S <b>F</b> PL <b>S</b> Y <b>I</b> T <b>I</b> N <b>M</b> S <b>Y</b> T <b>S</b> NA <b>V</b> Y <b>V</b> W <b>I</b> I <b>S</b> AS <b>L</b> M <b>I</b> NG <b>N</b> NY <b>T</b> P <b>NG</b> LY <b>Y</b> DA <b>E</b> L N <b>G</b> A <b>P</b> T <b>S</b> F <b>N</b> SL <b>N</b> CT <b>L</b> G <b>L</b> Y <b>I</b> S <b>NG</b> S N <b>C</b> T <b>L</b> G <b>L</b> Y <b>I</b> S <b>NG</b> S <b>I</b> TP <b>V</b> PS <b>L</b> Y <b>T</b> F G <b>V</b> PI <b>A</b> Y <b>NG</b> I <b>E</b> N <b>L</b> T <b>I</b> L <b>T</b> N <b>N</b> F <b>S</b> V <b>I</b> L G <b>I</b> EN <b>L</b> T <b>I</b> L <b>T</b> N <b>N</b> F <b>S</b> V <b>I</b> L	6

Table A2. Putative *N*-glycosylated predicted secretory proteins.

Species	Proteins
<i>A. fulgidus</i>	AF0078, 0101, 0141, 0142, 0254, 0274, 0378, 0487, 0637, 0663, 0681, 0711, 0736, 0740, 0759, 0816, 0817, 0912, 0924, 0992, 1025, 1054, 1055, 1096, 1118, 1177, 1231, 1276, 1354, 1411, 1464, 1574, 1658, 1659, 1660, 1661, 1675, 1763, 1884, 1945, 1948, 2017, 2048, 2079, 2080, 2081, 2090, 2119, 2163, 2188
<i>A. pernix</i>	APE0016, 0018, 0174, 0268, 0419, 0592, 0595, 0620, 0622, 0654, 0725, 0814, 0825, 0858, 0876, 0916, 0958, 1013, 1143, 1171, 1303, 1334, 1458, 1463, 1521, 1574, 1575, 1604, 1625, 1675, 1676, 1820, 1901, 1905, 1944, 2014, 2073, 2095, 2175, 2186, 2204, 2422, 2434, 2545, 2594, 2610,
<i>Halobacterium</i> sp. NRC-1	VNG0200, 0204, 0219, 0220, 0262, 0281, 0399, 0539, 0620, 0681, 0816, 0818, 0819, 0931, 0953, 1009, 1036, 1100, 1181, 1337, 1372, 1659, 1675, 1951, 2148, 2516, 2567, 2573, 2589, 6064, 6178, 6323
<i>M. jannaschii</i>	MJ0418, 0492, 0505, 0753, 0755, 0756, 0784, 0815, 0822, 0832.1, 0833, 0835.1, 0872, 0891, 0893, 0903, 0954, 1128, 1284, 1289, 1291, 1292, 1396, 1400, 1402, 1472, 1488, 1502
<i>M. thermoautotrophicus</i>	MTH57, 60, 61, 97, 117, 247, 302, 318, 330, 353, 354, 412, 424, 479, 508, 513, 523, 567, 570, 596, 597, 604, 612, 635, 670, 715, 719, 795, 806, 845, 910, 911, 924, 968, 977, 995, 1038, 1216, 1290, 1330, 1400, 1513, 1566, 1653, 1772, 1777, 1782
<i>P. abyssi</i>	PAB0061, 0133, 0275, 0337, 0378, 0485, 0602, 0741, 0847, 0936, 0982, 0984, 1279, 1379, 1380, 1384, 1456, 1651, 1687, 1790, 1823, 1825, 2366
<i>P. horikoshii</i>	PH004, 0036, 0066, 0188, 0244, 0298, 0310, 0399, 0462, 0480, 0548, 0549, 0550, 0551, 0555, 0564, 0572, 0602, 0649, 0733, 0777, 0875, 0954, 0996, 1070, 1130, 1184, 1190, 1191, 1192, 1289, 1303, 1336, 1341, 1495, 1569, 1668, 1690, 1740, 1790, 1929, 1964
<i>S. solfararius</i>	SSO0037, 0117, 0147, 0330, 0483, 0489, 0519, 0522, 0541, 0650, 0999, 1053, 1093, 1131, 1167, 1320, 1872, 1981, 2088, 2187, 2420, 2570, 2611, 2681, 2684, 2712, 2776, 2795, 2819, 2846, 2969, 2971, 2972, 3104, 3138, 3177
<i>T. acidophilum</i>	Ta0129, 0261, 0297, 0308, 0367, 0380, 0496, 0510, 0553, 0556, 0591, 0721, 0741, 0851, 0951, 1029, 1052, 1171, 1236, 1403, 1421, 1462, 1499
<i>T. volcanium</i>	TVN0085, 0128, 0205, 0231, 0366, 0371, 0544, 0610, 0611, 0711, 0840, 0948, 0969, 1143, 1200, 1290, 1452, 1469