Glycoproteomics of *Haloferax volcanii* reveals an extensive glycoproteome and concurrence of different *N*-glycosylation pathways

Stefan Schulze[1], Friedhelm Pfeiffer[2], Benjamin A. Garcia[3] and Mechthild Pohlschroder[1*]

[1] Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

[2] Computational Biology Group, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

[3] Epigenetics Institute, Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

* corresponding author: Mechthild Pohlschroder, pohlschr@sas.upenn.edu

Abstract

Glycosylation is one of the most complex post-translational protein modifications. Its importance has been established not only for eukaryotes but also for a variety of prokaryotic cellular processes, such as biofilm formation, motility and mating. However, comprehensive glycoproteomic analyses are largely missing in prokaryotes. Here we extend the phenotypic characterisation of *N*-glycosylation pathway mutants in *Haloferax volcanii* and provide a detailed glycoproteome for this model archaeon through the mass spectrometric analysis of intact glycopeptides. Using in-depth glycoproteomic datasets generated for the wild-type and mutant strains as well as a reanalysis of datasets within the Archaeal Proteome Project, we identify the largest archaeal glycoproteome described so far. We further show that different *N*-glycosylation pathways can modify the same glycosites under the same culture conditions. The extent and complexity of the *Hfx. volcanii N*-glycoproteome revealed here provides new insights into the roles of *N*-glycosylation in archaeal cell biology.

Introduction

Post-translational modifications (PTMs) increase the complexity of the proteome and may affect functional activity, localization, and interactions of proteins. Among the plethora of PTMs, glycosylation is one of the most complex, and is encountered in all three domains of life, eukaryotes, bacteria, and archaea. The wide distribution and importance of glycosylation in eukaryotes, including its crucial roles in human diseases, has been well-established over the past decades (Aebi, 2013; Apweiler et al., 1999; Freeze, 2013; Varki, 2017). However, much less is known about prokaryotic protein glycosylation, as protein glycosylation was long thought to be specific to eukaryotes. Prokaryotic protein glycosylation was first detected in the S-layer glycoprotein (SLG) of the archaeon *Halobacterium salinarum* (Mescher and Strominger, 1976) and soon thereafter also in *Haloferax volcanii* (Sumper et al., 1990). From there on, more and more prokaryotes, including bacteria, were found to carry glycan modifications of proteins. By now it is well established that protein glycosylation is common in prokaryotes and its importance in a variety of biomedically and biotechnologically relevant processes has been revealed such as biofilm formation, pathogenicity, viral protection, and mating (Harvey et al., 2018; Nothaft and Szymanski, 2010; Schäffer and Messner, 2017; Valguarnera et al., 2016). However, besides a few abundant surface proteins and appendages, little attention has been paid to which proteins are glycosylated and under which conditions (Choudhary et al., 2019; Eichler et al., 2013; Halim and Anonsen, 2017).

In archaea, glycosylation of SLG, type IV pili and archaella has been confirmed for various species, including non-haloarchaea like *Methanococcus maripaludis* and *Sulfolobus acidocaldarius* (Guan et al., 2016; Jarrell et al., 2014; Kelly et al., 2020; Meyer and Albers, 2013). Efforts to identify additional archaeal glycoproteins are limited to the characterization of purified proteins (Vinogradov et al., 2012), the lectin affinity based enrichment of glycoproteins (Palmieri et al., 2013) and periodic acid-Schiff staining of proteins (as summarized in the ProGlycProt database (Choudhary et al., 2019)). Comprehensive glycoproteomic analyses that permit the identification of intact glycopeptides on a proteome-wide scale are missing so far.

Two main types of glycosylation have been encountered in archaea so far: *O*- and *N*-glycosylation. For *O*-glycosylation, which is characterized by the linkage of a glycan to the hydroxyl oxygen of Ser or Thr, only few proteins with short *O*-glycans (mono- and disaccharides) have been identified (Mescher and Strominger, 1976; Sumper et al., 1990) and the archaeal *O*-glycosylation pathway is still elusive. In contrast, *N*-glycosylation pathways, linking glycans to the amide nitrogen of Asn, have been studied in considerable detail in several species and a plethora of *N*-glycan structures of varying length has been identified (Jarrell et al., 2014; Meyer and Albers, 2013). Glycan structures are highly variable, even between closely related species (Shalev et al., 2018), and this also holds true for the biosynthetic enzymes that are responsible for the step by step assembly of the glycan on its lipid carrier. The assembled lipid-linked oligosaccharide is then transferred onto the protein by two general steps: the transfer of the glycan to the extracellular side by a flippase and the attachment to the N-X-S/T sequon of target proteins by an oligosaccharyltransferase (OST).

In *Hfx. volcanii*, two distinct *N*-glycosylation pathways have been described so far. One leads to the attachment of a Glc-GlcA-GalA-MeGlcA tetrasaccharide (with Glc, glucose; GlcA, glucuronic acid; GalA, galacturonic acid; MeGlcA, methyl-glucuronic acid) which can be further elongated by a Man (Mannose) residue after transfer onto the protein (Jarrell et al., 2014; Kandiba et al., 2016). This pathway can be disturbed by deleting *aglB*, the gene coding for the OST. The second pathway results in the transfer of a SHex-Hex-Hex-Rha tetrasaccharide (with SHex, sulfated hexose; Hex, hexose; Rha, rhamnose) and can be disrupted by the deletion of *agl15*, which encodes a flippase (Eichler et al., 2013). The OST corresponding to this pathway remains to be

identified but transfer of the glycan is independent of AglB. Notably, Agl15-dependent glycans have only been found to modify SLG and only under low-salt growth conditions (Eichler et al., 2013; Guan et al., 2012). The enzymes involved in the biosynthesis of this glycan exist, however, under normal growth conditions as revealed by their proteomic identification (Schulze et al., 2020). Finally, a third *N*-glycan of a more complex type, containing *N*-acetylglucosamine and repeating units of MeSQ-Hex (with MeSQ, methyl-sulfoquinovose) has been found attached to the SLG but the corresponding biosynthesis pathway awaits identification (Parente et al., 2014).

A variety of phenotypes have been described for mutants of the AglB-dependent glycosylation pathway. *Hfx. volcanii ΔaglB* mutants exhibit a slight growth defect and were observed to be more prone to shedding of the SLG (Abu-Qarn et al., 2007; Tamir and Eichler, 2017). They also appear to be unable to synthesize stable archaealla, a phenotype not dependent on SLG instability as specific mutations of glycosites within the archaellins render the cells non-motile and devoid of detectable archaella (Tripepi et al., 2012). Lack of AglB-dependent glycosylation was also shown to result in aggregation of type IV pili and early induction of microcolony formation (Esquivel et al., 2016). Moreover, a recent study revealed a decreased mating efficiency for *ΔaglB* that was further decreased in a *ΔaglB/Δagl15* double deletion strain (Shalev et al., 2017). While mating of *Hfx. volcanii* has been shown to be independent of type IV pili and archaella (Tripepi et al., 2010), and *N*-glycosylation of the SLG has been suggested to be involved (Shalev et al., 2017), little is known about the many other predicted glycoproteins on the surface of *Hfx. volcanii* that might play key roles in mating and other critical surface associated functions.

Here we extended the phenotypic analysis of the *ΔaglB* strain and the less well characterized *Δagl15* mutant and carried out a systematic glycoproteomics study to identify as complete of a *Hfx. volcanii* glycoproteome as possible. These studies identified the largest number of glycoproteins yet identified in any archaeon, and also revealed that proteins can harbor the so-called low-salt, Agl15-dependent *N*-glycan under normal salt conditions. Furthermore, AglB- and Agl15-dependent *N*-glycosylation occurred not only under the same conditions, they were also found to be able to modify the same *N*-glycosylation sites.

Results

Cell biological assays reveal phenotypes for **ΔaglB** and **Δagl15** strains under normal salt conditions

In order to further characterize the biological implications of an *aglB* deletion, but especially to gain more insight into the phenotypic effects of an *agl15* deletion, we analyzed the colony morphology and color, the motility, growth, and shape of both *ΔaglB* and *Δagl15* strains (Figure 1). Colonies that lack Agl15 were darker and smaller than wildtype colonies (WT, H53), while *ΔaglB* colonies, which were smaller as well, showed the same color as the WT (Figure 1A). Consistent with previous studies, *ΔaglB* mutants are non-motile (Tripepi et al., 2012). In contrast, *Δagl15* mutants exhibit the same motility as the WT (Figure 1B).

A growth curve revealed that the *Δagl15* strain grew to a slightly higher optical density at 600 nm (OD_{600}) than the WT strain. The *ΔaglB* strain exhibited a slightly slower growth than the WT and *Δagl15* strains during mid-logarithmic phase, but reached the same OD_{600} at stationary phase as the WT (Figure 1C). *Hfx. volcanii* has been described to change from rod-shaped cells in early-logarithmic growth phase to disk-shaped cells in the late-logarithmic growth phase (Duggin et al., 2015; Li et al., 2019; Silva et al., 2020). While we observed the same trend for all strains analyzed here, interestingly, the cell shape of both *N*-glycosylation pathway mutants differed from that of the WT strain at different stages of the growth curve (Figure 1C). In comparison to the WT, the *Δagl15* mutant exhibited a higher ratio of rod- to disk-shaped cells in the mid-logarithmic growth phase and even still retained a few rod-shaped cells in the late-logarithmic growth phase. In contrast, a higher percentage of disk-shaped cells was formed by the *ΔaglB* mutant compared to the WT in the early-logarithmic growth phase, and during mid-logarithmic growth, exclusively disk-shaped cells could be found in *ΔaglB* cultures.

In summary, these results revealed multiple biological phenotypes for *Hfx. volcanii* strains lacking AglB or Agl15-dependent *N*-glycosylation indicating that both pathways are involved in a variety of cellular processes. Since all assays were performed using media with optimal salt concentrations (18% salt water corresponding to 2.7 M salts) this suggests that, consistent with the ArcPP data, Agl15-dependent glycosylation is not restricted to low salt conditions. Importantly, the phenotypes of both deletion mutants were distinct from each other, indicating differing roles in the same cellular processes.

The diversity and multiplicity of the biological effects caused by the interference with *N*-glycosylation pathways may involve additional *N*-glycoproteins besides the well-characterized SLG, archaellins, and pilins. Thus, in the following we aimed to provide a more complete picture about the *Hfx. volcanii* glycoproteome.

Glycoprotein staining reveals minimal differences between the WT, **ΔaglB** and **Δagl15** strains

In order to determine which *Hfx. volcanii* proteins can be glycosylated, we devised a comprehensive glycoproteomic analysis of samples taken under salt conditions optimal for the growth of this haloarchaeon. This analysis was performed for the WT, *Δagl15* and *ΔaglB* strains, thus including negative controls for each described *N*-glycosylation pathway, allowing to control for false positive identifications. Cells were fractionated into cytosol (Cyt), membrane (Mem) and culture supernatant (SN) to achieve a high proteome coverage by reducing the sample complexity, and to obtain information about the cellular localization of the identified glycoproteins. Separation of these fractions by LDS-PAGE followed by Coomassie staining demonstrated that for the *ΔaglB* mutant a larger portion of the SLG was shed into the SN than for the WT and *Δagl15* strains (Figure 2A). This is in line with previous reports indicating a reduced S-layer stability on the cells lacking AglB (Abu-Qarn et al., 2007; Tamir and Eichler, 2017). Besides this, a similar banding pattern was observed for all strains. Similarly, a glycoprotein stain showed no apparent differences between the strains either (Figure 2B). It may be surprising that the disruption of an *N*-glycosylation pathway does not affect the banding pattern of glycoproteins. This observation is, however, in line with previous reports using the conceptually similar periodic acid-Schiff staining, which showed only a small decrease in signal intensity for SLG from the *ΔaglB* mutant (Shalev et al., 2017). Similar signal intensities for the Pro-Q Emerald 300 glycoprotein staining employed here may indicate the presence of Agl15-dependent glycosylation in the *ΔaglB* mutant and vice versa. Alternatively, since proteins from the Cyt fraction were stained as well, widespread *O*-glycosylation could mask differences in *N*-glycosylation patterns. Finally, unspecific staining of non-glycosylated proteins cannot be completely excluded.

2
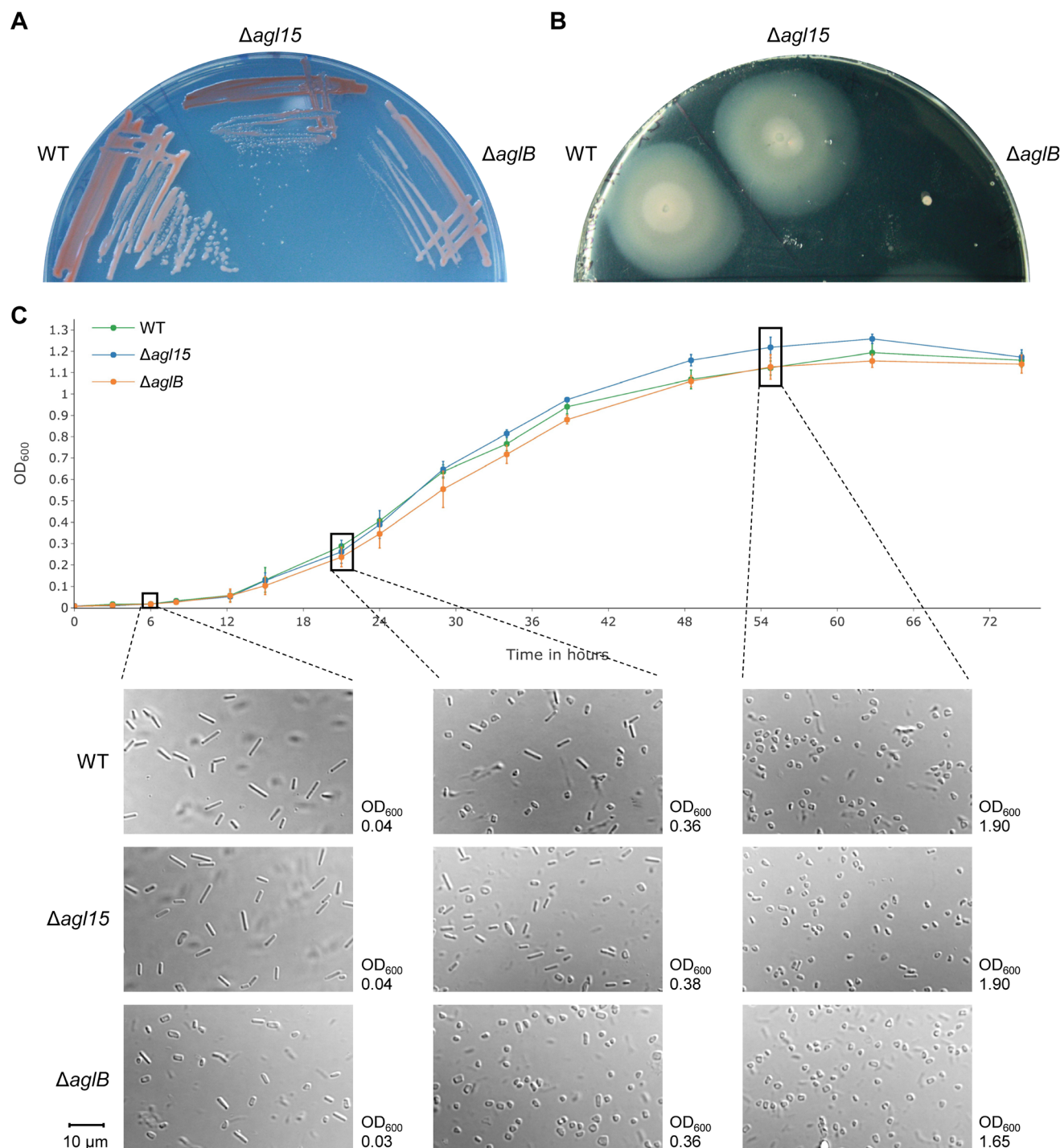
Figure 1. Colony morphology, motility, growth and cell shape phenotypes of *N*-glycosylation pathway mutants. A, WT, Δ*agl15* and Δ*aglB* strains were streaked out on Hv-Cab plates with 1.5% agar. Colonies of Δ*agl15* and Δ*aglB* mutants appear smaller, and for the Δ*agl15* mutant also darker, than WT colonies. **B**, WT, Δ*agl15* and Δ*aglB* strains were stabbed into motility plates (Hv-Cab with 0.35% agar) and imaged after 5 days of incubation. While the WT and Δ*agl15* strains show normal motility, the Δ*aglB* mutant is non-motile. **C**, The growth of WT (green), Δ*agl15* (blue) and Δ*aglB* (orange) strains was analyzed by measuring the OD_{600} of cultures over the course of three days (top). Data points and error bars represent the mean and standard deviation of three biological replicates. Samples of early- (bottom left), mid- (bottom middle) and late-logarithmic (bottom right) growth phase cultures for each strain were imaged using DIC microscopy. The majority of WT cells in the early-logarithmic growth phase is rod-shaped and the ratio of rod- to disk-shaped cells decreases over time. A higher ratio of rod to disk shaped cells was observed in mid- and late-logarithmic growth phase for the Δ*agl15* mutant in comparison to the WT while almost no rod shaped cells were visible for the Δ*aglB* strain. It should be noted that OD_{600} measurements for cell shape samples were performed with a path length of 1.5 cm while growth curve measurements were performed in 96-well plates with 250 µl of culture. Images in A, B, and C (bottom) are representative for at least four biological replicates. The scale bar (bottom left) indicates 10 µm and applies to all DIC microscopy images.
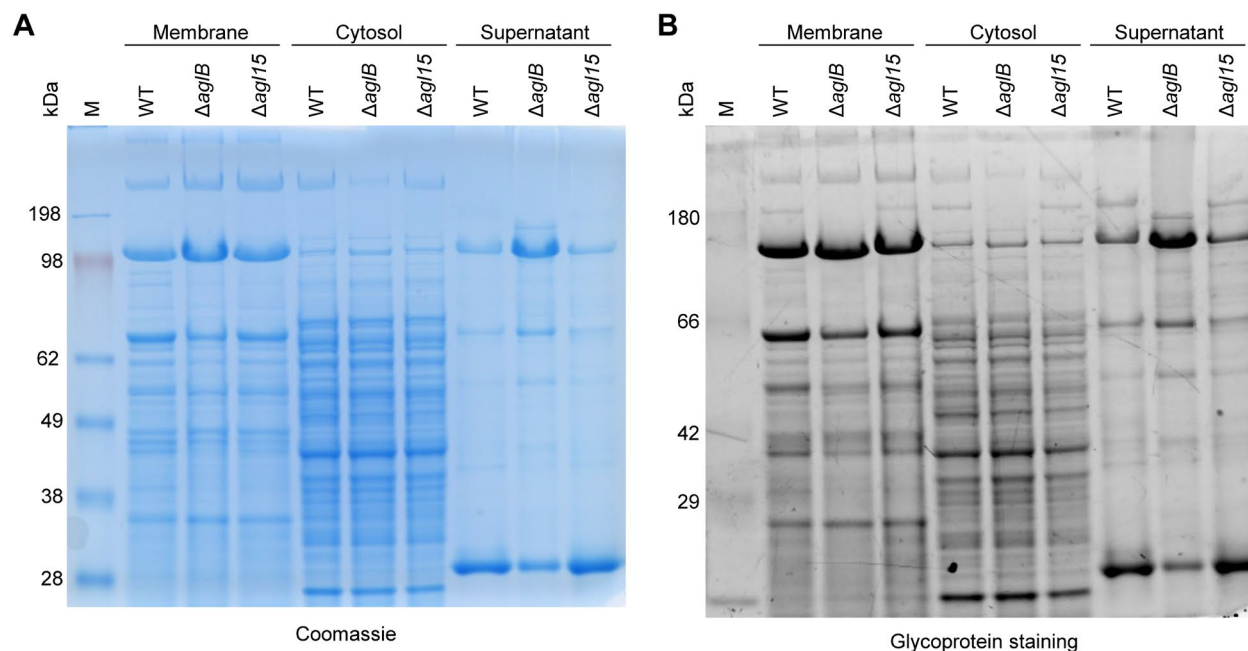
Figure 2. Proteins from cellular fractions of WT, Δagl15 and ΔaglB exhibit comparable glycoprotein staining. Samples from WT, Δagl15 and ΔaglB strains grown to mid-logarithmic and early-stationary growth phase (mixed using equal culture volumes) were fractionated into membrane, cytosol and culture supernatant. Equal protein amounts for each strain (7.5 µg for membrane and cytosol, 5 µg for supernatant) were separated by LDS-PAGE and stained using Coomassie brilliant blue (**A**) or Pro-Q Emerald 300 glycoprotein staining (**B**).

Proteomic analysis of *Hfx. volcanii* reveals the largest archaeal glycoproteome described to date

For the mass spectrometric analysis of the fractions obtained from WT, Δagl15 and ΔaglB strains, samples from cultures in mid- and late-logarithmic growth phases were mixed using equal culture volumes. This allows for a comprehensive qualitative assessment of glycosylation in *Hfx. volcanii* across multiple growth phases in which different glycoproteins might be expressed. In order to identify intact glycopeptides, we used the combination of multiple protein database search engines as previously described within the ArcPP project (Schulze et al., 2020). Importantly, the glycans of each step in the two known *N*-glycosylation pathways of *Hfx. volcanii* (in total 10 glycans) (Eichler et al., 2013; Jarrell et al., 2014), as well as di-hexose as the only described *O*-glycosylation for *Hfx. volcanii* so far (Sumper et al., 1990), were included in the search as potential modifications. Since this substantial expansion of the search space could lead to an increase in false positive identifications, a stringent filtering of search results was performed. Besides a posterior error probability (PEP) $\leq$ 1% on the PSM level, (glyco-)peptides were required to be identified by at least two spectra and to have a peptide false discovery rate (FDR) $\leq$ 1%.

Protein database search engines only take into account the fragmentation of the peptide backbone. Therefore, spectra corresponding to glycopeptide identifications were searched subsequently for glycopeptide-specific ions. In analogy to the b- and y-ion series for the peptide backbone, which are due to preferential cleavage of the peptide bond, glycans are frequently cleaved at the intergylcosidic bond (Domon and Costello, 1988). This leads to a Y-ion series (representing the peptide plus the remainder of the glycan, the agylcan version being Y0) and a B-ion series (or oxonium-ion series) representing the cleaved-off part of the glycan (the liberated terminal sugar being B1). For safe glycopeptide identifications, the detection of at least one B-ion and two Y-ions was considered mandatory. While stringent criteria are expected to strongly

reduce the likelihood of false positive identifications, it should be noted that they become more easily satisfied when glycans are longer. Since the glycans of *O*-glycopeptides are very short, we focused in this work on the detailed analysis of *N*-glycopeptides. However, *O*-glycopeptides were nevertheless identified (see below).

Through this extensive glycoproteomic approach, a total of 194 *N*-glycopeptides, spanning 102 *N*-glycosites were identified for 39 *N*-glycoproteins (Figure 3, Table 1). These results demonstrate that *N*-glycosylation occurs in *Hfx. volcanii* to a much higher extent than previously known, as only five *N*-glycoproteins were described before (SLG, ArlA1, ArlA2, PilA1, PilA2 (Esquivel et al., 2016; Sumper et al., 1990; Tripepi et al., 2012)). Notably, the reliability of this search and filtering approach is highlighted by the lack of AglB- and Agl15-dependent glycans in the respective deletion mutants. Furthermore, all *N*-glycoproteins identified in this dataset (PXD021874) are predicted to be secreted or transmembrane proteins (Table 1), which was expected since the OST is located on the extracellular side of the plasma membrane. Exceptions from other datasets (e.g. Ths3 with predicted cytosolic localization, identified in dataset PXD011056) are described and discussed below. Overall, our approach allowed for the reliable large-scale identification of archaeal *N*-glycopeptides.

The vast majority of *N*-glycopeptides was identified in both the WT and Δagl15 strains (Figure 3A,B), suggesting a high similarity of the *N*-glycoproteome between the two strains. Furthermore, almost all identified *N*-glycopeptides harbor AglB-dependent glycans (Table 1). This is in line with previous reports establishing the AglB-dependent pathway as the predominant *N*-glycosylation pathway in *Hfx. volcanii* (Eichler et al., 2013; Kaminski et al., 2013). However, interestingly, Agl15-dependent *N*-glycopeptides were identified in samples from the WT strain, revealing that Agl15-dependent glycosylation does not only occur under low salt conditions. The peptides harboring Agl15-dependent glycans correspond to the SLG, which has previously been studied extensively. Asn residues at positions 13, 83, 274, and 279 had

been shown to be modified with AglB-dependent glycans (Kandiba et al., 2016), while N498 was found harboring an Agl15-dependent glycan (Kaminski et al., 2013) and N732 carrying a glycan from an as yet unknown *N*-glycosylation pathway (Parente et al., 2014) (Figure 3C). Our analysis now identified an additional *N*-glycosite, N370, harboring AglB-dependent glycans. Furthermore, N274 and N279 were found to be modified with AglB- as well as Agl15-dependent glycans, while N498 carried AglB-dependent glycosylation (Figure 3D, Figure S1). Together these results show not only that Agl15-dependent *N*-glycosylation can

occur under the same conditions as AglB-dependent *N*-glycosylation but also that both pathways can modify the same *N*-glycosites.

Peptides containing both *N*-glycosites N274 and N279 were identified harboring either two AglB- or two Agl15-dependent *N*-glycans. Unfortunately, we could not assess whether peptides containing two *N*-glycosites could be modified with one AglB- and one Agl15-dependent *N*-glycan, since the employed protein database search engines do not support the simultaneous search for two different modifications of the same amino acid.
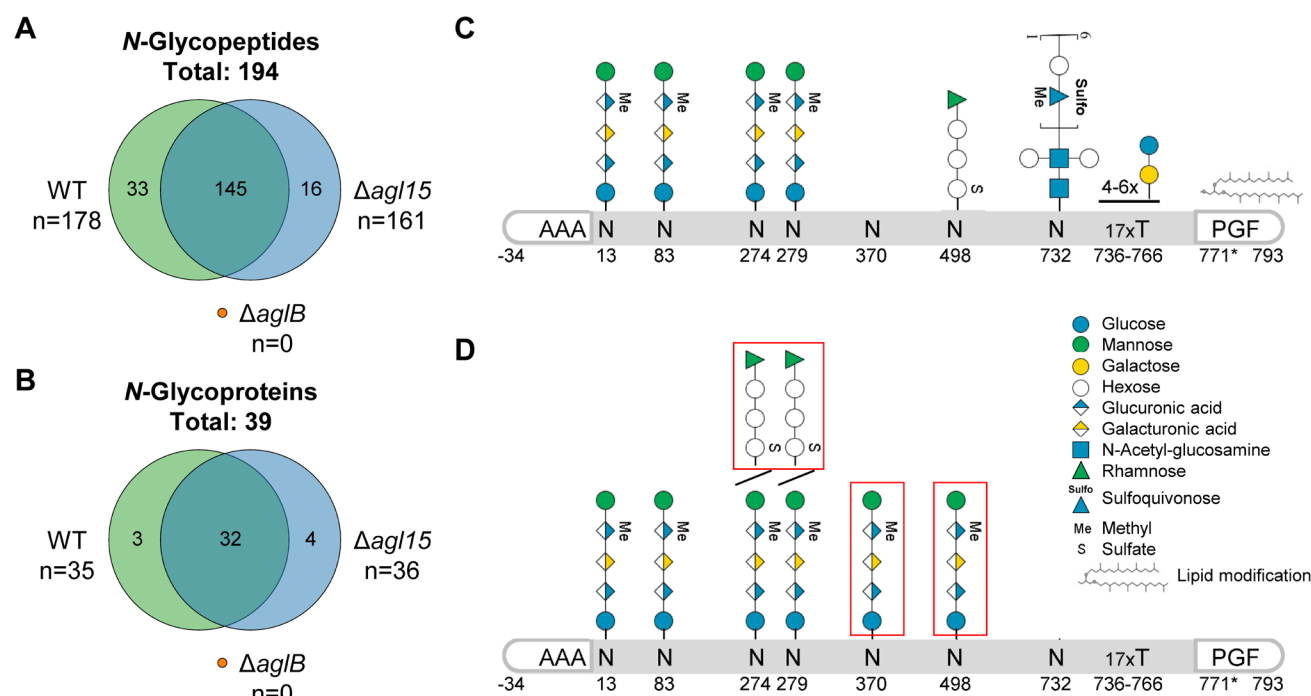
Figure 3. Glycoproteomic analysis of *Hfx. volcanii* reveals concurrent AglB- and Agl15-dependent *N*-glycosylation. Cellular fractions of WT, Δ*agl15* and Δ*aglB* strains were analyzed by LC-MS/MS. The overlap of identified *N*-glycopeptides (**A**) and *N*-glycoproteins (**B**) between the three strains is represented as Venn diagrams. No *N*-glycopeptides were identified for Δ*aglB*. Since *Hfx. volcanii* SLG has been extensively studied previously, the glycosites (numbers indicating the amino acid position after signal peptide cleavage) and corresponding glycans that have been described so far ((Kaminski et al., 2013; Kandiba et al., 2016; Parente et al., 2014; Sumper et al., 1990), figure adapted from (Pohlschroder et al., 2018)) are depicted schematically (**C**). *N*-glycosites and corresponding *N*-glycans that have been identified in this study are shown schematically (**D**). For some *N*-glycosites, multiple *N*-glycans were identified, indicated by diagonal lines. Furthermore, *N*-glycans that have not been identified previously are highlighted by red boxes. It should be noted that *N*-glycopeptides with shorter versions of the AglB-dependent pentasaccharide have been described previously and identified here as well, but are not depicted separately.

Glycoproteomic analysis of the ArcPP datasets further extends glycoprotein identifications

After establishing that our workflow resulted in the reliable identification of glycopeptides, we extended our glycoproteomics analysis to suitable datasets from the ArcPP (Schulze et al., 2020), while at the same time including our newly generated proteomic dataset in ArcPP. Following the same filtering criteria as described above (except for the mandatory detection of oxonium-ions in datasets that did not include low m/z ranges in MS2 spectra), a total of 233 *N*-glycopeptides spanning 114 *N*-glycosites corresponding to 45 *N*-glycoproteins were identified (Figure 4A, Table 1). This represents a remarkable increase in the number of identified *N*-glycoproteins and -peptides of 15% and 20%, respectively, compared to the separate analysis of PXD021874. These results highlight the value of such a combined reanalysis for the identification of complex PTMs like glycosylation. Furthermore, it allows for comparisons between the datasets. When taking into account all peptides, independent of their glycosylation, the vast majority of *N*-

glycoproteins could be identified in all datasets that analyzed whole proteomes (Figure 4B). However, most *N*-glycopeptides were only identified in the dataset generated for this study (PXD021874). Datasets with more than the median eight *N*-glycoprotein identifications (PXD021874, PXD011015, PXD011050) used both trypsin and GluC for protein digestion before MS measurements (Figure 4A), indicating that the cleavage by GluC aided in the detection of *N*-glycopeptides. Further increases in *N*-glycopeptide identifications between PXD021874 and PXD011015/PXD011050 can likely be attributed to cell fractionation, peptide fragmentation using stepped collisional energies, and high sensitivity of the employed instrument. Notably, taking together glycosylated and non-glycosylated peptides and proteins, PXD021874 represents the most comprehensive *Hfx. volcanii* proteome dataset within the ArcPP (Figure S2). Consequently, the inclusion of PXD021874 in the ArcPP increased the number of identified *Hfx. volcanii* proteins from 2,912 to 3,036 (72% and 74.5% of the theoretical proteome, respectively), with a median protein sequence coverage of 54%.

Table 1. Summary of identified *N*-glycoproteins. For each protein that was identified to be *N*-glycosylated in this study, the HVO ID, name and description is given together with the number of identified *N*-glycosites, the *N*-glycan type(s), the number of corresponding PSMs, the dataset(s) in which it was identified to be *N*-glycosylated, and the predicted processing. Datasets are given as PRIDE IDs and it should be noted that PXD021874 corresponds to the dataset generated for this manuscript, while all other PRIDE IDs correspond to datasets of the ArcPP. Cyt, cytosolic; TM, transmembrane domain; Sec, Sec pathway; Tat, twin arginine translocation pathway; Pil, type IV pilin pathway; SPI, signal peptidase I; lipobox, conserved cleavage site motif for lipoproteins; SPIII, signal peptidase III; ArtA, archaeosortase A substrate.

HVO ID	Name	Description	*N*-Glyco-sites	*N*-Glycan type(s)	PSMs	Dataset(s)	Predicted processing
HVO_0307	-	conserved hypothetical protein	3	AglB	201	PXD021874; PXD006877; PXD010824; PXD011012; PXD011050; PXD011056	Sec (SPI)
HVO_0504	-	DUF192 family protein	1	AglB	32	PXD021874	Sec (SPI)
HVO_0778	Ths3	thermosome subunit 3	1	Agl15	4	PXD011056	Cyt
HVO_0892	NosD	ABC-type transport system periplasmic substrate-binding protein	4	AglB	193	PXD021874; PXD011012; PXD011218	Sec (SPI)
HVO_0972	PilA1	pilin PilA	3	AglB	424	PXD021874; PXD006877; PXD009116; PXD010824; PXD011012; PXD011015; PXD011050; PXD011056; PXD014974	Pil (SPIII)
HVO_1014	CoxB1	cox-type terminal oxidase subunit II	2	AglB	13	PXD021874	Sec (SPI)
HVO_1030	-	DUF4382 domain protein	2	AglB	55	PXD021874; PXD011050	Sec (lipobox)
HVO_1176	-	conserved hypothetical protein	2	AglB	54	PXD021874; PXD009116; PXD011218; PXD013046; PXD014974	Sec (lipobox)
HVO_1210	ArlA1	archaellin A1	2	AglB	63	PXD021874; PXD011012; PXD011015; PXD011050	Pil (SPIII)
HVO_1211	ArlA2	archaellin A2	1	AglB	6	PXD021874; PXD011012; PXD011050	Pil (SPIII)
HVO_1259	-	conserved hypothetical protein	2	AglB	47	PXD021874	TM N-term
HVO_1530	AglB	dolichyl-monophosphooligosaccharide--protein glycotransferase AglB	2	AglB	256	PXD021874; PXD010824; PXD011012; PXD011050; PXD011056; PXD014974	>=2 TM
HVO_1624	-	conserved hypothetical protein	1	AglB	3	PXD021874	Tat (lipobox)
HVO_1673	-	conserved hypothetical protein	2	AglB	116	PXD021874; PXD009116; PXD011012; PXD011050; PXD011218	Sec (lipobox)
HVO_1749	-	conserved hypothetical protein	2	AglB	153	PXD021874; PXD010824; PXD011012; PXD011050; PXD011056	Pil (SPIII)
HVO_1802	-	peptidase M10 family protein	1	AglB	2	PXD021874	Sec (lipobox)
HVO_1806	-	conserved hypothetical protein	1	AglB	8	PXD021874	Sec (lipobox)
HVO_1870	-	M50 family metalloprotease	2	AglB	97	PXD021874; PXD010824; PXD011050	>=2 TM
HVO_1944	-	probable transmembrane glycoprotein / HTH domain protein	1	AglB	16	PXD021874	Sec (SPI)
HVO_1945	-	conserved hypothetical protein	4	AglB	183	PXD021874	Tat (SPI)
HVO_1976	SecD	protein-export membrane protein SecD	2	AglB	33	PXD021874; PXD010824; PXD011012; PXD013046; PXD014974	Sec (SPI)
HVO_1988	-	GATase domain protein	1	AglB	31	PXD021874	Sec (SPI)
HVO_2062	PilA2	pilin PilA	2	AglB	56	PXD021874; PXD011012; PXD011015; PXD011050	Pil (SPIII)*
HVO_2066	-	conserved hypothetical protein	1	AglB, Agl15	27	PXD006877	Sec (SPI)
HVO_2070	-	conserved hypothetical protein	2	AglB	100	PXD021874; PXD011012; PXD011015; PXD013046	Sec (SPI)

|							
HVO_2071	-	probable secreted glycoprotein	4	AglB	134	PXD021874; PXD006877; PXD011012; PXD011050	Sec (SPI)
HVO_2072	SLG	S-layer glycoprotein	6	AglB, Agl15	2364	PXD021874; PXD006877; PXD007061; PXD009116; PXD010824; PXD011012; PXD011015; PXD011050; PXD011056; PXD011218; PXD013046; PXD014974	Sec (SPI)*, ArtA
HVO_2074	-	probable secreted glycoprotein	1	AglB	2	PXD021874; PXD011050	Sec (SPI)
HVO_2076	-	probable secreted glycoprotein (nonfunctional)	3	AglB	42	PXD021874	Sec (SPI)
HVO_2081	-	pectin lyase domain protein	3	AglB	52	PXD021874; PXD011050	Sec (SPI)
HVO_2082	-	conserved hypothetical protein	2	AglB	117	PXD021874; PXD006877; PXD011012; PXD011050	Sec (SPI)
HVO_2084	-	ABC-type transport system permease protein (probable substrate macrolides)	1	AglB	50	PXD021874	>=2 TM
HVO_2160	-	probable secreted glycoprotein	19	AglB	1561	PXD021874; PXD006877; PXD007061; PXD009116; PXD010824; PXD011012; PXD011050; PXD011056; PXD011218; PXD013046; PXD014974	Sec (SPI), ArtA
HVO_2161	-	probable secreted glycoprotein	1	AglB	5	PXD021874; PXD011012	Sec (SPI)
HVO_2167	-	conserved hypothetical protein	1	AglB	12	PXD021874; PXD006877	Sec (SPI)
HVO_2172	-	conserved hypothetical protein	3	AglB	82	PXD021874	Sec (SPI)
HVO_2173	-	DUF1616 family protein	4	AglB	83	PXD021874	>=2 TM
HVO_2533	-	conserved hypothetical protein	5	AglB	103	PXD021874; PXD007061; PXD009116; PXD011012; PXD011056; PXD011218; PXD013046	Sec (SPI), ArtA
HVO_2535	-	conserved hypothetical protein	1	AglB	2	PXD021874	Sec (lipobox)
HVO_2634	-	conserved hypothetical protein	1	AglB	40	PXD011012	Sec (SPI)
HVO_A0039	-	conserved hypothetical protein	2	AglB	93	PXD021874; PXD007061; PXD013046	Sec (SPI)
HVO_A0466	-	conserved hypothetical protein	1	AglB	9	PXD021874	Sec (SPI)
HVO_A0499	-	conserved hypothetical protein	2	AglB	27	PXD021874	Sec (SPI)
HVO_B0194	-	LppX domain protein	4	AglB	46	PXD021874	Sec (SPI)
HVO_C0054	-	hypothetical protein	3	AglB	29	PXD021874	Tat (SPI)
HVO_0307	-	conserved hypothetical protein	3	AglB	201	PXD021874; PXD006877; PXD010824; PXD011012; PXD011050; PXD011056	Sec (SPI)

*The processing for this entry has corrected upon manual curation

The identification of additional N-glycopeptides confirmed and extended our findings about the concurrence of AglB- and Agl15-dependent glycosylation. The SLG peptide containing N-glycosites N274 and N279 was found modified with AglB- and Agl15-dependent glycans in an additional dataset (PD011012, Table 1). Furthermore, both N-glycan types were found on N-glycopeptides corresponding to a second protein, HVO_2066 (PXD006877, Table 1). However, the vast majority of identified N-glycopeptides harbored exclusively AglB-dependent N-glycans, and only one N-glycopeptide with exclusively Agl15-dependent N-glycans was found. Surprisingly, the latter originates from the thermosome subunit 3 (Ths3), which is the only identified N-glycoprotein with a predicted cytosolic localization (Table 1). The identification of Ths3 as an N-glycoprotein is well-supported by four corresponding MS2 spectra through almost complete b- and y-ion series as well as a complete series of glycan-specific Y-ions (Figure S3). Therefore, this is unlikely to represent a false positive identification. Thermosome subunits were described to be surface-associated in *Sulfolobus shibatae* (Trent et al., 2003), and the identification of thermosome subunits in lectin affinity enrichments from *Methanosarcina mazei* and *Methanosarcina acetivorans* (Francoleon et al., 2009) further supports the surface association and N-glycosylation of Ths3. Thus, our identification of N-glycosylated peptides of Ths3 provides additional evidence for the hypothesis that thermosome subunits are not restricted to a cytosolic localization.
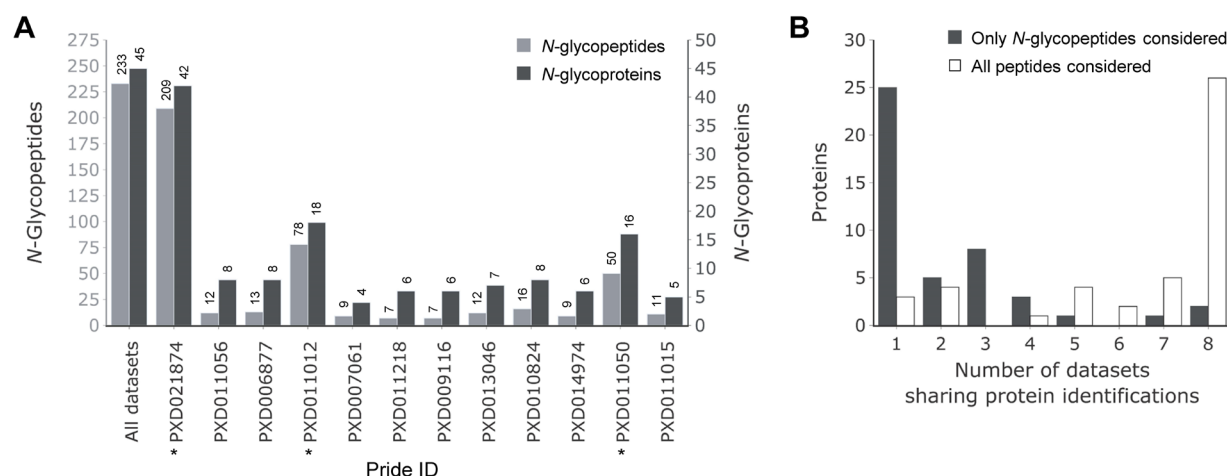
Figure 4. Glycoproteomic analysis of ArcPP datasets extends the *Hfx. volcanii* N-glycoproteome. Datasets included in the ArcPP, supplemented by dataset PXD021874 from the current study, were reanalyzed including *Hfx. volcanii* glycans as potential modifications. **A**, The number of identified *N*-glycopeptides (light grey) and *N*-glycoproteins (dark grey) for each dataset is shown as a barplot (sorted by the total number of identified proteins, see Figure S2). Datasets, for which the enzymatic digest was performed with trypsin as well as GluC, are marked with an asterisk (*). **B**, For each identified *N*-glycoprotein, the number of whole proteome datasets which share this identification was determined. The number of *N*-glycoproteins identified in the given number of datasets is represented as a bar plot. This analysis was performed taking into account either only *N*-glycopeptides (dark grey) or all peptides identified for a protein that was determined to be *N*-glycosylated in any of the datasets (white).

Unknown functions but high degree of phylogenetic conservation within Halobacteria for most identified N-glycoproteins

After revealing this remarkably extensive and complex *N*-glycoproteome of *Hfx. volcanii*, we continued our study with a phylogenetic analysis of all 45 identified *N*-glycoproteins. While the functions and *N*-glycosylation of SLG, ArlA1, ArlA2, PilA1 and PilA2 have been established (Esquivel et al., 2016; Tamir and Eichler, 2017; Tripepi et al., 2012), we were able to identify additional *N*-glycoproteins with known functions that previously were not considered to be *N*-glycosylated. Interestingly, besides SecD (membrane protein involved in protein-export) this includes the OST AglB, which itself is *N*-glycosylated in an AglB-dependent manner. Moreover, for a few other *N*-glycoproteins, functions can be inferred from characterized homologs and/or conserved domains: CoxB1, a subunit of a cox-type terminal oxidase (HVO_1014); a thermosome subunit (Ths3, HVO_0778); one substrate-binding protein and one permease component of ABC-type transport systems (HVO_0892 and HVO_2084); as well as an M10 family metallopeptidase and an M50 family metalloprotease (HVO_1802 and HVO_1870).

However, a large part of the identified *N*-glycoproteins are uncharacterized (two-thirds of *N*-glycoproteins compared to one-third of the theoretical proteome), lacking characterized homologs, that are closely enough related to imply similar functions. The vast majority of these proteins is even devoid of InterPro domain assignments, and therefore belongs to the genomic dark matter. Therefore, for all identified *N*-glycoproteins, the current annotation has been evaluated in order to inspect if current knowledge allows the assignment of protein functions (Table S1). During this analysis, proteins were also subjected to gene synteny analysis using SyntTax (Oberto, 2013). Furthermore, all proteins were categorized with respect to the taxonomic range of their orthologs employing OrthoDB (Table S1) (Zdobnov et al., 2020). Within the taxonomic range, the number of genera coding for an ortholog was determined, revealing that for most *N*-glycoproteins only a few orthologs were found outside the class Halobacteria (Figure 5). However, results indicated a higher degree of conservation within the Halobacteria. Therefore, within this class, the same phylogenetic analysis was done at the species level and orthologs for the majority of species and *N*-glycoproteins were identified (Figure S4).

Notably, several *N*-glycosylated proteins are encoded in genomic vicinity to each other (Table S1). Besides the known cluster of *arlA1/arlA2*, this includes a region spanning from *hvo_2062* (*pilA2*) to *hvo_2084* (ABC-type transport system component), which encode for 10 *N*-glycoproteins, one of which is SLG (*hvo_2072*). Furthermore, a cluster from *hvo_2160* to *hvo_2173* encodes five *N*-glycoproteins, with HVO_2160 being the largest *Hfx. volcanii* protein that also contains the highest number of confirmed *N*-glycosites (19). However, the reason for this clustering remains to be elucidated, as the function of most proteins encoded within these clusters is unknown. This is also the case for three additional gene clusters corresponding to *N*-glycoproteins of unknown function.

## Evidence for N-glycosylation at non-canonical N-glycosites

In vitro experiments previously suggested that AglB-dependent *N*-glycosylation in archaea may not be strictly specific to the canonical N-X-S/T sequon (Zeitler et al., 1998). We therefore checked for non-canonical *N*-glycosites by omitting the requirement for N-X-S/T sequons in the final filtering of glycopeptide identifications. Indeed, this revealed the confident identification of five *N*-glycopetides corresponding to three non-canonical *N*-glycosites from three different proteins (Table 2). Interestingly, this includes the heavily *N*-glycosylated HVO_2160 as well as PilA6, which was previously suggested to be *N*-glycosylated based on an electrophoretic mobility shift in Δ*aglB* extracts (Esquivel et al., 2016), a result that thus far had not been confirmed by MS. Two additional candidates for non-canonical *N*-glycopeptides were identified from the pyruvate kinase PykA (HVO_0806) and the 3-methyl-2-oxobutanoate hydroxymethyltransferase PanB2 (HVO_2703). However, these proteins are predicted to be cytosolic and the identifications are based on only two PSMs each and lack glycan-specific B-ions, since the *m/z*-range of MS2 spectra in the corresponding dataset (PXD006877) does not allow for the identification of mono- and disaccharide B-ions. This renders the identification of non-canonical *N*-glycopeptides in PykA

8

and PanB2 less reliable in comparison to those of HVO_2160 and PilA6 or the identified canonical *N*-glycosites.

Furthermore, it should be noted that *N*-glycans shorter than the AglB-dependent tetra-/pentasaccharide and Agl15-dependent tetrasaccharide were identified. While this is in line with previous reports

of shorter *N*-glycans linked to SLG (Kandiba et al., 2016), it is unclear whether these represent technical artefacts (e.g. from sample preparation or in-source fragmentation) or biological products (e.g. from *N*-glycan trimming or low substrate specificity of AglB). Therefore, these *N*-glycopeptides have not been further analyzed in this study.
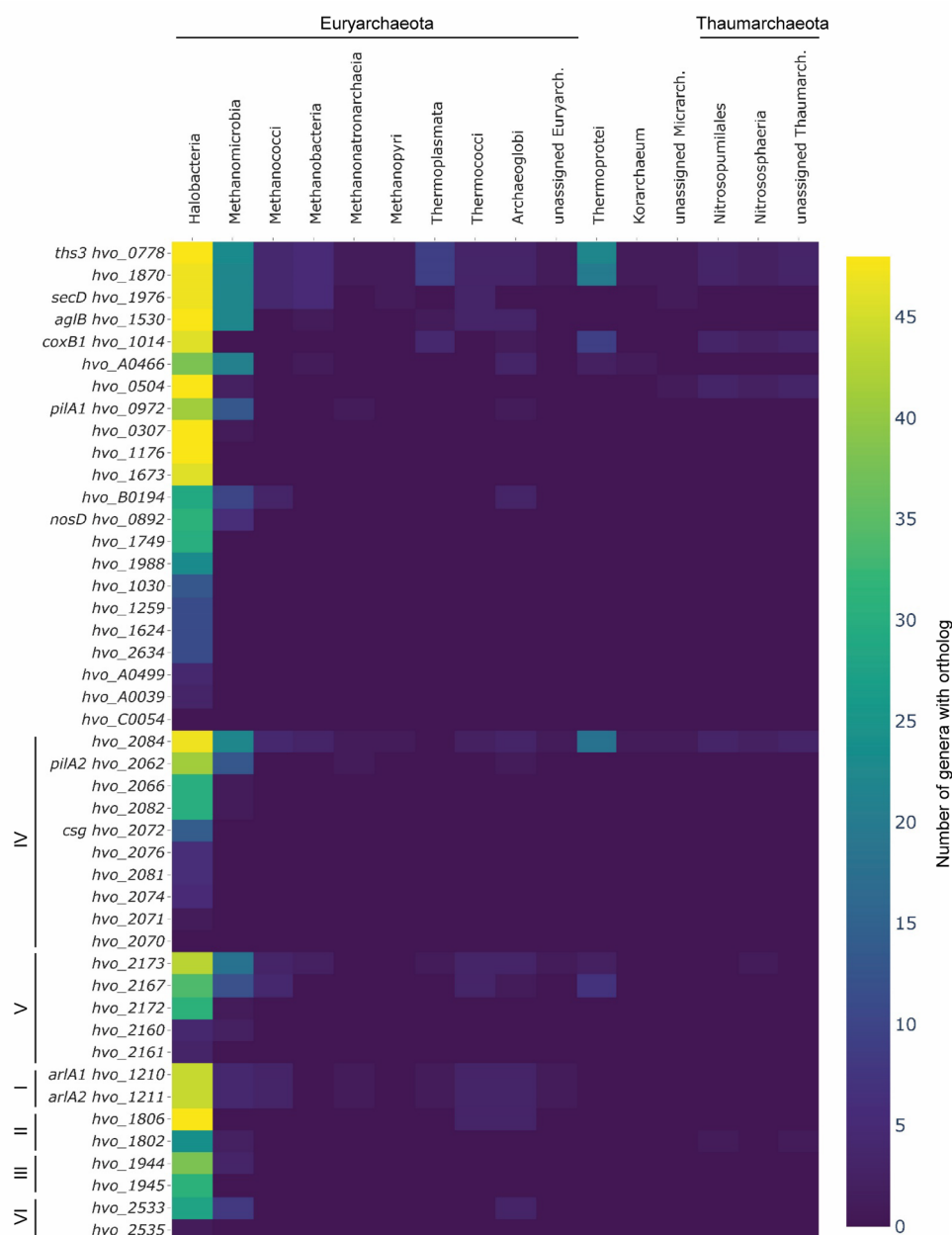
Figure 5. Phylogenetic analysis of identified *N*-glycoproteins reveals conservation within Halobacteria. The HVO IDs of all identified *N*-glycoproteins have been subjected to orthoDB analysis in order to determine orthologues across the archaeal domain. For each *N*-glycoprotein encoding gene the number of genera within the different archaeal taxonomic classes is given as a heatmap ranging in color from yellow (48 genera with orthologous proteins) to purple (0 genera). Genes were sorted by the overall number of genera with orthologous proteins, while clusters with multiple *N*-glycoprotein encoding genes (see TableS1) were grouped separately from the remaining genes.

Table 2. Summary of identified *N*-glycopeptides with non-canonical *N*-glycosite. For each non-canonical *N*-glycopeptide that was identified, the HVO ID, name and description is given together with the peptide sequence (*N*-glycosite is marked in bold), the *N*-glycan type, the number of corresponding PSMs, the dataset(s) in which it was identified to be *N*-glycosylated, and the predicted processing. Cyt, cytosolic; TM, transmembrane domain; Sec, Sec pathway; Pil, type IV pilin pathway; SPI, signal peptidase I; SPIII, signal peptidase III.

HVO ID	Name	Description	*N*-Glycan type	*N*-Glycopeptide sequence	Datasets	PSMs	Predicted processing
HVO_0806	PykA	pyruvate kinase	Agl15	IERAGAVD**N**LDEIIQAAHGVMVAR	PXD006877	2	Cyt
HVO_2160	-	probable secreted glycoprotein	AglB	MPS**N**ANIMGVTPGSR	PXD021874	25	Sec (SPI)
HVO_2173	-	DUF1616 family protein	AglB	LVRGEPASLVLGVG**N**QE	PXD021874	2	>=2 TM
HVO_2703	PanB2	3-methyl-2-oxobutanoate hydroxymethyl-transferase	AglB	AHAEAGAFSLVLEHVPA**N**LAK	PXD006877	2	Cyt
HVO_A0633	PilA6	pilin PilA	AglB	VVWTSESGS**N**SATLQR	PXD021874	20	Pil (SPIII)

Identification of likely *O*-glycoproteins

Knowledge about archaeal *O*-glycosylation is so far limited to the early description of disaccharides *O*-linked to the SLG in *Hbt. salinarum* and *Hfx. volcanii* (Mescher and Strominger, 1976; Sumper et al., 1990). In order to shine new light onto the *O*-glycoproteome of *Hfx. volcanii*, we included the search for potential di-hexose modifications of Ser and Thr in our analysis. While we used the same filtering approach as applied to the identification of *N*-glycopeptides, the shorter length of *O*-glycans reduces the possibilities of detecting *O*-glycan-derived fragment ions. Nevertheless, we identified 23 *O*-glycopeptides corresponding to 20 *O*-glycoproteins (Table 3). Interestingly, the predictions of cellular localisation indicated that both cytosolic and secreted proteins can be *O*-glycosylated in *Hfx. volcanii*. The *O*-glycosylation pathway(s) in archaea remains to be elucidated, hence preventing us from using a deletion mutant lacking *O*-glycans for the stringent evaluation of potential false positives. Nevertheless, the identified *O*-glycopeptides are promising candidates.

Table 3. Summary of identified *O*-glycoprotein candidates. For each protein that has been identified to be likely *O*-glycosylated in this study, the HVO ID, name and description is given together with the number of identified *O*-glycopeptides, the number of corresponding PSMs and dataset(s), and the predicted processing. Cyt, cytosolic; Sec, Sec pathway; Tat, twin arginine translocation pathway; SPI, signal peptidase I; lipobox, conserved cleavage site motif for lipoproteins; ArtA, archaeosortase A substrate.

HVO ID	Name	Description	*O*-Glyco-peptides	Datasets	PSMs	Predicted processing
HVO_0154	-	conserved hypothetical protein	3	PXD021874	20	Tat (lipobox)
HVO_0306	-	probable transmembrane glycoprotein / HTH domain protein	1	PXD021874	13	Sec (SPI)
HVO_0349	RpoA1	DNA-directed RNA polymerase subunit A'	1	PXD006877	3	Cyt
HVO_0359	Tef1a1	translation elongation factor aEF-1 alpha / peptide chain release factor aRF-3	1	PXD006877	7	Cyt
HVO_0654	Rpl43e	50S ribosomal protein L43e	1	PXD006877	4	Cyt
HVO_0677	AspS	aspartate--tRNA(Asp/Asn) ligase	1	PXD011056	6	Cyt
HVO_0778	Ths3	thermosome subunit 3	1*	PXD013046	8	Cyt
HVO_0869	GltB	glutamate synthase (ferredoxin) large subunit	1	PXD006877	2	Cyt
HVO_1148	Rps15	30S ribosomal protein S15	1	PXD006877	4	Cyt
HVO_1198	-	UspA domain protein	1	PXD006877	8	Cyt
HVO_1597	-	conserved hypothetical protein	1	PXD021874	4	Tat (lipobox)
HVO_2071	-	probable secreted glycoprotein	2*	PXD021874	12	Sec (SPI)
HVO_2160	-	probable secreted glycoprotein	1*	PXD021874	5	Sec (SPI), ArtA
HVO_2172	-	conserved hypothetical protein	1*	PXD021874	3	Sec (SPI)
HVO_2226	TrpD2	probable phosphoribosyltransferase (homolog to anthranilate phosphoribosyltransferase)	1	PXD006877	6	Cyt
HVO_2413	Tef1a2	translation elongation factor aEF-1 alpha / peptide chain release factor aRF-3	1	PXD006877	2	Cyt
HVO_2487	Asd	aspartate-semialdehyde dehydrogenase	1	PXD006877	3	Cyt

HVO_2580	NadB	L-aspartate oxidase	1	PXD006877	3	Sec (SPI)
HVO_A0380	DppA8	ABC-type transport system periplasmic substrate-binding protein (probable substrate dipeptide/oligopeptide)	1	PXD021874	2	Tat (lipobox)
HVO_B0050	CobN	ATP-dependent cobaltochelatase subunit CobN	1	PXD006877	2	Cyt

* For these proteins, *N*-glycopeptides were identified as well.

Discussion

Protein glycosylation is involved in a variety of cellular processes in prokaryotes. In this study, we extended the phenotypic analysis of *N*-glycosylation pathway mutants in the model archaeon *Hfx. volcanii*. Our results revealed that *N*-glycosylation is not only involved in cell growth and motility, but also affects colony morphology and cell shape. Furthermore, we could show that the deletion of *agl15*, which is required for an *N*-glycosylation type that was so far described to occur only under low-salt conditions, has phenotypic effects under optimal salt conditions as well.

In order to gain insights into the underlying molecular basis for these effects, we developed a stringent workflow for the in-depth glycoproteomic analysis of *Hfx. volcanii*. With the identification of 45 *N*-glycoproteins, this approach resulted in the most extensive archaeal *N*-glycoproteome described so far. Importantly, this glycoproteomic workflow is applicable to a broad range of archaea with known *N*-glycosylation pathways and even allows for the reanalysis of existing MS data as shown by the incorporation of ArcPP datasets. Furthermore, it revealed the surprising concurrence of the AglB- and Agl15-dependent *N*-glycosylation pathways. The identification of *N*-glycopeptides corresponding to both pathways implies their activity under the same optimal salt conditions. Remarkably, datasets in which AglB- and Agl15-dependent *N*-glycopeptides were identified comprised a broad range of growth conditions (mid-logarithmic and early-stationary growth phase, planktonic and biofilm cells, oxidative stress conditions). While quantitative analyses are beyond the scope of this manuscript, these results indicate that neither AglB- nor Agl15-dependent *N*-glycosylation is limited to a specific growth condition. Additionally, our qualitative analysis showed that the two independent *N*-glycosylation pathways are able to modify the same *N*-glycosites.

It should be noted that Agl15-dependent *N*-glycans were not identified in samples from the Δ*aglB* mutant. This was surprising because previous studies indicated that Agl15-dependent glycosylation of the SLG increases in the absence of AglB-dependent glycosylation (Kaminski et al., 2013). However, these studies were performed with purified SLG, allowing for a higher sensitivity than the cellular fractions analyzed here. The decreased stability of the SLG in the Δ*aglB* mutant might have further hampered the identification of Agl15-dependent *N*-glycopeptides in the Δ*aglB* mutant. Furthermore, the *N*-glycosite harboring Agl15-dependent glycans analyzed by Kaminski et al. (N498) differs from those identified here (N274 and N279). Therefore, it cannot be excluded that changes in the abundance of Agl15-dependent *N*-glycopeptides depend on the *N*-glycosite. It is worth noting that even in the dataset PXD011056, which contains samples from cultures grown under low-salt conditions, no increase in the identification of Agl15-dependent *N*-glycopeptides could be seen. In line with this, the vast majority of enzymes corresponding to the Agl15-dependent *N*-glycosylation pathway are less abundant under low-salt conditions compared to optimal salt conditions, as shown by the quantification performed by Jevtić et al. (PXD011056, (Jevtić et al., 2019)). While different low-salt regimes were applied in the various studies, it can be concluded that Agl15-dependent *N*-glycosylation is not specific to low-

salt conditions. This is in line with the physiological phenotypes observed here for the Δ*agl15* mutant under optimal salt conditions.

Overall, our results highlight a remarkable complexity of the glycosylation pathways in *Hfx. volcanii* which poses questions about their regulation. Interestingly, AglB itself was found to be *N*-glycosylated. The eukaryotic counterpart of AglB, the OST complex subunit STT3, has previously been shown to be *N*-glycosylated as well, with at least one *N*-glycan playing an essential role in the assembly and function of the complex (Bai et al., 2018; Li et al., 2005; Wild et al., 2018). Despite the evolutionary proximity of archeal AglB and eukaryotic STT3 (Nikolayev et al., 2020) the *N*-glycosylation of AglB itself had not been previously revealed. While the archaeal OST acts as a single-subunit enzyme instead of a complex, its *N*-glycan could nevertheless serve regulatory functions and/or interact with other proteins.

The importance of *N*-glycosylation for S-layer stability, motility and adhesion via the modification of SLG, archaellins and type IV pilins, respectively, has been shown previously (Esquivel et al., 2016; Tamir and Eichler, 2017; Tripepi et al., 2012). However, the extent of *N*-glycosylation in *Hfx. volcanii* revealed here indicates that additional proteins could be involved in these processes and that a variety of other cellular functions may be affected. For example, while *N*-glycosylation has been shown to be crucial for mating (Shalev et al., 2017), SLG has been suggested to play a role in this process mainly because it is the sole component of the *Hfx. volcanii* S-layer and is known to be highly *N*-glycosylated. However, our results show that a multitude of cell surface proteins in *Hfx. volcanii* are *N*-glycosylated, some of which have been linked to mating. A recent transcriptomic study of mating in *Hfx. volcanii* included eight genes, for which we showed *N*-glycosylation of their gene products, in a cluster that indicated elevated transcript levels during the initial timepoints of mating (Makkay et al., 2020). Besides SLG, AglB and CoxB1, this comprised various proteins of unknown function (HVO_1749, HVO_1981, HVO_2071, HVO_2533, HVO_C0054). Interestingly, HVO_2533 is also a predicted ArtA-substrate and *artA* deletion strains have been shown to be mating deficient (Abdul Halim et al., 2013). Other confirmed or predicted ArtA-substrates that are *N*-glycosylated are SLG and HVO_2160, the largest *Hfx. volcanii* protein.

The lack of ArtA has also been associated with a defect in transitioning from rod- to disk-shaped cells (Abdul Halim et al., 2016; Abdul-Halim et al., 2020). Since our results revealed shape phenotypes of Δ*aglB* as well as Δ*agl15* mutants, the *N*-glycosylation of ArtA-substrates may be important for their function and could therefore be involved in the regulation of *Hfx. volcanii* cell shape. Alternatively, the *N*-glycosylation of ABC transporters that we identified here could have an effect on (micro-)nutrient acquisition in *Hfx. volcanii*, which has been linked to cell shape as well (Silva et al., 2020). This may also explain differing growth curves between the WT and *N*-glycosylation pathway mutant strains. Finally, the *N*-glycosylation of SecD, a component of the secretory pathway, could affect cell surface biogenesis and thereby cell shape as well as other surface-related processes.

In conclusion, the extensive and complex *N*-glycoproteome of *Hfx. volcanii* that we revealed here shines new light on a variety of cellular functions. While the specific effects of *N*-glycosylation on the various identified proteins remain to be studied in detail, their identification

11

represents an essential first step in analyzing the roles of *N*-glycosylation in archaea. The importance of this is highlighted not only by the various phenotypes of *N*-glycosylation pathway mutants but also by the phylogenetic conservation of identified *N*-glycoproteins with unknown functions within Halobacteria.

Materials and Methods

Strains and growth conditions. *H. volcanii* H53 as well as the Δ*aglB* and Δ*agl15* mutants were grown at 45 °C in liquid or on solid agar semi-defined casamino acid (Hv-Cab) medium (Silva et al., 2020). H53 cultures were supplemented with tryptophan (+Trp) and uracil (+Ura) at a final concentration of 50 μg ml^{-1}. The Δ*aglB* and Δ*agl15* strains were derived from the background strain WR536 (Δ*pyrE2*; Δ*trpA*) and carried a *trpA* insertion in the respective gene (Abu-Qarn and Eichler, 2006); they were therefore grown on Hv-Cab +Ura.
For direct comparison of colony morphology and motility phenotypes, all strains were grown on Hv-Cab +Trp +Ura, but phenotypes of deletion mutants were confirmed on Hv-Cab +Ura. Colony morphology was imaged after 5 days of incubation at 45 °C using a Nikon D3500 DX.

Motility assay. Motility assays were performed as described previously (Tripepi et al., 2010) using Hv-Cab media containing 0.35% agar supplemented with Ura and Trp as required for the analyzed strain. Halos around the stab-inoculation site were imaged after 5 days of incubation at 45 °C using a Nikon D3500 DX.

Growth curve. Liquid Hv-Cab starting cultures (5 mL) were inoculated from single colonies and grown at 45 °C under shaking (250 rpm). After reaching an OD_{600} of 0.3 - 0.5 (measured in the culture tube, path length approximately 1.5 cm), all cultures were diluted to an OD_{600} of 0.05 in 20 mL to obtain comparable conditions at the start of the growth curve. Subsequently, OD_{600} was monitored for 74 h; these measurements were performed by pipetting 250 μL of culture into a 96-well plate and measuring the absorption with a Biotek PowerWaveX2 microplate spectrophotometer. When cultures reached an OD above 0.75, samples were diluted 1:5 before measurement.

Cell shape assessment. Liquid Hv-Cab (5 mL) was inoculated with single colonies and grown at 45 °C under shaking (250 rpm). At early logarithmic (OD_{600} of 0.03 - 0.05), mid logarithmic (OD_{600} of 0.35 - 0.5) and late logarithmic (OD_{600} of 1.6 - 2.0) growth phase, 1 mL of culture was centrifuged (5 min, 6000 g) and resuspended in 20 μL of medium (for early logarithmic phase, correspondingly higher volumes were used for higher ODs). Resuspended cells were observed on a Leica DMi8 microscope at 100x magnification, using differential inference microscopy. Images were taken with a Leica DFC9000GT camera attached to the microscope.

Cell fractionation. A 5 mL Hv-Cab starting culture (inoculated from single colonies) was grown at 45 °C under shaking (250 rpm) until it reached an OD_{600} of 0.03 - 0.05 (measured in the culture tube, path length approximately 1.5 cm). The culture was then diluted to 25 mL (OD_{600} of 0.01) and was kept growing under the same conditions. Once the culture reached an OD_{600} of 0.03, a 10 mL sample was taken, while the remaining culture continued growing for another 24 h, reaching early stationary phase (OD_{600} > 1.7), at which point a second 10 mL sample was taken. Each sample was processed as follows immediately after it has been taken.
(i) The culture was centrifuged at 6,000 g for 10 min to separate the cells from the supernatant (SN). The SN fraction was transferred into new tubes and centrifuged again two more times, once at 10,000 g for 10 min and once at 16,000 g for 20 min, to remove contaminating cells and potential large cell debris. Afterwards, the SN from both samples (mid-

logarithmic and early-stationary growth phase) were combined and concentrated to approximately 150 μL in centrifugal filter units with a 3 kDa molecular mass cutoff (Amicon ultra centrifugal filters, 0.5 ml, Millipore, centrifugation at 14,000 g).
(ii) The cell pellets from step (i) of both samples were resuspended and combined in 1 mL phosphate-buffered saline (PBS; 2.14 M NaCl, 2.68 mM KCl, 10.14 mM Na_2HPO_4, 1.76 mM KH_2PO_4, pH 7.4) containing 10 mM EDTA as well as 1 mM 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride and 1 mM phenylmethylsulfonyl fluoride as protease inhibitors. Cells were lysed by freezing (-80 °C) and thawing (on ice) them three times. Cellular DNA was digested by adding 10 μg of DNAse I and incubating the mixture at 37 °C for 30 min. The lysate was centrifuged at 10,000 g for 5 min at 4 °C to pellet unlysed cells. Subsequently, the SN was centrifuged for 30 min at 300,000 g and 4 °C in a Beckman TL-100 ultracentrifuge. The SN of this step was transferred to a new tube and centrifuged again (using the same settings) to remove potential membrane contaminations. The resulting SN corresponds to the Cyt fraction.
(iii) The pellet after the first ultracentrifugation step was washed in 600 μL of ice-cold PBS and centrifuged again using the same settings. After carefully removing the SN, the pellet representing the membrane (Mem) fraction was resuspended in 300 μL 100 mM Tris/HCl buffer (pH 7.4) containing 2% SDS.
Each cellular fraction was transferred to centrifugal filter units with a 3 kDa molecular mass cutoff, washed three times with 400 μL H_2O, followed by a protein solubilization step with 400 μL 100 mM Tris/HCl buffer (pH 7.4) containing 2% SDS, incubation at 55 °C for 15 min, and two additional washes with 400 μL H_2O. Afterwards, the protein concentration of each fraction was determined using a bicinchoninic acid assay (BCA Protein Assay Kit by Thermo Scientific). Samples were stored at -80 °C until further use.

SDS-PAGE and glycoprotein staining. Protein samples of each fraction (7.5 μg for Mem and Cyt, 5 μg for SN) were mixed with NuPAGE loading buffer and dithiothreitol (DTT, final concentration 50 mM), incubated at 55 °C for 15 min and then separated by electrophoresis on a 4-12% v/v Bis-Tris NuPAGE gel (1.5 mm, 10 well) using MOPS SDS running buffer (NuPAGE). Two gels were run in parallel, one of which was stained overnight with Coomassie brilliant blue for protein detection (destaining with H_2O), while the other was stained for glycoproteins using the Pro-Q Emerald 300 glycoprotein staining kit (Invitrogen) following manufacturer's instructions. Gels were imaged using an Amersham Imager 600 (GE Healthcare Life Sciences).

Mass spectrometric analysis. Protein samples of each fraction (30 μg for Mem and Cyt, 10 μg for SN) were processed and digested using S-Trap mini spin columns (ProtiFi) and following the manufacturer's instructions. Samples were mixed with an equal volume of solubilization buffer (10% SDS in 100 mM triethylammonium bicarbonate (TEAB), pH 7.55) and subsequently reduced by adding 20 mM DTT (15 min, 55 °C) and alkylated by adding 40 mM iodoacetamide (30 min, room temperature, dark). After acidifying the solution with phosphoric acid (final concentration 1.2%), S-Trap binding buffer (90% methanol, 100 mM TEAB, pH 7.1) was added at a ratio of 1:7 (sample volume : binding buffer volume). The mixture was loaded onto the S-Trap spin column and centrifuged at 4,000 g for 30 s. Captured proteins were washed three times with 400 μL S-Trap binding buffer. Proteases were added at a ratio of 1:25 (sample protein weight : enzyme weight) in 125 μL digestion buffer (50 mM TEAB for trypsin; 10 mM phosphate buffer, pH 8.0 for GluC). Digestion was performed overnight at 37 °C. Peptides were eluted in three steps: 80 μL H_2O, followed by 80 μL 0.2 % formic acid, and finally 80 μL 50% acetonitrile (ACN). The spin column was centrifuged

at 1,000 g for 60 s at each step. Eluted peptides from the different steps were pooled and dried in a vacuum centrifuge.

All samples were desalted using homemade C18 stage-tips (3 M Empore Discs) as previously described (Rappsilber et al., 2007), with minor modifications. Briefly, columns were conditioned with 100 µL ACN and equilibrated two times with 100 µL 0.1% trifluoroacetic acid (TFA). Samples were resuspended in 0.1% TFA and loaded onto the stage-tip. After washing three times with 100 µL 0.1% TFA, peptides were eluted with two times 80 µL 0.1% formic acid in 60% acetonitrile and dried in a vacuum centrifuge.

Peptides were reconstituted in 0.1% formic acid (solvent A) and analyzed with a Dionex Ultimate 3000 UPLC (Thermo Scientific) coupled via a nanospray source to a Q-Exactive HFX tandem MS (Thermo Scientific) operated in positive ion mode. Approximately 1 µg of peptides were desalted on a trap column (Acclaim PepMap100 C18, 5 µm, 100 Å, 300 µm i.d. × 5 mm, Thermo Scientific) and then separated on an in-house packed column (75 µm i.d. x 25 cm fused silica capillary packed with 3 µm ReproSil-Pur C18 beads (Dr. Maisch)). A 60 min gradient was used with a flow rate of 400 µL mL^{-1}. Solvent B (0.1% formic acid in 80% ACN) was increased from 4% to 21% over 40 min, followed by an increase to 40% over 20 min. Afterwards, to wash and re-equilibrate the column, solvent B was increased to 90% over 2 min, kept constant at 90% for 6 min and then decreased to 4% over 2 min with subsequent constant flow at 4% for at least 10 min.

Data-dependent acquisition was performed, acquiring MS1 spectra over a range of 400 to 2000 *m/z* with a resolution of 60,000, an AGC target of 1e6, and a maximum injection time of 100 ms. The top 20 peaks were selected for HCD fragmentation with stepped normalized collision energy (25, 30, and 35 NCE). MS2 were acquired with a fixed first mass (130 *m/z*), a resolution of 15,000, an AGC target of 1e5, and a maximum injection time of 200 ms. A dynamic exclusion list (10 s) was used and charge states 1 and >6 were excluded.

Bioinformatic analysis of MS data. The analysis of MS data followed the general workflow established by the ArcPP (Schulze et al., 2020), using the Python framework Ursgal (Kremer et al., 2016), with modifications as follows. Briefly, MS raw files were converted into mzML files using ThermoRawFileParser (Hulstaert et al., 2020), followed by conversion into MGF format by pymzML (Kösters et al., 2018). Protein database searches were performed against the ArcPP *Hfx. volcanii* theoretical proteome (4074 proteins, https://doi.org/10.5281/zenodo.3565631), supplemented with common contaminants, and decoys using shuffled peptide sequences. MSFragger (version 2.3) (Kong et al., 2017), X!Tandem (version vengeance) (Craig and Beavis, 2004) and MS-GF+ (version 2019.07.03) (Kim and Pevzner, 2014) were employed as search engines. Search parameters for the reanalysis of ArcPP datasets matched those described previously (Schulze et al., 2020). For the dataset PXD021874 (generated for this study), a precursor mass tolerance of +/-10 ppm, and a fragment mass tolerance of 10 ppm were used, two and three missed cleavages were allowed for data derived from samples digested with Trypsin and GluC, respectively, and oxidation of M, N-terminal acetylation were included as potential modifications, while carbamidomethylation of C was required as fixed modification. In addition, for all datasets, all known compositions of AglB- and Agl15-dependent *N*-glycans have been added as potential modifications of N (Table 4). Hex(2) was included as a potential modification of S or T for the search of *O*-glycopeptides. Since most protein database search engines do not support the search for multiple different modifications of the same amino acid, each *N*-glycan composition was searched for separately. Unfortunately, this means that we could not assess whether peptides containing two *N*-glycosites could be modified with one AglB- and one Agl15-dependent.

Results from all cellular fractions of one sample were merged before statistical post-processing with Percolator (version 3.4) (The et al., 2016). Subsequently, results from different search engines and from all searches for different glycan compositions were combined using the combined PEP approach as described before (Kremer et al., 2016; Schulze et al., 2020). After removal of PSMs with a combined PEP > 1%, results were sanitized, i.e. for spectra with multiple differing PSMs, only the PSM with the best combined PEP was accepted. Peptide and protein FDRs were calculated as described previously (Schulze et al., 2020), while treating each peptidoform (peptide + modification) separately and only peptidoforms with an FDR ≤ 1% and proteins with an FDR ≤ 0.5% were accepted.

Spectra corresponding to glycopeptide spectrum matches were searched for glycopeptide-specific fragment ions using the glycopeptide_fragmentor node in Ursgal. The fragment ions mass tolerance was set to 20 ppm and at least one B-ion and two Y-ions were required to be found in at least one spectrum of each glycopeptide. Finally, glycopeptides were required to be identified in at least two replicates.

Table 4. Glycans included as potential modifications in protein database searches. Given are the glycan composition (Hex, hexose; HexA, hexuronic acid; MeHexA, methyl-hexuronic acid; SO3Hex, sulfated hexose; dHex, deoxyhexose), the chemical composition, the Unimod ID and the amino acid that has been specified as potential site of modification.

Amino acid	Glycan composition	Chemical composition	Unimod ID
N	Hex(1)	$C_6H_{10}O_5$	41
N	Hex(1)HexA(1)	$C_6H_8O_6$	1427
N	Hex(1)HexA(2)	$C_{18}H_{26}O_{17}$	-
N	Hex(1)HexA(2)MeHexA(1)	$C_{25}H_{36}O_{23}$	-
N	Hex(2)HexA(2)MeHexA(1)	$C_{31}H_{46}O_{28}$	-
N	SO3Hex(1)	$C_6H_{10}O_8S_1$	-
N	SO3Hex(1)Hex(1)	$C_{12}H_{20}O_{13}S_1$	-
N	SO3Hex(1)Hex(2)	$C_{18}H_{30}O_{18}S_1$	-
N	SO3Hex(1)Hex(2)dHex(1)	$C_{24}H_{40}O_{22}S_1$	-
S/T	Hex(2)	$C_{12}H_{20}O_{10}$	512

Prediction of protein subcellular localization and protein domains. Secretion signals and the probable subcellular localization were determined as previously described (Schulze et al., 2020). Candidates for ArtA processing were taken from (Haft et al., 2012). Further protein domains were predicted with InterPro (Blum et al., 2021) and arCOGlet categories for each protein were obtained based on the latest release (December 2015) (Makarova et al., 2015).

Phylogenetic analyses. For analysis of phylogenetic distributions, we used OrthoDB v10.1 (Zdobnov et al., 2020). The locus tag (e.g. HVO_2072 for SLG) was used to interrogate OrthoDB and groups were extracted at the taxonomic levels Halobacteria, Euryarchaeota, and Archaea (Oct-2020). At each level, OrthoDB lists the number of species having an ortholog and the number of species analyzed at that level. For each group and level, a FASTA file is provided and the FASTA files at levels Halobacteria and Archaea were downloaded for a more detailed, script-based analysis. At the level Halobacteria, the analysis was performed at the species level. Four species were represented by two strains, and only one of them was considered, leaving a total of 161 species at the level Halobacteria. For the domain Archaea, the analysis

was performed at the genus level. OrthoDB codes consist of a species identifier (the NCBI TaxID with a strain serial number) and a protein identifier. For transformation of TaxID to species/genus, the NCBI taxonomy was inquired (https://www.ncbi.nlm.nih.gov/taxonomy) (accessed 22-Oct-2020).

For gene synteny analysis, the SyntTax server was used (Oberto, 2013) (accessed Oct-2020). At the time of usage, the server analyzed 94 genomes from the taxonomic class Halobacteria. It should be noted that only the main chromosomes are analyzed within the SyntTax server.

Data availability

MS raw files of the dataset generated for this study have been uploaded to the ProteomeXchange Consortium via the PRIDE partner repository (Perez-Riverol et al., 2019) with the data set identifier PXD021874. Protein database search result files have been added to the ArcPP result repository (https://doi.org/10.5281/zenodo.3724742) while analysis scripts have been included in the ArcPP GitHub repository (https://github.com/arcpp/ArcPP).

Acknowledgments

Conflict of interest

The authors declare no conflict of interest.

Supplemental Material

Figure S1. MS2 spectra of SLG *N*-glycosites N274 and N279 strongly support modification by AglB- as well as Agl15-dependent *N*-glycans. Annotated spectra for SLG *N*-glycopeptides comprising the peptide sequence VGIANSSATNTSGSSTGPTVE with AglB- (**A**) and Agl15-dependent (**B**) *N*-glycans attached to the *N*-glycosites N274 and N279. Measured raw peaks are shown in grey, annotated a- and b-ions in purple, y-ions in yellow and *N*-glycopeptide-specific Y- and B-ions in cyan. Insets illustrate the peptide sequence coverage through a- or b-ions (purple) and y-ions (yellow) (in both cases detected ions shown as wide bar, missing ions shown as line), as well as the coverage of Y- and B-ions (detected ions shown in cyan).

Figure S2. More proteins and peptides have been identified in PXD021874 than in any other dataset of the ArcPP. The number of identified peptides (light grey) and proteins (dark grey) for each dataset is shown as a barplot (sorted by the total number of identified proteins).

Figure S3. Agl15-dependent *N*-glycosylation of Ths3 is strongly supported by MS2 fragment ion series. Shown is an annotated spectrum for the Ths3 *N*-glycopeptide with sequence KSEVDTEYNITSVDQLTAAIDAEDSELR harboring an Agl15-dependent *N*-glycan. Measured raw peaks are shown in grey, annotated a- and b-ions in purple, y-ions in yellow and *N*-glycopeptide-specific Y- and B-ions in cyan. Insets illustrate the peptide sequence coverage through a- or b-ions (purple, detected ions shown as wide bar, missing ions shown as line) and y-ions (yellow), as well as the coverage of Y- and B-ions (detected ions shown in cyan).

Figure S4. Distribution of *N*-glycoproteins across Halobacteria. For each identified *N*-glycoprotein (x-dimension) the percentage of species within the class of Halobacteria that carry an orthologue is presented. Furthermore, for each species (y-dimension), the percentage of *N*-glycoproteins for which an orthologue is present within the species' genome is shown. Percentages are shown as heat, ranging from yellow (100%) to purple (0%).

Figure S5. *N*-glycopeptides with non-canonical *N*-glycosites are strongly supported by MS2 fragment ion series. Annotated spectra for non-canonical *N*-glycosites within the peptide sequences VVWTSESGSNSATLQR (**A**) and MPSNANIMGVTPGSR (**B**) corresponding to the protein pilA6 and HVO_2160, respectively, are shown. Both *N*-glycopeptides are modified by an AglB-dependent glycan and the site of attachment is indicated. Measured raw peaks are shown in grey, annotated a- and b-ions in purple, y-ions in yellow and *N*-glycopeptide-specific Y- and B-ions in cyan. Insets illustrate the peptide sequence coverage through a- or b-ions (purple, detected ions shown as wide bar, missing ions shown as line) and y-ions (yellow), as well as the coverage of Y- and B-ions (detected ions shown in cyan).

Table S1. Analysis of phylogeny, gene synteny and protein domains of identified *N*-glycoproteins. For each protein that was identified to be *N*-glycosylated, the HVO ID, description, InterPro domains and arCOGlet classification is given together results of phylogenetic and gene synteny analysis for the genes encoding the identified *N*-glycoproteins.

References

Abdul Halim MF, Karch KR, Zhou Y, Haft DH, Garcia BA, Pohlschroder M. 2016. Permuting the PGF Signature Motif Blocks both Archaeosortase-Dependent C-Terminal Cleavage and Prenyl Lipid Attachment for the Haloferax volcanii S-Layer Glycoprotein. *Journal of Bacteriology* **198**:808–815. doi:10.1128/JB.00849-15

Abdul Halim MF, Pfeiffer F, Zou J, Frisch A, Haft D, Wu S, Tolić N, Brewer H, Payne SH, Paša-Tolić L, Pohlschroder M. 2013. Haloferax volcanii archaeosortase is required for motility, mating, and C-terminal processing of the S-layer glycoprotein. *Mol Microbiol* **88**:1164–1175. doi:10.1111/mmi.12248

Abdul-Halim MF, Schulze S, DiLucido A, Pfeiffer F, Filho AWB, Pohlschroder M. 2020. Lipid Anchoring of Archaeosortase Substrates and Midcell Growth in Haloarchaea. *mBio* **11**. doi:10.1128/mBio.00349-20

Abu-Qarn M, Eichler J. 2006. Protein N-glycosylation in Archaea: defining Haloferax volcanii genes involved in S-layer glycoprotein glycosylation. *Molecular Microbiology* **61**:511–525. doi:https://doi.org/10.1111/j.1365-2958.2006.05252.x

Abu-Qarn M, Yurist-Doutsch S, Giordano A, Trauner A, Morris HR, Hitchen P, Medalia O, Dell A, Eichler J. 2007. Haloferax volcanii AglB and AglD are Involved in N-glycosylation of the S-layer Glycoprotein and Proper Assembly of the Surface Layer. *Journal of Molecular Biology* **374**:1224–1236. doi:10.1016/j.jmb.2007.10.042

Aebi M. 2013. N-linked protein glycosylation in the ER. *Biochim Biophys Acta* **1833**:2430–2437. doi:10.1016/j.bbamcr.2013.04.001

Apweiler R, Hermjakob H, Sharon N. 1999. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**:4–8. doi:10.1016/s0304-4165(99)00165-8

Bai L, Wang T, Zhao G, Kovach A, Li H. 2018. The atomic structure of a eukaryotic oligosaccharyltransferase complex. *Nature* **555**:328–333. doi:10.1038/nature25755

Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, Richardson L,

Salazar GA, Williams L, Bork P, Bridge A, Gough J, Haft DH, Letunic I, Marchler-Bauer A, Mi H, Natale DA, Necci M, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A, Finn RD. 2021. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research* **49**:D344–D354. doi:10.1093/nar/gkaa977

Choudhary P, Nagar R, Singh V, Bhat AH, Sharma Y, Rao A. 2019. ProGlycProt V2.0, a repository of experimentally validated glycoproteins and protein glycosyltransferases of prokaryotes. *Glycobiology* **29**:461–468. doi:10.1093/glycob/cwz013

Craig R, Beavis RC. 2004. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**:1466–1467. doi:10.1093/bioinformatics/bth092

Domon B, Costello CE. 1988. Structure elucidation of glycosphingolipids and gangliosides using high-performance tandem mass spectrometry. *Biochemistry* **27**:1534–1543. doi:10.1021/bi00405a021

Duggin IG, Aylett CHS, Walsh JC, Michie KA, Wang Q, Turnbull L, Dawson EM, Harry EJ, Whitchurch CB, Amos LA, Löwe J. 2015. CetZ tubulin-like proteins control archaeal cell shape. *Nature* **519**:362–365. doi:10.1038/nature13983

Eichler J, Arbiv A, Cohen-Rosenzweig C, Kaminski L, Kandiba L, Konrad Z. 2013. N-glycosylation in Haloferax volcanii: adjusting the sweetness. *Front Microbiol* **4**. doi:10.3389/fmicb.2013.00403

Esquivel RN, Schulze S, Xu R, Hippler M, Pohlschroder M. 2016. Identification of Haloferax volcanii Pilin N-Glycans with Diverse Roles in Pilus Biosynthesis, Adhesion, and Microcolony Formation. *J Biol Chem* **291**:10602–10614. doi:10.1074/jbc.M115.693556

Francoleon DR, Boontheung P, Yang Y, Kim U, Ytterberg AJ, Denny PA, Denny PC, Loo JA, Gunsalus RP, Ogorzalek Loo RR. 2009. S-layer, Surface-Accessible, and Concanavalin A Binding Proteins of Methanosarcina acetivorans and Methanosarcina mazei *J Proteome Res* **8**:1972–1982. doi:10.1021/pr800923e

Freeze HH. 2013. Understanding human glycosylation disorders: biochemistry leads the charge. *J Biol Chem* **288**:6936–6945. doi:10.1074/jbc.R112.429274

Guan Z, Delago A, Nußbaum P, Meyer B, Albers S-V, Eichler J. 2016. N-glycosylation in the thermoacidophilic archaeon Sulfolobus acidocaldarius involves a short dolichol pyrophosphate carrier. *FEBS Lett* **590**:3168–3178. doi:10.1002/1873-3468.12341

Guan Z, Naparstek S, Calo D, Eichler J. 2012. Protein glycosylation as an adaptive response in Archaea: growth at different salt concentrations leads to alterations in Haloferax volcanii S-layer glycoprotein N-glycosylation. *Environ Microbiol* **14**:743–753. doi:10.1111/j.1462-2920.2011.02625.x

Haft DH, Payne SH, Selengut JD. 2012. Archaeosortases and Exosortases Are Widely Distributed Systems Linking Membrane Transit with Posttranslational Modification. *Journal of Bacteriology* **194**:36–48. doi:10.1128/JB.06026-11

Halim A, Anonsen JH. 2017. Microbial glycoproteomics. *Current Opinion in Structural Biology*, Carbohydrates: A feast of structural glycobiology • Sequences and topology: Computational studies of protein-protein interactions **44**:143–150. doi:10.1016/j.sbi.2017.03.005

Harvey H, Bondy-Denomy J, Marquis H, Sztanko KM, Davidson AR, Burrows LL. 2018. Pseudomonas aeruginosa defends against phages through type IV pilus glycosylation. *Nat Microbiol* **3**:47–52. doi:10.1038/s41564-017-0061-y

Hulstaert N, Shofstahl J, Sachsenberg T, Walzer M, Barsnes H, Martens L, Perez-Riverol Y. 2020. ThermoRawFileParser: Modular, Scalable, and Cross-Platform RAW File Conversion. *J Proteome Res* **19**:537–542. doi:10.1021/acs.jproteome.9b00328

Jarrell KF, Ding Y, Meyer BH, Albers S-V, Kaminski L, Eichler J. 2014. N-linked glycosylation in Archaea: a structural, functional, and genetic analysis. *Microbiol Mol Biol Rev* **78**:304–341. doi:10.1128/MMBR.00052-13

Jevtić Ž, Stoll B, Pfeiffer F, Sharma K, Urlaub H, Marchfelder A, Lenz C. 2019. The Response of Haloferax volcanii to Salt and Temperature Stress: A Proteome Study by Label-Free Mass Spectrometry. *PROTEOMICS* **19**:1800491. doi:https://doi.org/10.1002/pmic.201800491

Kaminski L, Guan Z, Yurist-Doutsch S, Eichler J. 2013. Two Distinct N-Glycosylation Pathways Process the Haloferax volcanii S-Layer Glycoprotein upon Changes in Environmental Salinity. *mBio* **4**. doi:10.1128/mBio.00716-13

Kandiba L, Lin C-W, Aebi M, Eichler J, Guerardel Y. 2016. Structural characterization of the N-linked pentasaccharide decorating glycoproteins of the halophilic archaeon Haloferax volcanii. *Glycobiology* **26**:745–756. doi:10.1093/glycob/cww014

Kelly JF, Vinogradov E, Stupak J, Robotham AC, Logan SM, Berezuk A, Khursigara CM, Jarrell KF. 2020. Identification of a novel N-linked glycan on the archaellins and S-layer protein of the thermophilic methanogen, Methanothermococcus thermolithotrophicus. *J Biol Chem* **295**:14618–14629. doi:10.1074/jbc.RA120.012790

Kim S, Pevzner PA. 2014. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nature Communications* **5**:5277. doi:10.1038/ncomms6277

Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI. 2017. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. *Nature Methods* **14**:513–520. doi:10.1038/nmeth.4256

Kösters M, Leufken J, Schulze S, Sugimoto K, Klein J, Zahedi RP, Hippler M, Leidel SA, Fufezan C. 2018. pymzML v2.0: introducing a highly compressed and seekable gzip format. *Bioinformatics* **34**:2513–2514. doi:10.1093/bioinformatics/bty046

Kremer LPM, Leufken J, Oyunchimeg P, Schulze S, Fufezan C. 2016. Ursgal, Universal Python Module Combining Common Bottom-Up Proteomics Tools for Large-Scale Analysis. *J Proteome Res* **15**:788–794. doi:10.1021/acs.jproteome.5b00860

Li G, Yan Q, Nita-Lazar A, Haltiwanger RS, Lennarz WJ. 2005. Studies on the N-Glycosylation of the Subunits of Oligosaccharyl Transferase in Saccharomyces cerevisiae. *J Biol Chem* **280**:1864–1871. doi:10.1074/jbc.M410969200

Li Z, Kinosita Y, Rodriguez-Franco M, Nußbaum P, Braun F, Delpech F, Quax TEF, Albers S-V. 2019. Positioning of the Motility Machinery in Halophilic Archaea. *mBio* **10**. doi:10.1128/mBio.00377-19

Makarova KS, Wolf YI, Koonin EV. 2015. Archaeal Clusters of Orthologous Genes (arCOGs): An Update and Application for Analysis of Shared Features between Thermococcales, Methanococcales, and Methanobacteriales. *Life (Basel)* **5**:818–840. doi:10.3390/life5010818

Makkay AM, Louyakis AS, Ram-Mohan N, Gophna U, Gogarten JP, Papke RT. 2020. Insights into gene expression changes under conditions that facilitate horizontal gene transfer (mating) of a model archaeon. *Sci Rep* **10**:22297. doi:10.1038/s41598-020-79296-w

Mescher MF, Strominger JL. 1976. Purification and characterization of a prokaryotic glycoprotein from the cell envelope of Halobacterium salinarium. *J Biol Chem* **251**:2005–2014.

Meyer BH, Albers S-V. 2013. Hot and sweet: protein glycosylation in Crenarchaeota. *Biochem Soc Trans* **41**:384–392. doi:10.1042/BST20120296

15

Nikolayev S, Cohen-Rosenzweig C, Eichler J. 2020. Evolutionary considerations of the oligosaccharyltransferase AglB and other aspects of N-glycosylation across Archaea. *Molecular Phylogenetics and Evolution* **153**:106951. doi:10.1016/j.ympev.2020.106951

Nothaft H, Szymanski CM. 2010. Protein glycosylation in bacteria: sweeter than ever. *Nature Reviews Microbiology* **8**:765–778. doi:10.1038/nrmicro2383

Oberto J. 2013. SyntTax: a web server linking synteny to prokaryotic taxonomy. *BMC Bioinformatics* **14**:4. doi:10.1186/1471-2105-14-4

Palmieri G, Balestrieri M, Peter-Katalinić J, Pohlentz G, Rossi M, Fiume I, Pocsfalvi G. 2013. Surface-exposed glycoproteins of hyperthermophilic Sulfolobus solfataricus P2 show a common N-glycosylation profile. *J Proteome Res* **12**:2779–2790. doi:10.1021/pr400123z

Parente J, Casabuono A, Ferrari MC, Paggi RA, De Castro RE, Couto AS, Giménez MI. 2014. A Rhomboid Protease Gene Deletion Affects a Novel Oligosaccharide N-Linked to the S-layer Glycoprotein of Haloferax volcanii. *J Biol Chem* **289**:11304–11317. doi:10.1074/jbc.M113.546531

Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, Pérez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yılmaz Ş, Tiwary S, Cox J, Audain E, Walzer M, Jarnuczak AF, Ternent T, Brazma A, Vizcaíno JA. 2019. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Research* **47**:D442–D450. doi:10.1093/nar/gky1106

Pohlschroder M, Pfeiffer F, Schulze S, Halim MFA. 2018. Archaeal cell surface biogenesis. *FEMS Microbiol Rev* **42**:694–717. doi:10.1093/femsre/fuy027

Rappsilber J, Mann M, Ishihama Y. 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**:1896–1906. doi:10.1038/nprot.2007.261

Schäffer C, Messner P. 2017. Emerging facets of prokaryotic glycosylation. *FEMS Microbiol Rev* **41**:49–91. doi:10.1093/femsre/fuw036

Schulze S, Adams Z, Cerletti M, De Castro R, Ferreira-Cerca S, Fufezan C, Giménez MI, Hippler M, Jevtic Z, Knüppel R, Legerme G, Lenz C, Marchfelder A, Maupin-Furlow J, Paggi RA, Pfeiffer F, Poetsch A, Urlaub H, Pohlschroder M. 2020. The Archaeal Proteome Project advances knowledge about archaeal cell biology through comprehensive proteomics. *Nature Communications* **11**:3145. doi:10.1038/s41467-020-16784-7

Shalev Y, Soucy SM, Papke RT, Gogarten JP, Eichler J, Gophna U. 2018. Comparative Analysis of Surface Layer Glycoproteins and Genes Involved in Protein Glycosylation in the Genus Haloferax. *Genes (Basel)* **9**. doi:10.3390/genes9030172

Shalev Y, Turgeman-Grott I, Tamir A, Eichler J, Gophna U. 2017. Cell Surface Glycosylation Is Required for Efficient Mating of Haloferax volcanii. *Front Microbiol* **8**. doi:10.3389/fmicb.2017.01253

Silva RT de, Abdul-Halim MF, Pittrich DA, Brown HJ, Pohlschroder M, Duggin IG. 2020. Improved growth and morphological plasticity of Haloferax volcanii. *bioRxiv* 2020.05.04.078048. doi:10.1101/2020.05.04.078048

Sumper M, Berg E, Mengele R, Strobel I. 1990. Primary structure and glycosylation of the S-layer protein of Haloferax volcanii. *J Bacteriol* **172**:7111–7118. doi:10.1128/jb.172.12.7111-7118.1990

Tamir A, Eichler J. 2017. N-Glycosylation Is Important for Proper Haloferax volcanii S-Layer Stability and Function. *Appl Environ Microbiol* **83**. doi:10.1128/AEM.03152-16

The M, MacCoss MJ, Noble WS, Käll L. 2016. Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. *J Am Soc Mass Spectrom* **27**:1719–1727. doi:10.1007/s13361-016-1460-7

Trent JD, Kagawa HK, Paavola CD, McMillan RA, Howard J, Jahnke L, Lavin C, Embaye T, Henze CE. 2003. Intracellular localization of a group II chaperonin indicates a membrane-related function. *PNAS* **100**:15589–15594. doi:10.1073/pnas.2136795100

Tripepi M, Imam S, Pohlschröder M. 2010. Haloferax volcanii Flagella Are Required for Motility but Are Not Involved in PibD-Dependent Surface Adhesion. *Journal of Bacteriology* **192**:3093–3102. doi:10.1128/JB.00133-10

Tripepi M, You J, Temel S, Önder Ö, Brisson D, Pohlschröder M. 2012. N-glycosylation of Haloferax volcanii flagellins requires known Agl proteins and is essential for biosynthesis of stable flagella. *J Bacteriol* **194**:4876–4887. doi:10.1128/JB.00731-12

Valguarnera E, Kinsella RL, Feldman MF. 2016. Sugar and Spice Make Bacteria Not Nice: Protein Glycosylation and Its Influence in Pathogenesis. *Journal of Molecular Biology*, Glycosylation in Cellular Processes and Disease **428**:3206–3220. doi:10.1016/j.jmb.2016.04.013

Varki A. 2017. Biological roles of glycans. *Glycobiology* **27**:3–49. doi:10.1093/glycob/cww086

Vinogradov E, Deschatelets L, Lamoureux M, Patel GB, Tremblay T-L, Robotham A, Goneau M-F, Cummings-Lorbetskie C, Watson DC, Brisson J-R, Kelly JF, Gilbert M. 2012. Cell surface glycoproteins from Thermoplasma acidophilum are modified with an N-linked glycan containing 6-C-sulfofucose. *Glycobiology* **22**:1256–1267. doi:10.1093/glycob/cws094

Wild R, Kowal J, Eyring J, Ngwa EM, Aebi M, Locher KP. 2018. Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science* **359**:545–550. doi:10.1126/science.aar5140

Zdobnov EM, Kuznetsov D, Tegenfeldt F, Manni M, Berkeley M, Kriventseva EV. 2020. OrthoDB in 2020: evolutionary and functional annotations of orthologs. *Nucleic Acids Res*. doi:10.1093/nar/gkaa1009

Zeitler R, Hochmuth E, Deutzmann R, Sumper M. 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**:1149–1155. doi:10.1093/glycob/8.12.1157