REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
15-01-2010	Final Report	01-01-2007-30-11-2009

4. TITLE AND SUBTITLE

Protein glycosylation in Archaea: A post-translational modification to enhance extremophilic protein stability

5a. CONTRACT NUMBER
FA9550-07-10057

5b. GRANT NUMBER
FA9550-07-10057

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dr. Jerry Eichler

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
AFOSR/NA

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Ben Gurion University of the Negev
POB 653
Beersheva 84105, ISRAEL

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR
875 N RANDOLPH ST
STE 3112
ARLINGTON, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
USAF, AFRL

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-SR-AR-TR-10-0091

12. DISTRIBUTION/AVAILABILITY STATEMENT
A

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Post-translational modifications account for much of the biological diversity generated at the proteome level. Of these, glycosylation is the most prevalent. Long-thought to be unique to Eukarya, it is now clear that both Bacteria and Archaea are also capable of N-glycosylation, namely the covalent linkage of oligosaccharides to select target asparagine residues. However, little had been known of this process in Archaea. As such, this project aimed at defining the N-glycosylation pathway of the halophilic archaeon Haloferax volcanii. Employing a combination of bioinformatics, genetic, biochemical and structural approaches, the funded research succeeded in identifying a group of clustered H. volcanii genes (the agl genes) encoding proteins involved in the assembly and attachment of a pentasaccharide to select asparagine residues of the S-layer glycoprotein, a reporter of N-glycosylation in this species. In addition to providing insight into N-glycosylation across evolution and the biology of extremophiles, these findings could be exploited to create archaeal strains expressing selected N-glycosylation enzyme modules. This would allow for a hamessing of the greater diversity associated with this post-translational modification in Archaea in the design of tailor-made glycoproteins.

15. SUBJECT TERMS

Extremophiles, Archaea, N-glycosylation, Post-translational modifications, Halophiles, Haloferax volcanii

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Daphna Tripto
			UU	26	19b. TELEPHONE NUMBER *(Include area code)* (+972) 8647-2425

AFSOR award FA9550-07-10057 - "Protein glycosylation in Archaea: A post-translational modification to enhance extremophilic protein stability"

Jerry Eichler, Ph.D., Dept. of Life Sciences, Ben Gurion University of the Negev, Beersheva, Israel

FINAL REPORT

Background

Long-thought to be an exclusively eukaryal trait, it is now clear that prokaryotes are also capable of attaching glycan moieties to selected asparagine residues i.e. performing N-glycosylation (1-3). In contrast to Bacteria, where N-glycosylation is still considered a rare event, numerous species have contributed to the long list of proteins containing Asn-linked glycan moieties in Archaea (4). Able to thrive in amongst the harshest environmental conditions on the planet, Archaea express proteins that remain properly folded and functional in the face of extremes of salinity, temperature, and other adverse physical conditions that would normally lead to protein denaturation, loss of solubility and aggregation (5,6). In many instances, glycosylation serves to maintain the stability of archaeal proteins in the drastic conditions in which they exist. For example, glycosylation enhances protein solubility in hypersaline conditions (7), offers protection to proteins exposed to highly acidic surroundings (8) and has been implicated in thermoarchaeal protein stabilization (9). Furthermore, it is likely that glycosylation in Archaea also assists in protein folding and mediates protein-protein interactions, as is the case in non-extremophilic organisms (10,11). Thus, a better understanding of archaeal protein glycosylation will offer novel insight into the molecular strategies adopted by extremophile-derived polypeptides in overcoming the challenges of their environments. While clearly important for addressing protein science from an evolutionary perspective, such information also carries enormous applied potential.

Analysis of the chemical composition of the Asn-linked polysaccharides decorating many archaeal proteins has revealed the use of a wider variety of sugar subunits than seen in either eukaryal or bacterial glycoproteins (4). Still, little is known of the steps or components involved in this post-translational modification in Archaea. Deciphering the archaeal N-glycosylation process will, therefore, allow for a harnessing of the agents responsible for the unequaled glycan diversity observed in archaeal glycosylated proteins for the design of tailor-made glycoproteins, optimized for stability in the face of different environmental challenges.

The limited available information on the archaeal N-glycosylation process reveals similarities to the parallel eukaryal system. *Halobacterium salinarum* and *Haloferax volcanii* membranes have been shown to contain phospho- and pyrophosphodolichol, reminiscent of the eukaryal glycan-charged lipid, linked to a variety of monosaccharides, including glucose, mannose, and N-acetylglucosamine (GlcNAc), as well as a tetrasaccharide moiety containing mannose, galactose, and rhamnose (12,13). In addition, archaeal homologues of selected eukaryal enzymes involved in N-glycosylation have been described. For example, a GlcNAc transferase in *H. salinarum* has been partially characterized (12), as was a dolichylphosphate mannose synthase purified from *Thermoplasma acidophilum* (14). In addition, a putative

dolichylphosphate glucose synthase was identified in *H. volcanii* homogenates (15). Antibiotics that block the eukaryal N-glycosylation process have also been shown to interfere with protein glycosylation in Archaea (15-18). Moreover, archaeal protein glycosylation has been localized to the external surface of the plasma membrane (19), the topological homologue of the lumen-facing leaflet of the ER membrane, the site of the later stages of eukaryal N-glycosylation. Accordingly, membranes from *Sulfolobus acidocaldarius* contain an externally-oriented pyrophosphatase, possibly involved in the hydrolysis and recycling of dolicholpyrophosphate (20).

Despite these similarities, numerous aspects of archaeal N-glycosylation find no parallel in Eukarya. First, while glycans are linked to Asn residues of eukaryal glycoproteins almost solely through GlcNAc, Archaea can also use other linking sugars (19). Secondly, the eukaryal N-glycosylation machinery modifies the Asn residue of an Asn-X-Ser/Thr motif, where X is any residue apart from Pro (10,21-23), essentially exclusively (24,25). In *H. salinarum*, however, it was reported that exchanging the Ser residue of an Asn-Ala-Ser motif within the S-layer glycoprotein sequence with a Val, Leu or Asn did not prevent N-glycosylation (26), suggesting that the workings of the archaeal oligosaccharide transferase differs from its eukaryal counterpart. In further distinction from the eukaryal N-glycosylation process, in which the lipid-linked oligosaccharide precursor is not chemically modified (27), transient and permanent modification of dolichol-linked glycans have been observed in *H. salinarum* and *Methanothermus fervidus*, respectively (28,29).

Thus, to better describe protein glycosylation across evolution as well as the contribution of protein glycosylation in enhancing archaeal protein stability and longevity in the face of extreme conditions, a detailed description of both the steps involved in the archaeal version of this post-translational modification as well as the target amino acid composition best preferred by the archaeal glycosylation machinery are required. In the funded research, work aimed at characterizing these aspects of protein glycosylation in Archaea was conducted, using the haloarchaea *Haloferax volcanii* as the model of study. With the availability of a completed genome sequence (30), advanced genetic, molecular biology and biochemical tools and techniques (31,32) and simple growth requirements, *H. volcanii* represents a strain of choice for molecular studies in Archaea. Based on our findings, as well as those of others, a detailed picture of the archaeal version of N-glycosylation is now emerging.

1. Schaffer, C., M. Graninger, and P. Messner. 2001. Proteomics 1:248-261.
2. Upreti, R.K., M. Kumar, and V. Shankar. 2003. Proteomics 3:363-379.
3. Szymanski, C.M. and B.W. Wren. 2005. Nat Rev Microbiol 3:225-237.
4. Eichler, J. and M.W.W. Adams. 2005. Microbiol Mol Biol Rev 69:393-425.
5. Rothschild, L. J. and R.L. Mancinelli. 2001. Nature 409:1092-1101.
6. Danson, M.J. and D.W. Hough. 1998. Trends Microbiol 6:307–314.
7. Mengele, R., and M. Sumper. 1992. J Biol Chem 267:8182-8185.
8. Zahringer, U., H. Moll, T. Hettmann, Y.A. Knirel, and G. Schafer. 2000. Eur J Biochem 267:4144-4149.
9. Albers, S.V., S.M. Koning, W.N. Konings, and A.J. Driessen. 2004. J Bioenerg Biomembr 36:5-15.
10. Helenius, A. and M. Aebi. 2004. Annu Rev Biochem 73:1019-1049.

11. Mitra, N., S. Sinha, T.N. Ramya, and A. Surolia. 2006. Trends Biochem Sci 31:156-163.
12. Mescher, M.F., U. Hansen, and J.L. Strominger. 1976. J Biol Chem 251:7289-7294.
13. Kuntz, C., J. Sonnenbichler, I. Sonnenbichler, M. Sumper, and R. Zeitler. 1997. Glycobiology 7:897-904.
14. Zhu, B.C. and R.A. Laine. 1996. Glycobiology 6:811-816.
15. Zhu, B.C., R.R. Drake, H. Schweingruber, and R.A. Laine. 1995. Arch Biochem Biophys 319:355-364.
16. Wieland, F., W. Dompert, G. Bernhardt, and M. Sumper. 1980. FEBS Lett 120:110-114.
17. Bayley, D.P., M.L. Kalmokoff, and K.F. Jarrell. 1993. Arch Microbiol 160:179–185.
18. Grogan, D.W. 1996. Can J Microbiol 42:1163–1171.
19. Lechner, J. and F. Wieland. 1989. Annu Rev Biochem 58:173-194.
20. Meyer, W. and G. Schafer. 1992. Eur J Biochem 207:741-746.
21. Kornfeld, R. and S. Kornfeld. 1985. Annu Rev Biochem 54:631-664.
22. Burda, P. and M. Aebi. 1999. Biochim Biophys Acta 1426:239-257.
23. Spiro, R. G. 2002. Glycobiology 12:43R-56R.
24. Kehry, M., C. Sibley, J. Schilling, and L. Hood. 1979. Proc Natl Acad Sci USA 76:2832-2936.
25. Singh, D.G., J. Lomako, W.M. Lomako, W.J. Whelan, H.E. Meyer, M. Serwe, and J.W. Metzger. 1995. FEBS Lett 376:61-64.
26. Zeitler, R., E. Hochmuth, R. Deutzmann, and M. Sumper. 1998. Glycobiology 8:1157-1164.
27. Varki, A. 1998. Trends Cell Biol 8:34-40.
28. Lechner, J., F. Wieland, and M. Sumper. 1985. J Biol Chem 260:860-866.
29. Kärcher, U., H. Schröder, E. Haslinger, G. Allmaier, R. Schreiner, F. Wieland, A. Haselbeck, and H. König. 1993. J Biol Chem 268:26821–26826.
30. www.tigr.org
31. www.microbiol.unimelb.edu.au/people/dyallsmith/HaloHandbook/
32. DasSarma, S. and Fleishmann, E.M., eds. 1995. Archaea: A laboratory manual – Halophiles, Cold Spring Harbor Laboratory Press, NY

Summary of Scientific Achievements

Analysis of AglB and AglD, components of the in *Haloferax volcanii* N-glycosylation pathway: The effects of deleting two genes previously implicated in *Haloferax volcanii* N-glycosylation on the assembly and attachment of a novel Asn-linked pentasaccharide decorating the *H. volcanii* S-layer glycoprotein were considered. Mass spectrometry revealed the pentasaccharide to comprise two hexoses, two hexuronic acids and an additional 190 Da saccharide. The absence of AglD prevented addition of the final hexose to the pentasaccharide, while cells lacking AglB were unable to N-glycosylate the S-layer glycoprotein. In AglD-lacking cells, the S-layer glycoprotein-based surface layer presented both an architecture and protease susceptibility different than the background strain. By contrast, the absence of AglB resulted in enhanced release of the S-layer glycoprotein. *H. volcanii* cells lacking these N-glycosylation genes, moreover, grew significantly less well at elevated salt levels than did cells of the background strain. Thus, these results offer

experimental evidence showing that N-glycosylation endows *H. volcanii* with an ability to maintain an intact and stable cell envelope in hypersaline surroundings, ensuring survival in this extreme environment.

AglE, a novel component of the *Haloferax volcanii* N-glycosylation pathway: By combining gene deletion together with mass spectrometry, AglE, originally identified as a homologue of murine Dpm1, was shown to play a role in the addition of the 190 Da sugar subunit of the novel pentasaccharide decorating the S-layer glycoprotein. Topological analysis of an AglE-based chimeric reporter assigns AglE as an integral membrane protein, with its N-terminus and putative active site facing the cytoplasm.

AglF, AglG and AglI are additional components of the *Haloferax volcanii* N-glycosylation pathway: As the next step in our efforts to identify components of the N-glycosylation pathway of the halophilic archaeon *Haloferax volcanii*, the involvement of three additional gene products in the biosynthesis of the pentasaccharide decorating the S-layer glycoprotein was demonstrated. The genes encoding AglF, AglI and AglG are found immediately upstream of the gene encoding the archaeal oligosaccharide transferase, AglB. Evidence showing that AglF and AglI are involved in the addition of the hexuronic acid found at position three of the pentasaccharide is provided, while AglG is shown to contribute to the addition of the hexuronic acid found at position two. Given their proximities in the *H. volcanii* genome, the transcription profiles of *aglF*, *aglI*, *aglG* and *aglB* were considered. While only *aglF* and *aglI* share a common promoter, transcription of the four genes is coordinated, as revealed by determining transcript levels in *H. volcanii* cells raised in different growth conditions. Such changes in N-glycosylation gene transcription levels offer additional support for the adaptive role of this post-translational modification in *H. volcanii*.

Assessing the topology of N-glycosylation in *Haloferax volcanii*: In Archaea, virturally nothing is known of the topology of enzymes involved in assembling those glycans subsequently N-linked to target proteins on the external surface of the cell. To remedy this situation, sub-cellular localization and topology predictive algorithms, protease accessibility and immunoblotting, together with cysteine modification following site-directed mutagenesis, were enlisted to define the topology of *Haloferax volcanii* proteins experimentally proven to participate in the N-glycosylation process. As such, AglJ and AglD, respectively involved in the earliest and latest stages of assembly of the pentasaccharide decorating the *H. volcanii* S-layer glycoprotein, were shown to present their soluble N-terminal domain, likely containing the putative catalytic site of each enzyme, to the cytosol. The same holds true for Alg5-B, Dpm1-A and Mpg1-D, proteins putatively involved in this post-translational event. The results thus point to the assembly of the pentasaccharide linked to select Asn residues of the *H. volcanii* S-layer glycoprotein as occurring within the cell.

Examination of the *Haloferax volcanii agl* gene cluster reveals novel genes involved in N-glycosylation: Our studies on the halophilic archaeon *Haloferax volcanii* have identified a gene cluster encoding the Agl proteins involved in the assembly and attachment of a pentasaccharide to select Asn residues of the surface (S)-layer glycoprotein in this species. However, because the automated tools used for rapid annotation of genome sequences, including that of *H. volcanii*, are not always accurate, a re-annotation of the *agl* cluster was undertaken in order to discover genes

not previously recognized. In the present report, re-analysis of the gene cluster that includes *aglB*, *aglE*, *aglF*, *aglG*, *aglI* and *aglJ*, known components of the *H. volcanii* protein N-glycosylation machinery, was undertaken. Using computer-based tools or visual inspection, together with transcriptional analysis and protein expression approaches, genes encoding AglP, AglQ and AglR are now described.

Cataloguing glycosyltransferases and oligosaccharyltransferases in Archaea: In this study, 56 available archaeal genomes were examined with the aim of identifying glycosyltransferases and oligosaccharyltransferases, including those putatively catalyzing this post-translational processing event. Such analysis reveals that while oligosaccharyltransferases, central components of the N-glycosylation pathway, are found across the range of archaeal phenotypes, the N-glycosylation machinery of hyperthermophilic archaea may well rely on fewer components than do the parallel systems of non-hyperthermophilic archaea. Moreover, genes encoding predicted glycosyltransferases of hyperthermophilic archaea tend to be far more scattered within the genome than is the case with non-hyperthermophilic species, where putative glycosyltransferase genes are often clustered around identified oligosaccharyltransferase-encoding sequences.

***Haloferax volcanii* AglF and AglM can function in a coordinated manner:** In this study, the *H. volcanii agl* gene cluster was expanded to include an additional sequence, *aglM*, shown to participate in the biosynthesis of hexuronic acids contained within a pentasaccharide decorating the S-layer glycoprotein, a reporter *H. volcanii* glycoprotein. In response to different growth conditions, changes in the transcription profile of *aglM* mirrored changes in the transcription profiles of *aglF*, *aglG* and *aglI*, genes encoding confirmed participants in the *H. volcanii* N-glycosylation pathway, thus offering support to the hypothesis that in *H. volcanii*, N-glycosylation serves an adaptive role. Following purification, biochemical analysis revealed AglM to function as a UDP-glucose dehydrogenase. In a coupled reaction with AglF, a previously identified glucose-1-phosphate uridyltransferase, UDP-glucuronic acid was generated from glucose-1-phosphate and UTP in a NAD^+-dependent manner. These experiments thus represent the first step toward in vitro reconstitution of the archaeal N-glycosylation process.

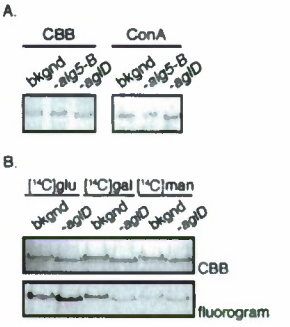
AglP is a S-adenosyl-L-methionine-dependent methyltransferase that participates in the N-glycosylation pathway of *Haloferax volcanii*: AglP, originally identified based on the proximity of its encoding gene to other *agl* genes whose products were shown to participate in N-glycosylation, was proposed, based on sequence homology, to serve as a methyltransferase. In the present report, gene deletion and mass spectrometry were employed to reveal that AglP is responsible for adding a 14 Da moiety to a hexuronic acid found at position four of the pentasaccharide decorating the *Hfx. volcanii* S-layer glycoprotein. Subsequent purification of a tagged version of AglP and development of an *in vitro* assay to test the function of the protein confirmed that AglP is a S-adenosyl-L-methionine-dependent methyltransferase.

A novel in vivo assay for the identification of residues important for the activity of *Haloferax volcanii* AglD, a component of the archaeal N-glycosylation pathway: In *Haloferax volcanii*, AglD adds the final hexose to the N-linked pentasaccharide decorating the S-layer glycoprotein. Not knowing the natural

substrate of the glycosyltransferase, together with the challenge of designing assays compatible with hypersalinity, have frustrated efforts at biochemical characterization of AglD activity. To circumvent these obstacles, an in vivo assay designed to identify amino acid residues important for AglD activity is described. In the assay, restoration of AglD function in a *Hfx. volcanii aglD* deletion strain transformed to express plasmid-encoded versions of AglD, generated through site-directed mutagenesis at positions encoding residues conserved in archaeal homologues of AglD, is reflected in the behavior of a readily detectable reporter of N-glycosylation. As such, Asp110, Asp112 were designated as elements of the DXD motif of AglD, a motif that interacts with metal cations associated with nucleotide-activated sugar donors, while Asp201 was assigned as the catalytic base of the enzyme.

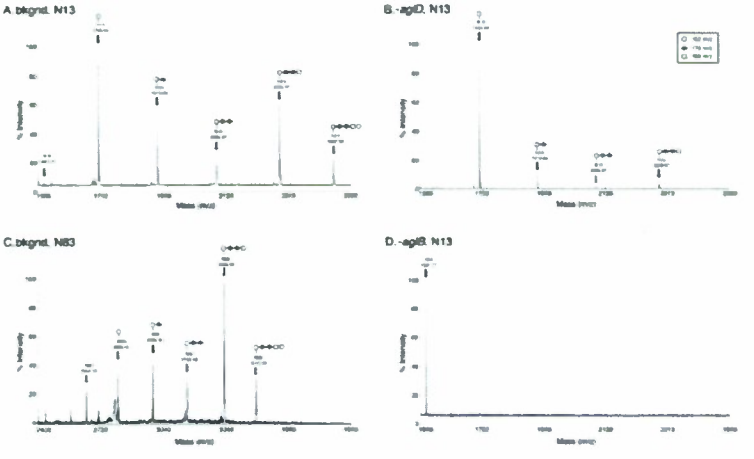
Detailed description of research findings

Analysis of AglB and AglD, components of the in *Haloferax volcanii* N-glycosylation pathway:

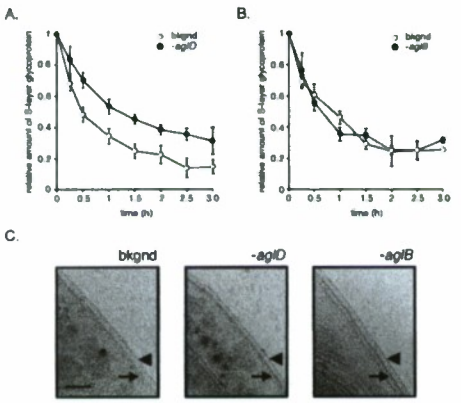


Glycosylation of the *H. volcanii* S-layer glycoprotein differs in the presence or absence of *aglD*. (a) The protein contents of *H. volcanii* WR536 background cells (bkgnd) as well as proteins of the same strain lacking either Alg5-B (*-alg5-B*) or AglD (*-aglD*) were separated by 7.5% SDS-PAGE and subjected to Coomassie staining (left panel) or overlay with HRP-conjugated ConA (right panel). (b) *H. volcanii* WR536 cells (bkgnd) or the same strain lacking AglD (*-aglD*) were grown in the presence of [^{14}C]-glucose, -galactose or -mannose and subjected to Coomassie staining (upper panel) or fluorography (lower panel). In both A and B, only the S-layer glycoprotein is shown.

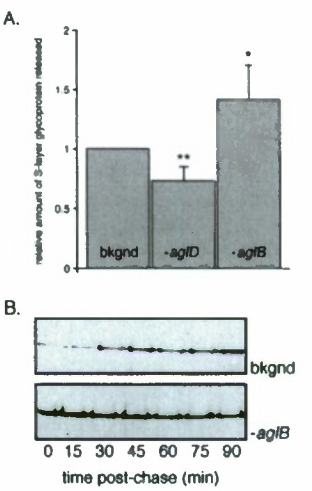
MALDI TOF analysis of *H. volcanii* S-layer glycoprotein-derived glycopeptides reveals the protein to be modified by a pentasaccharide requiring AglB and AglD. Shown are MALDI spectra obtained from the Asn-13-containing tryptic glycopeptides from (a) the background and (b) the *algD* deletion strains. (c) MALDI spectrum of the Asn-83-containing tryptic/Glu-C glycopeptide from the background strain. (d) MALDI spectrum of the unmodified Asn-13-containing peptide from the *aglB* deletion strain. The nature of the glycopeptide-associated sugar residues is shown in the inset of panel B. Tryptic peptides were separated by offline nanoLC prior to analysis.

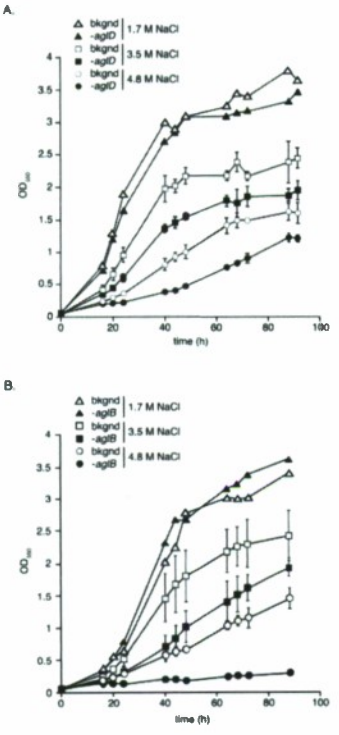


Perturbed S-layer glycoprotein N-glycosylation affects *H. volcanii* S-layer architecture. (a) Background strain WR536 cells (open circles) or AglD-lacking cells of the same strain (full circles) were challenged with 1 mg/ml trypsin at 37°C. Aliquots were removed immediately prior to incubation with trypsin and at subsequent 15-30 min intervals after addition of the protease and examined by 7.5% SDS-PAGE. The amount of S-layer glycoprotein remaining with time, relative to that amount found immediately prior to proteolysis, was densitometrically quantified. Each point represents the average of 3 experiments ± SEM. Background cells: open circles; AglD-lacking cells: full circles. (b) As in A, only AglB-lacking cells were examined instead of AglD-lacking cells. (c) Electron micrographs of cells envelopes. Left panel: *H. volcanii* WR536 cells (bkgnd); middle panel: the same cells lacking AglD (*-aglD*); right panel: the same cells lacking AglB (*-aglB*). In each panel, the arrow points to the plasma membrane while the arrowhead points to the outer edge of the surface layer. The bar in the left panel corresponds to 100 nm.



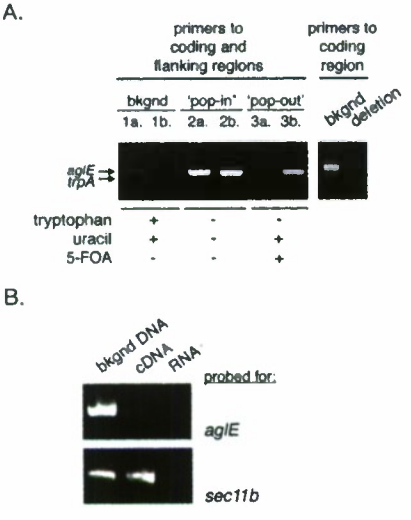
The S-layer glycoprotein is more readily released from *H. volcanii* cells lacking N-glycosylation capability. (a) The levels of S-layer glycoprotein released into the growth media from *H. volcanii* WR536 cells (bkgnd) or the same cells lacking either AglB (*-aglB*) or AglD (*-aglD*) after 10 h of growth were densitometrically quantified following TCA precipitation, SDS-PAGE and Coomassie staining. The amount of S-layer glycoprotein released from the background strain was set at a value of 1. Each bar represents the average of five separate growths for each cell type. Error bars correspond to standard deviation. The asterisk corresponds to a value being statistically significant to $p<0.01$, while the double asterisk corresponds to a statistical significance of $p<0.001$, as determined by student's t test. (b) Background *H. volcanii* cells or the same cells lacking AglB were pulse radiolabelled with [^{35}S] Cys/Met for 5 min and chased with excess Cys/Met. The growth medium of each culture was removed, the protein content precipitated, separated by 7.5% SDS-PAGE and visualized by fluorography. Upper panel: background (bkgnd) cells; lower panel: AglB-lacking (*-algB*) cells. Shown is a representative of four repeats.



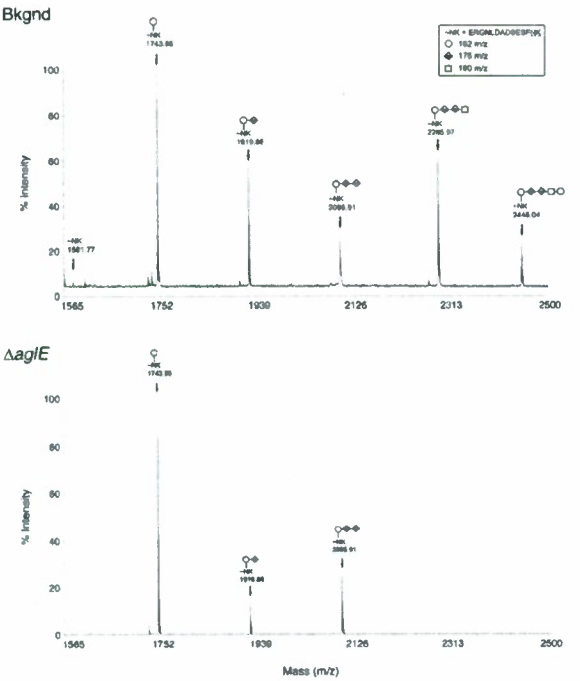


H. volcanii cells lacking AglB or AglD grow slower than the background strain at elevated NaCl concentrations. Background strain WR536 cells (open symbols) or AglB- or AglD-lacking cells of the same strain (full symbols) were grown in rich medium containing 1.75 M NaCl (triangles), 3.5 M NaCl (squares) or 4.8 M NaCl (circles). Growth was measured spectrophotometrically at an optical density of 550 nm. (a) Growth curves of background (bkgnd) and AglD-lacking (*-algD*) cells. (b) Growth curves of background (bkgnd) and AglB-lacking (*-algB*) cells. The growth curves of cells grown in medium containing 1.75 M NaCl represent the results of a single growth, while the growth curves of cells grown in medium containing 3.5 or 4.8 M NaCl represent the results of 2-3 growths ± standard deviation.

AglE, a novel component of the *Haloferax volcanii* N-glycosylation pathway:

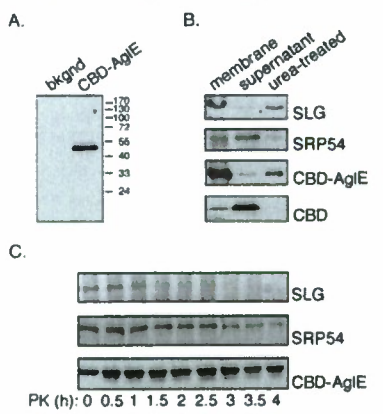


Deletion of the *dpm1-B* (*aglE)* gene does not affect cell viability. A. Left panel: PCR amplification was performed using a forward primer directed at the 5' *dpm1-B* flanking region and a reverse primer directed at a sequence within the *dpm1-B* coding region (lanes 1a, 2a and 3a) or a sequence within the *trpA* sequence (lanes 1b, 2b and 3b), together with genomic DNA from cells of the WR536 background strain (bkgnd, lanes 1a and 1b), from plasmid-incorporating cells ('pop-in', lanes 2a and 2b) or from cells that had replaced the *dpm1-B* gene ('pop-out', lanes 3a and 3b), as template. Left panel: PCR amplification was performed using primers directed against the *dpm1-B* coding region together with with genomic DNA from cells of the WR536 background strain (bkgnd) or the *dpm1-B*-deleted strain (deletion). B. RT-PCR was performed using primers directed at *dpm1-B* (upper panel) or *sec11b* (lower panel) and genomic DNA from WR536 background cells, or cDNA or RNA from *dpm1-B*-deleted cells as template.

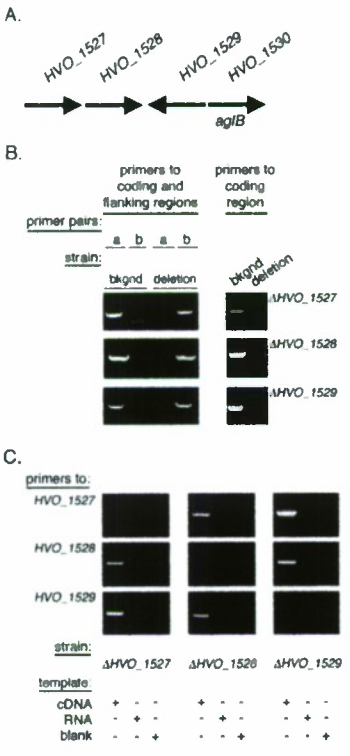


MALDI-TOF analysis of an Asn-13-containing, *H. volcanii* S-layer glycoprotein-derived glycopeptides shows a role for AglE in N-glycosylation. The MALDI-TOF spectra of the Asn-13-containing tryptic peptides derived from the S-layer glycoprotein of the WR536 background (upper panel) and *dpm1-B* (*aglE*)-deleted strains (lower panel) are shown. The components of the glycopeptide-associated sugar residues are shown in the inset box, while the glycan moieties decorating the peptide peaks are marked on the MALDI-TOF spectra, accordingly.

Expression and topology of the CBD-AglE construct. A. The protein contents of cells of the WR536 background strain or the same cells expressing CBD-AglE were precipitated, separated by SDS-PAGE and immunoblotted, using antibodies directed against the cellulose-binding domain. The positions of molecular weight markers are shown on the right. B. *H. volcanii* cells expressing CBD-AglE were broken by sonication, membrane and supernatant fractions were obtained by ultracentrifugation and the different fractions were immunoblotted with antibodies against the S-layer glycoprotein (SLG), SRP54 or CBD. Membrane fractions treated with 6 M urea and ultracentrifuged through a 40% sucrose cushion are shown in the right lanes of each panel. The same treatments were performed with *H. volcanii* cells expressing the soluble CBD moiety alone (bottom panel). C. *H. volcanii* cells (1 ml) expressing CBD-AglE were subjected to proteolysis with 1 mg/ml proteinase K at 55°C. 100 µl aliquots were removed every 30 minutes, TCA (15%) precipitated, separated by 12% SDS-PAGE and either Coomassie-stained (S-layer glycoprotein, upper panel) or immunoblotted with antibodies against SRP54 (middle panel) or CBD (lower panel). Antibody binding was detected by enhanced chemiluminescence.

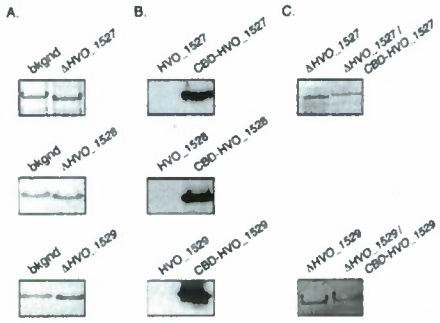


AglF, AglG and AglI are additional components of the *Haloferax volcanii* N-glycosylation pathway:

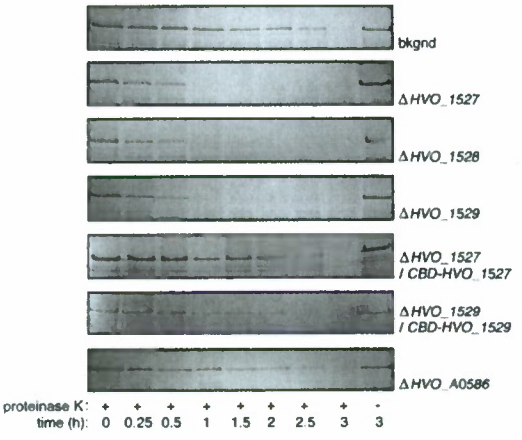


HVO_1527, *1528* and *1529* are not essential for *H. volcanii* survival. A. Schematic representation showing the orientations of *HVO_1527*, *1528*, *1529* and *1530* as well as their putative (*HVO_1527*, *1528*, *1529*) or proven (*HVO_1530*) annotations. B. Left panels: PCR amplification was performed using a forward primer directed at the *HVO_1527*, *1528* or *1529* 3' flanking regions and a reverse primer directed at a sequence within the *HVO_1527*, *1528* or *1529* coding regions (primer pair a) or a sequence within the *trpA* sequence (primer pair b), together with genomic DNA from cells of the WR536 background strain (bkgnd) or from cells that had replaced the *HVO_1527*, *1528* or *1529* gene (deletion; top, middle and bottom panels, respectively), as template. Right panel: PCR amplification was performed using primers directed against the *HVO_1527*, *1528* or *1529* coding regions, together with genomic DNA from cells of the WR536 background strain (bkgnd) or the *HVO_1527*, *1528* or *1529*-deleted strains (deletion; top, middle and bottom panels, respectively). C. RT-PCR was performed using primers directed at *HVO_1527* (top row of panels), *HVO_1528* (middle row of panels) or *HVO_1529* (bottom row of panels) and cDNA (left lane of each panel), RNA (middle lane of each panel) from *HVO_1527*, *1528* or *1529*-deleted strains (left, middle and right columns of panels, respectively) as template. In the right lane of each panel, no nucleic acid template was added to the reaction (blank). The identities of PCR products were confirmed by sequencing.

The absence of HVO_1527, 1528 or 1529 affects S-layer glycoprotein migration on SDS-PAGE. A. Equivalent aliquots of *H. volcanii* WR356 cells (bkgnd), the same cells lacking *HVO_1527* (Δ*HVO_1527*; top panel), *HVO_1528* (Δ*HVO_1528*; middle panel) or *HVO_1529* (Δ*HVO_1529*; bottom panel) or cells of the deletion strain transformed with a plasmid encoding CBD-tagged versions of the deleted gene (Δ*HVO_1527/CBD- HVO_1527*, top panel; Δ*HVO_1528/CBD- HVO_1528*, middle panel, Δ*HVO_1529/CBD- HVO_1529*, bottom panel) were separated by 5% SDS-PAGE and Coomassie blue-stained. The position of the S-layer glycoprotein is shown. B. The expression of CBD-tagged HVO_1527 (top panel), HVO_1528 (middle panel) and HVO_1529 (top panel) in the complemented deletion strains, as confirmed by immunoblot after separation on 15% SDS-PAGE using anti-CBD antibodies. C. Equivalent aliquots of *H. volcanii* WR356 cells lacking *HVO_1527* (Δ*HVO_1527*), *HVO_1529* (Δ*HVO_1529*) or

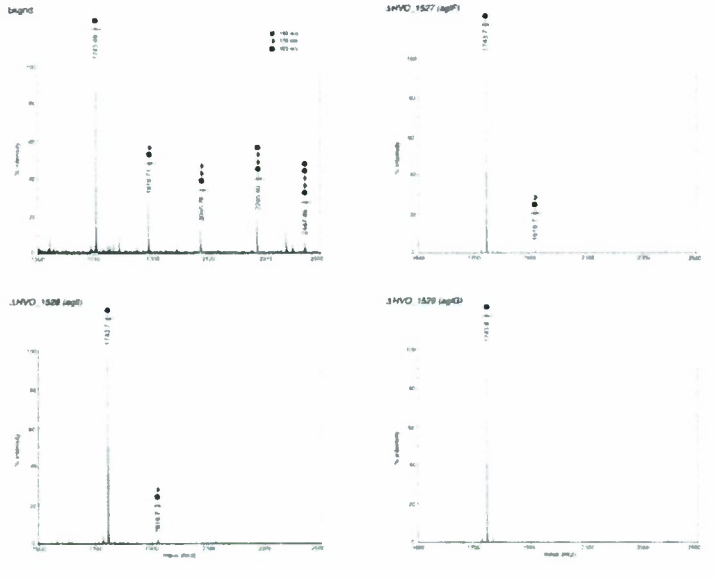


cells of the deletion strain transformed with a plasmid encoding a CBD-tagged version of the deleted gene (Δ*HVO_1527/CBD- HVO_1527;* Δ*HVO_1529/CBD-HVO_1529*) were separated by 5% SDS-PAGE and Coomassie blue-stained. The position of the S-layer glycoprotein is shown.

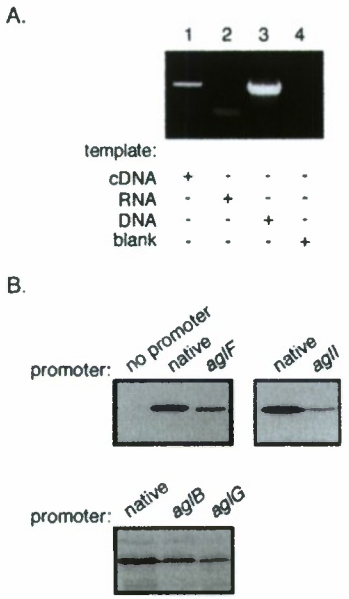


The S-layer surrounding *H. volcanii* cells is protease-sensitive in cells lacking *HVO_1527*, *1528* or *1529*. Background strain WR536 (top panel), and *HVO_1527*-, *1528*- or *1529*-lacking cells of the same strain (second, third and fourth panels, respectively) were challenged with 1 mg/ml proteinase K at 42°C. In the fifth and sixth panels, cells lacking *HVO_1527* or *HVO_1529*, respectively, transformed to express a CBD-tagged version of the deleted gene, were similarly challenged. Aliquots were removed immediately prior to incubation with proteinase K and at 15-30 min intervals following addition of the protease for up to 3 h and examined by 7.5% SDS-PAGE. In a control experiment, *H. volcanii* cells deleted of a seemingly non-related gene (*HVO_A0586*) were similarly challenged (bottom panel).

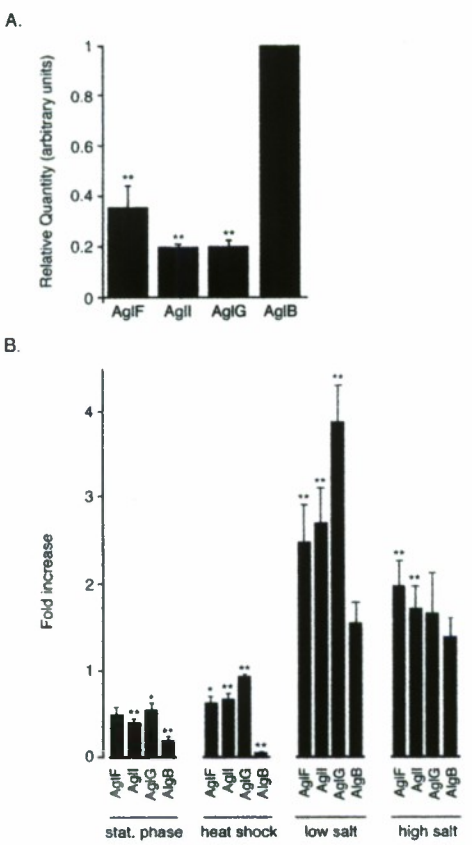
The products of *H. volcanii HVO_1527*, *1528* and *1529* are involved in the biogenesis of the pentasaccharide decorating S-layer glycoprotein Asn residues. The MALDI-TOF spectra of the Asn-13-containing tryptic peptide derived from the S-layer glycoprotein of the WR536 background cells (upper left panel) and or cells from the *HVO_1527*- (Δ*HVO_1527 (aglF)*; upper right panel), *HVO_1528*- (Δ*HVO_1528 (aglI)*; lower left panel) or *HVO_1529*- (Δ*HVO_1529 (aglG)*; lower right panel) deleted strains are shown. The components of the peptide-associated glycan are shown as an inset in the upper left panel, while the sugar subunits decorating the peptide peaks detected are indicated on the MALDI-TOF spectra.



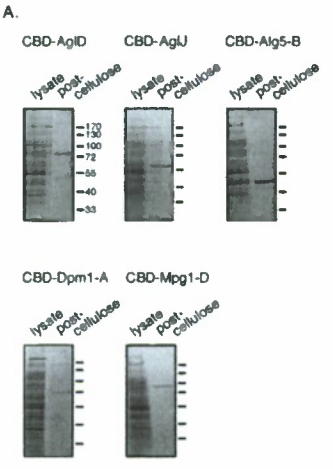
Functional characterization of the promoter regions of *H. volcanii aglB*, *aglF*, *algG* and *aglI*. A. RT-PCR reveals the co-transcription of *aglF* and *aglI*. PCR amplification was performed using a forward primer directed against the start of the coding region of *aglF* and a reverse primer against the end of the coding region of *aglI* together with cDNA (lane 1), RNA (lane 2) or DNA (lane 3) from *H. volcanii* strain WR536 background cells as template, or no nucleic acid template (lane 4). B. Upper panels: *H. volcanii* strain WR536 cells were transformed to express GFP, as directed by plasmid pJAM-1020 in which the promoter region had been removed, in which the native promoter was present, or when the region upstream to *aglF* or *aglI* replaced the native promoter of the plasmid. Lower panel: The 118 bp region separating *aglG* and *aglB* served as promoter (*aglB* lane). In lane *aglG*, the same region, this time in the reverse orientation, served as the promoter in plasmid pJAM-1020. GFP expression was visualized by immunoblotting using anti-GFP antibodies.



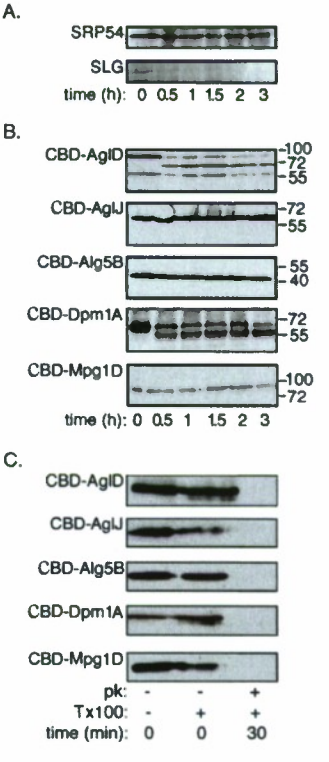
The transcription of *aglB*, *aglF*, *algG* and *aglI* is coordinated. A. Real-time RT-PCR was employed to assess the relative amounts of *aglB*, *aglF*, *algG* and *aglI* mRNA in *H. volcanii* strain WR536 cells grown to mid-exponential phase in rich medium. Values shown represent the average of three experiments ± standard deviation, expressed relative to the level of *aglB* mRNA, taken as 1. Differences from the level of *aglB* RNA marked with the double asterisk are statistically significant to $P < 0.01$, as determined by Student's t test. B. Real-time RT-PCR was employed to assess the fold increase in *aglB*, *aglF*, *algG* and *aglI* mRNA in *H. volcanii* strain WR536 cells grown to stationary phase (stat.), subjected to heat shock, or raised in low or high salt-containing medium, relative to those levels detected in cells grown to mid-exponential phase in rich medium. The values shown represent the average of three-five experiments. Bars marked with the double asterisk are statistically distinct to a significance of $P < 0.01$, while those marked with single asterisks are statistically distinct to a significance of $P < 0.05$, as determined by Student's t test.



Assessing the topology of N-glycosylation in *Haloferax volcanii*:

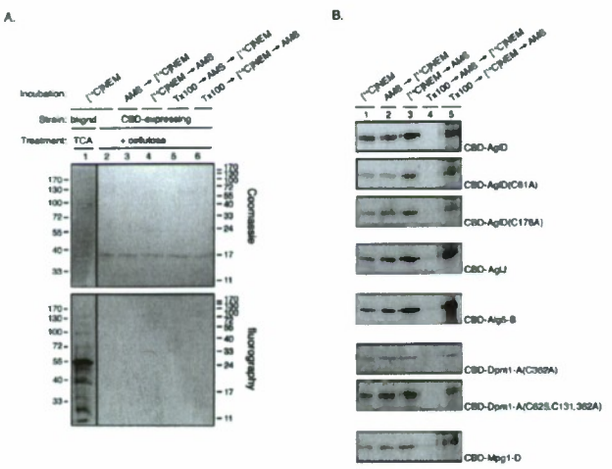


Cellulose-based purification of CBD-tagged *H. volcanii* enzymes proven or putatively involved in N-glycosylation. A. *H. volcanii* cells transformed to express CBD-AglD, -AlgJ, -Alg5-B, -Dpm1-A or Mpg1-D were lyzed and the CBD-tagged proteins purified with cellulose. Following separation by 10% SDS-PAGE, the proteins were visualized by Coomassie staining. B. Purified CBD-AglD, -AglJ, -Alg5-B, -Dpm1-A and Mpg1-D are recognized by anti-CBD antibodies in an immunblot protocol. Molecular mass markers (in kDa) are denoted on the right of each panel (A, with values shown only for the first panel) or the set of panels (B).



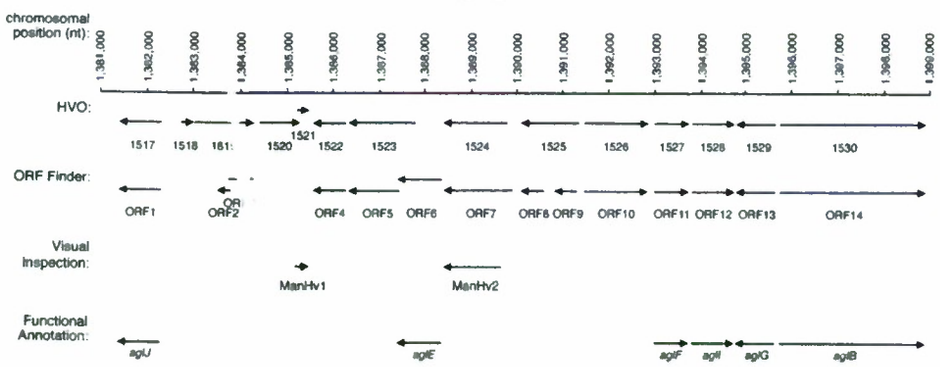
Proteinase K digestion of CBD-incorporating chimeras. A. In control experiments, *H. volcanii* cells were challenged with proteinase K (1 mg/ml, 55°C) for up to 3 h and the amount of surviving SRP54, a cytosolic marker, and the S-layer glycoprotein (SLG), a membrane protein marker, were assessed by immunoblotting using anti-SRP54 antibodies and Coomassie staining, respectively, following separation by 10% SDS-PAGE. B. *H. volcanii* cells transformed to express CBD-AglD, -AglJ, -Alg5-B, -Dpm1-A or Mpg1-D were similarly challenged with proteinase K and the profile of CBD-tagged proteins over the period of digestion was obtained by immunoblotting using anti-CBD antibodies. The positions of molecular mass markers (in kDa) are shown on the right of each panel. C. In control experiments, *H. volcanii* cells transformed to express CBD-AglD, -AglJ, -Alg5-B, -Dpm1-A or Mpg1-D were challenged with proteinase K (pk), in some cases following incubation with 1% Triton X-100 (Tx100), for 30 min.

Modification of cysteines as a tool for determining topology. A. Incubation of *H. volcanii* cells (Coomassie panel, lane 1) with [^{14}C]-NEM resulted in the labeling of a large number of proteins (fluorography panel, lane 1). CBD from *H. volcanii* cells were incubated with [^{14}C]-NEM, AMS and/or 1 % Triton X-100 in various combinations (fluorography panel, lanes 2-6). The cytoplasmically localized CBD moiety was not labeled by [^{14}C]-NEM (fluorography panel, lane 2), just as incubation with AMS either before (lane 3) or after [^{14}C]-NEM treatment (lane 4) did not result in CBD radiolabeling. Similarly, no radiolabeling was achieved when the cells were pre-treated with 1 % Triton X-100 and then challenged with AMS either before (lane 5) or after [^{14}C]-NEM treatment (lane 6). Molecular mass markers on left (170, 130, 100, 72, 55, 40 and 33 kDa) apply to lane 1, while those markers on the right (170, 130, 100, 72, 55, 40, 33, 24, 17 and 11 kDa) apply to lanes 2-6. B. *H. volcanii* cells transformed to express CBD-AglD, -AglJ, -Alg5-B, -Dpm1-A or Mpg1-D, as well CBD-AglD(C61A), CBD-AglD(C176A), CBD-Dpm1-A(C362A) and CBD-Dpm1-A(C62A,C131,362A) mutants were incubated with [^{14}C]-NEM alone (lane 1), or before (lane 3) or after (lane 2) treatment with AMS. In other experiments, the cells were incubated with 1% Triton X-100 (Tx100) prior to incubation with AMS and then [^{14}C]-NEM (lane 4) or vice versa (lane 5). Cysteine-mutated proteins were generated by site-directed mutagenesis.

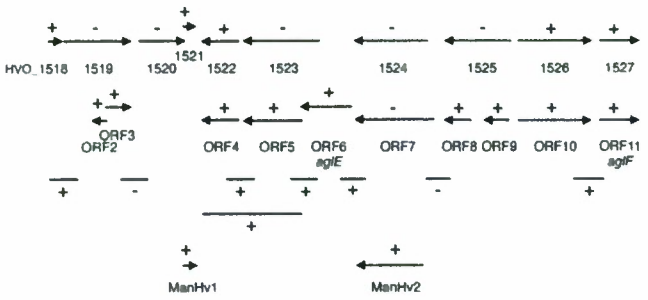


Examination of the *Haloferax volcanii agl* gene cluster reveals novel genes involved in N-glycosylation:

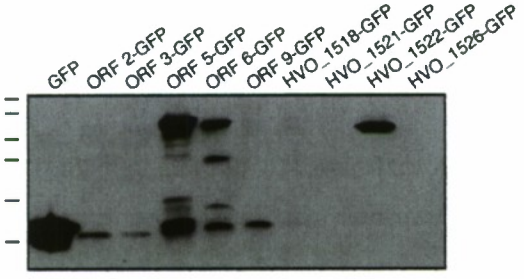
Schematic depiction of the *agl* cluster of the annotated *H. volcanii* genome. Shown are annotations automatically generated by the Glimmer (found at http://archaea.ucsc.edu/cgi-bin/hgGateway?db=haloVolc1; listed as HVO numbers) or ORF Finder (listed as ORF numbers) algorithms, as well as by visual inspection. Functional annotations are reflected as *agl* gene names.



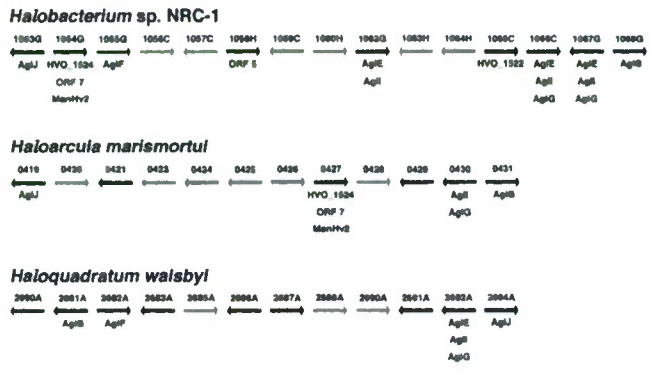
Schematic depiction of the transcripts generated by RT-PCR. Plus (+) symbols reflect sequences that generated transcripts, while minus (-) symbols reflect those sequences that failed to generate transcripts. Primers raised against either the proposed 5' and 3' regions of the gene of interest or sequences spanning the 3' region of a sequence and the 5' region of the downstream sequence, designed to address co-transcription of the sequences of interest, were employed.



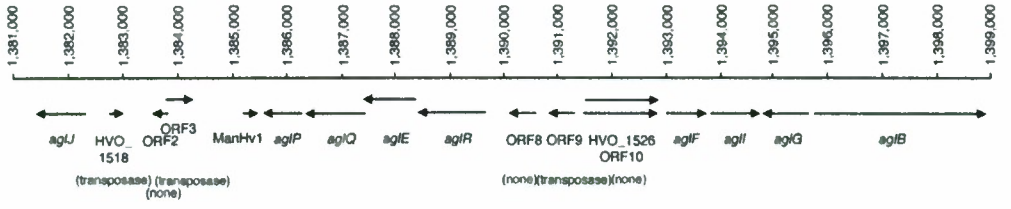
The appearance of GFP-containing fusion proteins reveals genes that are expressed. *H. volcanii* cells were transformed with plasmids encoding a sequence of interest fused at its 3' end to DNA encoding GFP. This construct was preceded by the 200 bp region immediately upstream to the sequence of interest. In a control experiment, plasmid pJAM1020 (19), directing the expression of GFP alone, was employed. Expression of the GFP-containing chimeras was then assessed by immunoblot analysis using antibodies against GFP. The positions of molecular weight markers (100, 70, 55, 40, 35, and 25 kDa) are shown on the left.



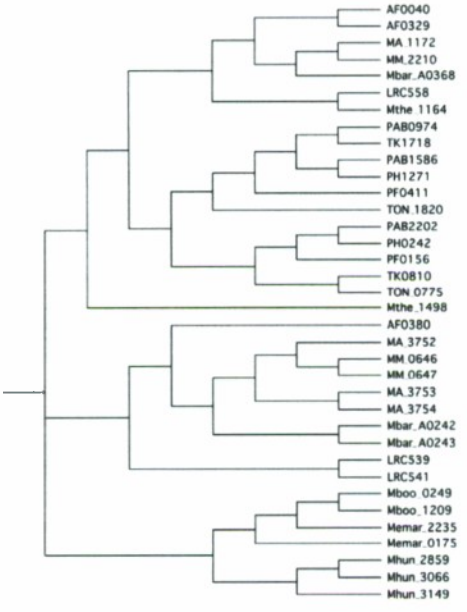
Schematic representation showing the presence of gene clusters containing *H. volcanii agl* gene homologues in other haloarchaea. Regions of the genomes of *Halobacterium* sp. NRC-1, *Haloarcula marismortui* and *Haloquadratum walsbyi* that include *H. volcanii agl* gene homologues, homologues to other previously annotated sequences found within the *H. volcanii agl* gene-containing island, homologues to *H. volcanii* ORF sequences defined in this study, or to other sequences automatically annotated as participating in protein glycosylation, are shown (black arrows). The grey arrows in each gene cluster represent sequences automatically annotated as hypothetical proteins or proteins assigned functions apparently unrelated to protein glycosylation. The arrows representing each ORF do not reflect the relative length of a given sequence; instead, they are of arbitrary but equal length. The number above each arrow corresponds to the identification number of that ORF in the genome sequence of that haloarcheaon.



The revised version of the *H. volcanii agl* gene cluster. Numbers along the horizontal bar denote nucleotide position in the genome. Proven or putative N-glycosylation genes are termed *agl* (*a*rchaeal *gl*ycosylation) sequences. For genes not assigned roles in N-glycosylation, proposed functions are given below each sequence



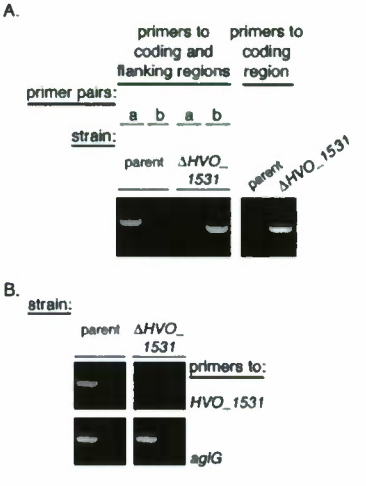
Cataloguing glycosyltransferases and oligosaccharyltransferases in Archaea:



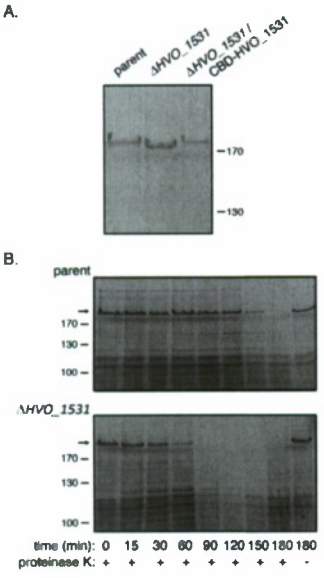
The archaeal AglB phylogenetic tree. The tree was generated by the Neighbor-Joining method with the Mega package. The bar represents two substitutions per 100 positions. Bootstrap values are indicated at branch nodes.

***Haloferax volcanii* AglF and AglM can function in a coordinated manner:**

Deletion of *HVO_1531* does not comprise *H. volcanii* cell viability. A. Left panel: PCR amplification was performed using a forward primer raised against a sequence within the coding region of *HVO_1531* (lane a) or *trpA* (lane b) and a reverse primer raised against the 3' downstream region of *HVO_1531*, together with genomic DNA from the WR536 parent strain (parent) or the *HVO_1531* deletion strain (*ΔHVO_1531*) as template. Right panel: PCR amplification was performed with primers directed against the *HVO_1531* open reading frame, together with genomic DNA from cells of the WR536 parent strain (parent) or the *HVO_1531* deletion strain (*ΔHVO_1531*) serving as template. B. RT-PCR was

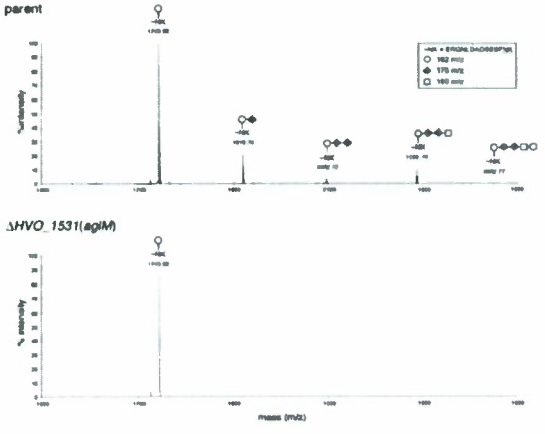


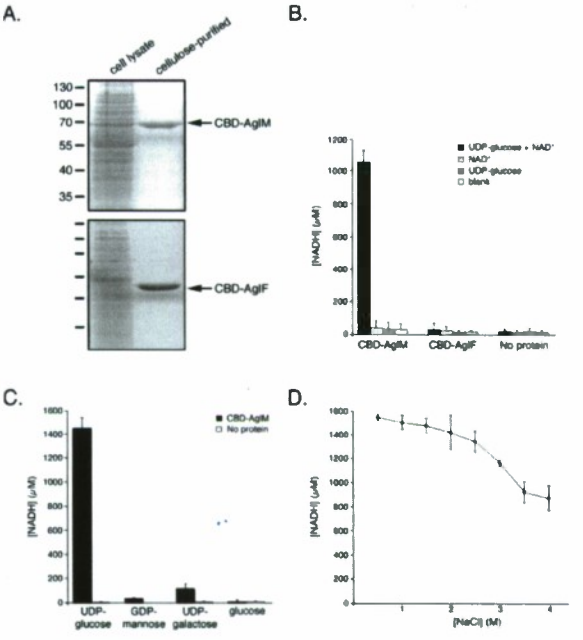
performed using primers directed against the *HVO_1531* (upper panel) or *aglG* (lower panel) open reading frames and cDNA (left lane) or RNA (right panel) from the WR536 parent strain (parent) or the *HVO_1531* deletion strain (*ΔHVO_1531).*



Cells deleted of *HVO_1531* present a modified S-layer. A. Equivalent amounts of *H. volcanii* WR536 cells (parent), *HVO_1531*-deleted cells *(ΔHVO_1531)* and *HVO_1531*-deleted cells transformed with plasmid pWL-CBD-HVO_1531 (*ΔHVO_1531 / CBD-HVO_1531*) were separated by 5% SDS-PAGE and Coomassie-stained. B. Parent strain WR536 (top panel) and *HVO_1531*-deleted cells (lower panel) were challenged with 1 mg/ml proteinase K at 42°C. Aliquots were removed immediately prior to incubation with proteinase K and at 15-30 min intervals following addition of the protease for up to 3 h and examined by 7.5% SDS-PAGE. The position of the S-layer glycoprotein in each panel is indicated by an arrow. In A and B, the positions of molecular weight markers are indicated on the right and left, respectively.

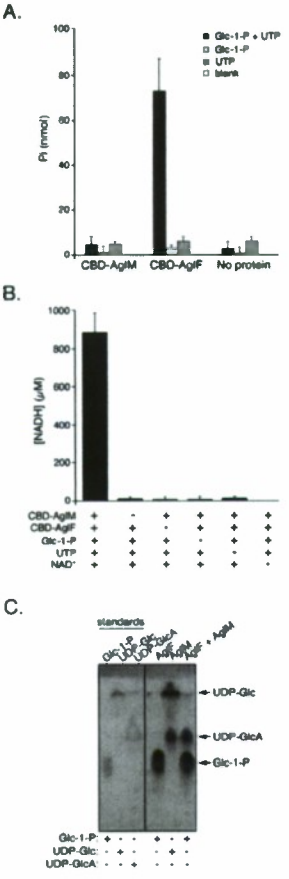
HVO_1531 (aglM) is involved in the biogenesis of the pentasaccharide decorating the *H. volcanii* S-layer glycoprotein. The MALDI-TOF spectra of the Asn-13-containing tryptic peptide derived from the S-layer glycoprotein of WR536 parent cells (upper panel) and cells of the *HVO_1531* deletion strain (lower panel) are shown. The components of the peptide-associated glycan are shown as an inset in the upper panel, while the sugar subunits decorating the peptide peaks detected are indicated on the MALDI-TOF spectra.



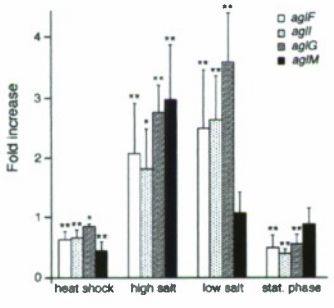


AglM functions as a NAD^+-dependent UDP-glucose dehydrogenase. A. *H. volcanii* cells transformed to express CBD-AglM or CBD-AglF were subjected to cellulose-based capture. The cell lysate and cellulose-bound proteins were separated on 10% SDS-PAGE and Coomassie-stained. Protein bands corresponding to CBD-AglM (~65 kDa; upper panel) and CBD-AglF (~50 kDa; lower panel) are indicated by arrows. The positions of molecular weight markers are shown on the left of each panel, with the molecular weight of each marker given in the upper panel. B. Cellulose-bound CBD-AglM, CBD-AglF or cellulose beads alone (No protein) were resuspended in reaction buffer and incubated in the presence of UDP-glucose and/or NAD^+ or no substrate (blank). The increase in NADH concentration was measured by comparing absorbance at 340 nm immediately after substrate addition and after a 1 h incubation at 42°C. C. Cellulose-bound CBD-AglM or cellulose beads (No protein) were resuspended in reaction buffer and incubated in the presence of NAD^+ and UDP-glucose, GDP-mannose, UDP-galactose or glucose. The increase in NADH concentration was measured by comparing absorbance at 340 nm immediately after substrate addition and after a 90 min incubation at 42°C. D. Cellulose-bound CBD-AglM was incubated in the presence of UDP-glucose and NAD^+ in reaction buffer containing 0.5-4 M NaCl. The increase in NADH concentration was measured by comparing absorbance at 340 nm immediately after substrate addition and after a 90 min incubation at 42°C. In B, C and D, the results show the average of three experiments ± standard deviation.

AglM and AglF are able to function in a sequential manner. A. Cellulose-bound CBD-AglM, CBD-AglF or cellulose beads (No protein) were resuspended in reaction buffer and incubated in the presence of glucose-1-phosphate (Glc-1-P) and/or UTP or no substrate (blank). The extent of phosphate release was measured by a malachite green-based assay immediately after substrate addition and after a 90 min incubation at 42°C. The results show the average of three experiments ± standard deviation. B. Cellulose-bound CBD-AglM and/or CBD-AglF were resuspended in reaction buffer in the presence or absence of glucose-1-phosphate (Glc-1-P), UTP and NAD^+. The increase in NADH concentration was measured by comparing absorbance at 340 nm immediately after substrate addition and after a 2 h incubation at 42°C. The results show the average of 3 experiments ± standard deviation. C. The reactions described in A and B and in the legend to the previous figure were repeated to assay AglF alone, AglF and AglM together and AglM alone, respectively. The assay products were separated by TLC, along with UTP, NAD^+, glucose-1-phosphate (Glc-1-P), UDP-glucose (UDP-Glc) and UDP-glucuronic acid (UDP-GlcA) standards.

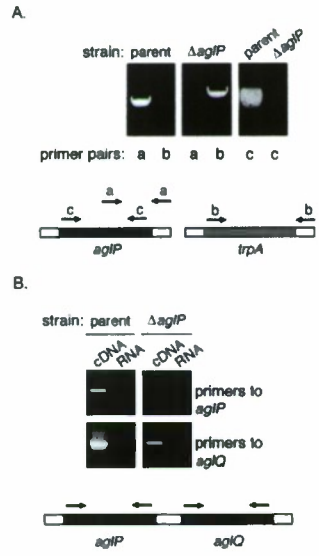


The transcription pattern of *aglM* is similar to that of *aglF, aglG* and *aglI* in the face of heat shock and hypersalinity. Real-time RT-PCR was employed to assess the fold increase in the levels of *aglF*, *aglI*, *algG* and *aglM* mRNA in *H. volcanii* strain WR536 cells subjected to heat shock (65°C, 45 min), grown in high or low salt-containing medium, or grown to stationary phase (stat.), relative to those levels detected in cells grown to mid-exponential phase in rich medium. The values shown represent the average of three-five experiments ± standard deviation. Bars marked with the double asterisk are statistically distinct to a significance of $P < 0.01$, while those marked with single asterisks are statistically distinct to a significance of $P < 0.05$, as determined by Student's t test.

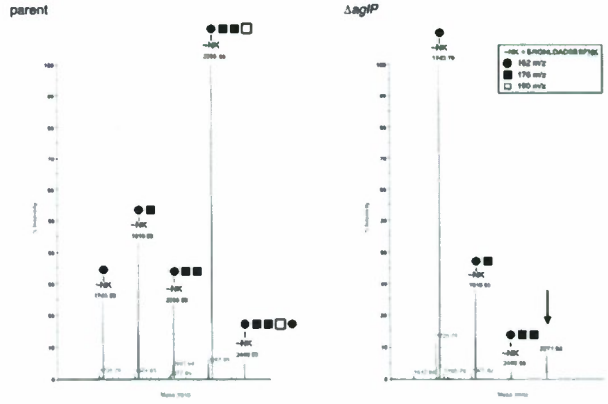


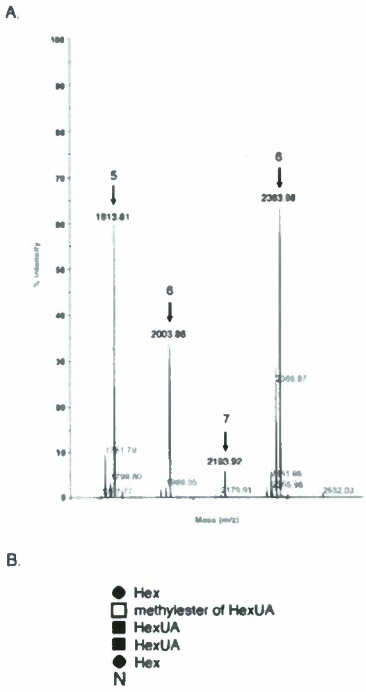
AglP is a S-adenosyl-L-methionine-dependent methyltransferase that participates in the N-glycosylation pathway of *Haloferax volcanii*:

aglP is not essential for *H. volcanii* survival. A. Left and middle panels: PCR amplification was performed using a forward primer directed to a sequence within the *aglP* coding region and a reverse primer directed at the *aglP* 3' flanking region (primer pair a) or using a forward primer directed to a sequence at the start of the *trpA* sequence and the same reverse primer as above (primer pair b), together with genomic DNA from cells of the parent strain (parent; left panel) or from cells where *aglP* had been replaced with *trpA* (Δ*aglP*; middle panel), as template. Right panel: PCR amplification was performed using primers pair c directed against the *aglP* coding region, together with genomic DNA from cells of the parent strain (parent) or the *aglP*-deleted strain (Δ*aglP*) as template. The positions to with the various primer pairs bind are shown in the drawing below the panels. Note that *aglP* and *trpA* are surrounded by the same flanking regions. B. RT-PCR was performed using primers directed at *aglP* (top row of panels) or *aglQ* (bottom row of panels) together with cDNA (left lane of each panel) or RNA (right lane of each panel) from the parent strain (parent; left column of panels) or *aglP*-deleted cells (Δ*aglP*; right columns of panels) as template. The identity of the PCR products was confirmed by sequencing. The positions to with the various primer pairs bind, as well as the relative positions of *aglP* and *aglQ* in the genome, are shown in the drawing below the panels.

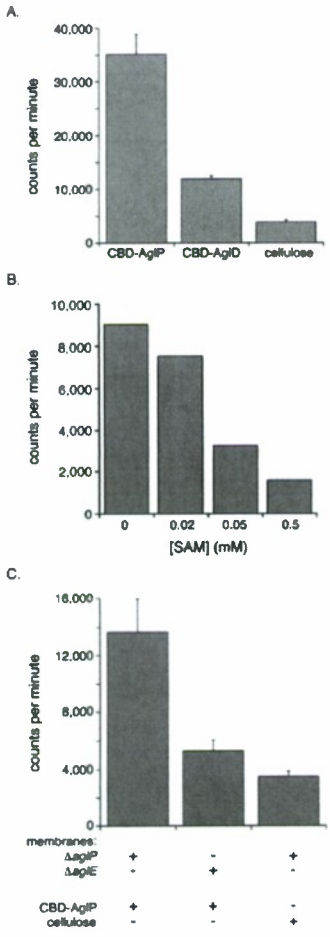


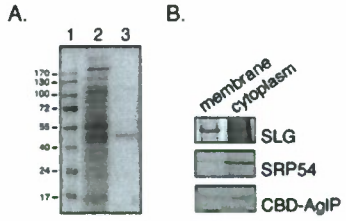
MALDI-TOF MS analysis of the Asn-13-containing *Hfx. volcanii* S-layer glycoprotein-derived tryptic glycopeptides reveals a role for AglP in N-glycosylation. The MALDI-TOF spectra of the Asn-13-containing tryptic peptide derived from the S-layer glycoprotein from the parent (left panel; parent) and *aglP*-deleted strains (right panel; Δ*aglP*) are shown. The components of the glycopeptide-associated sugar residues are shown in the inset box, while the glycan moieties decorating the peptide peaks are marked on the spectra, accordingly. In the Δ*aglP* panel, the arrow shows the position of the peptide modified by the tetrasaccharide that includes a novel subunit at position four.





Methyl esterification, using methanolic HCl, of the Asn-13-containing *H. volcanii* S-layer glycoprotein-derived tryptic glycopeptide from cells lacking AglP confirms the 190 Da saccharide of the peptide-linked pentasaccharide is a methylated ester of hexuronic acid. A. The Asn-13-containing *H. volcanii* S-layer glycoprotein-derived glycopeptide from cells lacking AglP was incubated with 1M MeOH-HCl. The molecular ions of the esterified peptide modified by the mono-, di- and trisaccharides, as well as by the tetrasaccharide terminating in a 176 Da unit, are indicated by arrows. The numbers above each arrow indicate the number of carboxylic acids converted to their methyl esters, with each such conversion adding 14 Da to the mass of the peptide. B. Schematic depiction of the pentasaccharide attached to the S-layer glycoprotein. N – modified asparagine residue; Hex – hexose; HexUA – hexuronic acid.





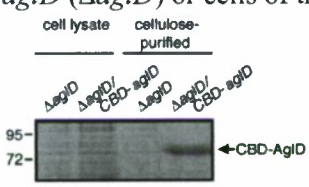
(On the right) Purification and sub-cellular localization of CBD-AglP. A. CBD-AglP can be purified from *H. volcanii* cells transformed to express the chimera. A Coomassie-stained SDS-PAGE gel showing molecular weight markers (lane 1), an aliquot of a total *H. volcanii* protein extract (lane 2) and cellulose-purified CBD-AglP (lane 3) is presented. B. CBD-AglP is localized to the cytosol. *H. volcanii* cells expressing CBD-AglP were broken by sonication, membrane and cytoplasmic fractions were obtained by ultracentrifugation and the different fractions were Coomassie-stained to reveal the presence of the S-layer glycoprotein (SLG, migrating as a 180-200 kDa protein (top panel) or immunoblotted with antibodies against SRP54 (migrating as a 50 kDa protein (middle panel) or CBD (lower panel).

(On the left) AglP is a SAM-dependent methyltransferase. A. The ability of cellulose-bound CBD-AglP to transfer a [^{3}H]-methyl group from [^{3}H]-methyl-SAM to membrane fragments prepared from *H. volcanii* Δ*aglP* cells was tested. In control experiments, cellulose-bound CBD-AglD or untreated

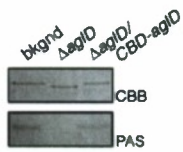
cellulose beads were employed. A representative experiment of four is shown, with each bar corresponding to the mean value of triplicate samples ± standard deviation. B. The experiment described in A was repeated with cellulose-bound CBD-AglP and increasing amounts (0-0.5 mM) of unlabeled SAM. A representative experiment of three is shown. In each column, the background reading obtained in the same experiment upon incubation of membrane fragments prepared from *H. volcanii* Δ*aglP* cells with untreated cellulose beads were subtracted. C. The ability of cellulose-bound CBD-AglP to perform [^{3}H]-methylation of membrane fragments prepared from *Hfx. volcanii* Δ*aglP* or Δ*aglE* cells upon [^{3}H]-methyl-SAM addition was tested. In control experiments, untreated cellulose beads were combined with membrane fragments prepared from *H. volcanii* Δ*aglP* cells and [^{3}H]-methyl-SAM. Each bar represents the average of 3 samples ± standard deviation.

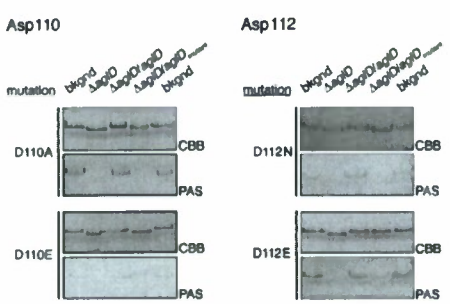
A novel in vivo assay for the identification of residues important for the activity of *Haloferax volcanii* AglD, a component of the archaeal N-glycosylation pathway:

Complementation of *H. volcanii* cells deleted of *aglD*. *H. volcanii* cells deleted of *aglD* (Δ*aglD*) or cells of the same strain transformed with a plasmid encoding CBD-AglD (Δ*aglD/CBD-aglD*) were disrupted by sonication (cell lysate). The lysates were then subjected to cellulose-based purification (cellulose-purified). The position of CBD-AglD is indicated (left), as are the positions of 95 and 72 kDa molecular weight markers (right).

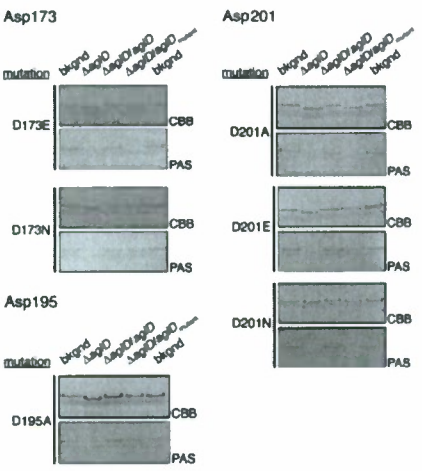


aglD-complemented *H. volcanii* cells regain the ability to properly glycosylate the S-layer glycoprotein. The protein contents of cells of the WR536 background strain (bkgnd), the same strain deleted of *aglD* (Δ*aglD*) or the AglD-lacking strain transformed with a plasmid encoding CBD-AglD (Δ*aglD/CBD-aglD*) were separated by 5% SDS-PAGE and the S-layer glycoprotein was detected by Coomassie stain (CBB) or periodic acid-Schiff reagent (PAS). In the presence of CBD-AglD, the migration and positive glycostaining of the S-layer glycoprotein is as observed in the background strain.



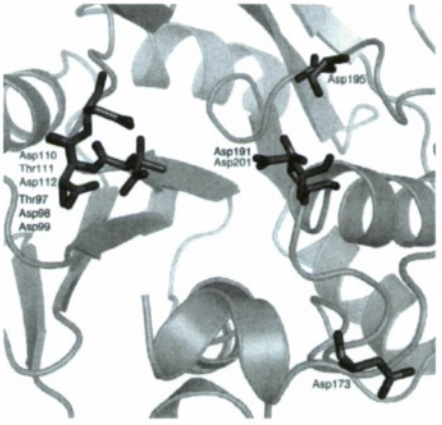


H. volcanii AglD Asp110 and Asp112 residues are assigned as the GT2 DXD motif involved in the catalytic activity of the enzyme. Site-directed mutagenesis was performed to generate CBD-AglD containing mutations of Asp110 (left column) or Asp112 (right column), as listed on the left of each panel. For each mutant, the upper and lower panels respectively show the Coomassie- and PAS-stained S-layer glycoprotein from the background strain (lanes 1 and 5), from the *aglD* deletion strain (lane 2), from the *aglD* deletion strain complemented to with a plasmid encoding CBD-AglD (lane 3) or from the *aglD* deletion strain complemented to with a plasmid encoding CBD fused to mutated AglD (lane 4).

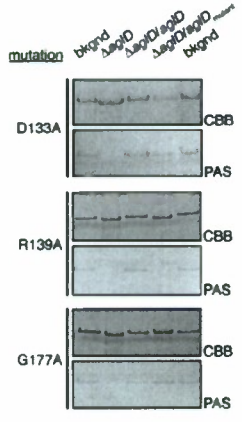


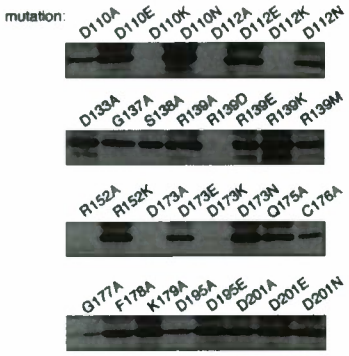
H. volcanii AglD Asp201 is assigned as the catalytic base of the enzyme. Site-directed mutagenesis was performed to generate CBD-AglD containing mutations of Asp173 (upper left column), Asp195 (lower left column) and Asp201 (right column), as listed on the left of each panel. For each mutant, the upper and lower panels respectively show the Coomassie- and PAS-stained S-layer glycoprotein from the background strain (lanes 1 and 5), from the *aglD* deletion strain (lane 2), from the *aglD* deletion strain complemented to with a plasmid encoding CBD-AglD (lane 3) or from the *aglD* deletion strain complemented to with a plasmid encoding CBD fused to mutated AglD (lane 4).

Homology modeling of *H. volcanii* AglD residues based on the available three-dimensional structure of *B. subtilis* SpsA. Structural modeling was performed by using the SWISS-MODEL program (http://swissmodel.expasy.org/) and visualized using PyMol (http://www.pymol.org/). *B. subtilis* SpsA Thr97, Asp98, Asp99 and Asp191 are shown in blue, while *H. volcanii* AglD Asp110, Thr111, Asp112, Asp173, Asp195 and Asp 201 are shown in brown. The ribbon diagram in the background corresponds to the three-dimensional structure of SpsA. The RMS value, reflecting the quality of the homology modeling, is 0.61 angstroms.



The conserved Arg139 residue is needed for AglD activity, unlike the conserved Asp133 or Gly177 residues. Site-directed mutagenesis was performed to generate CBD-AglD containing mutations of Asp133, Arg139 and Gly177, as listed on the left of each panel. For each mutant, the upper and lower panels respectively show the Coomassie- and PAS-stained S-layer glycoprotein from the background strain (lanes 1 and 5), from the *aglD* deletion strain (lane 2), from the *aglD* deletion strain complemented to with a plasmid encoding CBD-AglD (lane 3) or from the *aglD* deletion strain complemented to with a plasmid encoding CBD fused to mutated AglD (lane 4).



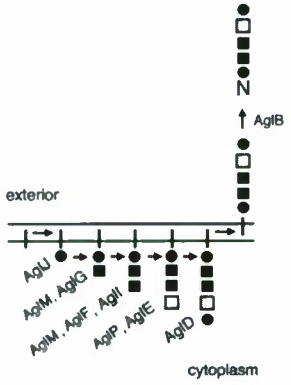


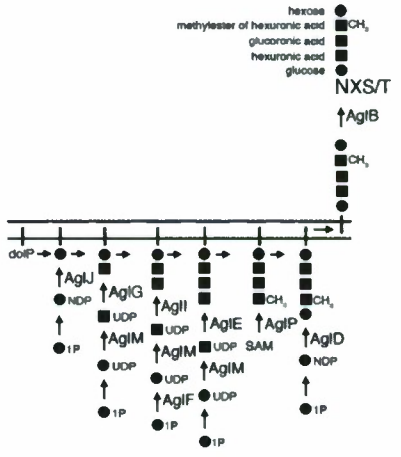
Expression levels of the various versions of CBD-AglD. *H. volcanii* cells expressing the various AglD mutants considered in this study were grown to OD_{550} 1.0 and their protein contents were separated on 10% SDS-PAGE. The CBD-AglD content of each strain was subsequently assessed by immunoblot using polyclonal anti-CBD antibodies. Antibody binding was detected using HRP-conjugated secondary antibodies and an enhanced chemiluminescence kit.

Impact of findings

Protein glycosylation is known to modulate numerous aspects of protein behaviour, ranging from folding and degradation to oligomerization and other protein-protein interactions. Long-thought to be restricted to Eukarya, it is becoming increasingly clear that prokaryotes are also capable of glycosylating a broad range of proteins. In Archaea, glycosylation has been cited as aiding proteins to remain stable in the face of physical challenges associated with the harsh environments in which these micoorganisms can exist. Accordingly, archaeal glycoproteins present a variety in the composition of their glycan moieties not detected in either Eukarya or Bacteria. However, in contrast to our relatively advanced understanding of the eukaryal glycosylation process and the growing body of information on bacterial glycosylation being amassed, little was known of the manner by which glycan moieties are added to selected archaeal proteins prior to these studies.

By combining bioinformatics, genetic, biochemistry and structural approaches, the funded research succeeded in identifying a group of clustered *H. volcanii* genes (the *agl* genes) encoding proteins involved in the assembly and attachment of a pentasaccharide to select asparagine residues of the S-layer glycoprotein, a reporter of N-glycosylation in this species.





Our findings are of importance on several fronts. In addition to providing insight into N-glycosylation across evolution, our investigations have led to better understanding of the process and significance of protein glycosylation under extreme conditions. Such information could be exploited to tap the enormous potential promised by extremophilic archaeal enzymes in a variety of biotechnological applications. Specifically, information derived from our finding could allow for the creation of archaeal strains expressing selected N-glycosylation enzyme modules, thereby allowing for a harnessing of the greater diversity associated with this post-translational modification in Archaea in the design of tailor-made glycoproteins. Indeed, the information gained in these studies can be extrapolated to other archaeal phenotypes and eventually be adapted to offer a flexible platform for using protein glycosylation to modulate extremophilic archaeal protein stability.

Publications resulting from research supported by this grant

Articles:

1. Abu-Qarn, M., Yurist-Doutsch, S., Giordano, A., Trauner A., Morris H.R., Hitchen P., Medalia, O., Dell, A. and 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. J. Mol. Biol., 374, 1224-1236.

2. Abu-Qarn, M., Giordano, A., Battaglia, F., Trauner A., Morris H.R., Hitchen P., Dell, A. and Eichler, J. (2008) Identification of AglE, a second glycosyltransferase involved in N-glycosylation of the *Haloferax volcanii* S-layer glycoprotein. J. Bacteriol., 190, 3140-3146.

3. Yurist-Doutsch, S., Chaban, B., VanDyke, D., Jarrell, K.F. and Eichler, J. (2008) Sweet to the extreme: Protein glycosylation in Archaea. Mol. Microbiol., 68, 1079-1084.

4. Yurist-Doutsch, S., Abu-Qarn, M., Battaglia, F., Morris, H.R., Hitchen, P.G., Dell, A. and Eichler, J. (2008) *aglF*, *aglG* and *aglI*, novel members of a gene cluster involved in the N-glycosylation of the *Haloferax volcanii* S-layer glycoprotein. Mol. Microbiol. 69, 1234-1245.

5. Abu-Qarn, M., Eichler, J. and Sharon, N. (2008) Not just for Eukarya anymore: N-glycosylation in Bacteria and Archaea. Curr. Opin. Struct. Biol.,18, 544-550.

6. Plavner, N. and Eichler, J. (2008) Defining the topology of the N-glycosylation pathway in the halophilic archaea *Haloferax volcanii*. J. Bacteriol., 190, 8045-8052.

7. Yurist-Doutsch, S., and Eichler, J. (2009) Manual annotation, transcriptional analysis and protein expression studies reveal novel genes in the *agl* cluster responsible for N-glycosylation in the halophilic archaeon *Haloferax volcanii*, J. Bacteriol., 191, 3068-3075.

8. Magidovich, H., and Eichler, J. (2009) Glycosyltransferases and oligosaccharyltransferases in Archaea: Putative components of the N-

glycosylation pathway in the third domain of life. FEMS Microbiol. Lett., 300, 122-130.

9. Yurist-Doutsch, S., Magidovich, H., Ventura, V.V., Hitchen, P.G., Dell, A. and Eichler, J. (2010) N-glycosylation in Archaea: On the coordinated actions of *Haloferax volcanii* AglF and AglM. Mol. Microbiol., in press.

10. Magidovich, H., Yurist-Doutsch, S., Konrad, Z., Ventura, V.V., Hitchen, P.G., Dell, A. and Eichler, J. AglP is a cytoplasmic S-adenosyl-L-methionine-dependent methyltransferase that participates in the N-glycosylation pathway of *Haloferax volcanii*. In revision.

11. Kaminski, L., and Eichler, J. A novel in vivo assay for the identification of residues important for the activity of *Haloferax volcanii* AglD, a component of the archaeal N-glycosylation pathway. Submitted.

Book chapters:

1. Eichler, J., Abu-Qarn, M., Konrad, Z., Magidovich, H., Plavner, N. and Yurist-Doutsch, S. (2010) The cell envelopes of haloarchaea: Staying in shape in a world of salt. In: Prokaryotic Cell Wall Compounds: Structure and Biochemistry (H. Konig, H. Claus and A. Varma, eds.), Springer, Heidelberg. In press.

Theses supported by this grant

M.Sc. thesis:
Noa Plavner, "The topology of proteins involved in N-glycosylation in the Archaea *Haloferax volcanii*", October, 2008.

Ph.D. thesis:
Mehtap Abu-Qarn, "Protein Glycosylation in Archaea", September, 2009.

Both theses are available online from The Kreitman School for Graduate Studies of Ben-Gurion University of the Negev (http://bgu4u.bgu.ac.il/html/thesis/).