N-glycan core synthesis by Alg24 strengthens the hypothesis of an archaeal origin of eukaryal N-glycosylation

Benjamin H. Meyer[1,2,4 #], Ben A. Wagstaff[2], Panagiotis S. Adam[3],

Sonja-Verena Albers[4], and Helge C. Dorfmueller[2#]

[1] Molecular Enzyme Technology and Biochemistry, Environmental Microbiology and Biotechnology, Faculty of Chemistry University Duisburg-Essen, Germany;
[2] Division of Molecular Microbiology, School of Life Sciences, University of Dundee, Dundee, DD1 5EH UK;
[3] Group for Aquatic Microbial Ecology, Environmental Microbiology and Biotechnology, Faculty of Chemistry University Duisburg-Essen, Germany;
[4] Molecular Biology of Archaea, Faculty of Biology, University of Freiburg, 79104 Freiburg, Germany

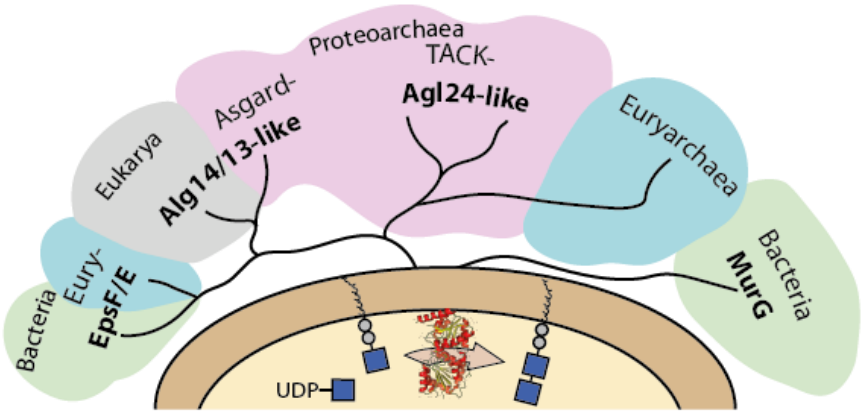
[#]Correspondence to B. H. Meyer, phone: +49 201 183 7084, e-mail: benjamin.meyer@uni-due.de and H. C. Dorfmueller, +44 1382 386203, e-mail: hczdorfmueller@dundee.ac.uk

Key words: **chitobiose, Archaea, N-glycosylation, N-glycan, Alg24, Alg13, Agl14, MurG**

Abbreviation: Agl, archaeal glycosylation enzyme; Alg, Asparagine linked glycosylation; CAZy, Carbohydrate-Active enZymes database; Dol, dolichol; Dol-P, dolichol-phosphate; DPANN, archaeal superphylum containing Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, Nanohaloarchaeota, Woesearchaeota and Pacearchaeota; Glc, glucose; GlcA, glucuronic acid; GlcNAc, N-acetylglucosamine; GFP, green fluorescent protein; GT, glycosyltransferase; LLO, lipid-linked oligosaccharide; MALDI-MS, matrix-assisted laser desorption ionization mass spectrometry; OST, oligosaccharyltransferase; SDS-PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; TACK, archaeal superphylum containing Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota; TMD, transmembrane domain; UDP, uridine diphosphate; Und, undecaprenol

Short title: Alg24 synthesizes the chitobiose N-glycan core

Cover Art:



Abstract

Protein N-glycosylation is one of the most common posttranslational modifications found in all three domains of life. The crenarchaeal N-glycosylation begins with the synthesis of a lipid-linked chitobiose core structure, identical to that in eukaryotes. Here, we report the first identification of a thermostable archaeal beta-1,4-N-acetylglucosaminyltransferase, named archaeal glycosylation enzyme 24 (Agl24), responsible for the synthesis of the N-glycan chitobiose core. Biochemical characterization confirmed the function as an inverting β-D-GlcNAc-(1→4)-α-D-GlcNAc-diphosphodolichol glycosyltransferase. Substitution of a highly conserved histidine residue, found also in the eukaryotic and bacterial homologs, demonstrated its importance for the function of Agl24. Furthermore, bioinformatics and structural modeling revealed strong similarities between Agl24 and both the eukaryotic Alg14/13 and a distant relation to the bacterial MurG, which catalyze the identical or a similar process, respectively. Our data, complemented by phylogenetic analysis of Alg13 and Alg14, revealed similar sequences in Asgardarchaeota, further supporting the hypothesis that the Alg13/14 homologs in eukaryotes have been acquired during eukaryogenesis.

Highlights

- First identification and characterization of a thermostable β-D-GlcNAc-(1→4)-α-D-GlcNAc-diphosphodolichol glycosyltransferase (GT family 28) in Archaea.
- A highly conserved histidine, within a GGH motif in Agl24, Alg14, and MurG, is essential for function of Agl24.
- Agl24-like homologs are broadly distributed among Archaea.
- The eukaryotic Alg13 and Alg14 are closely related to the Asgard homologs, suggesting their acquisition during eukaryogenesis.

Introduction

Asparagine (N)-linked glycosylation is one of the commonest co- and posttranslational protein modifications found in all three domains of life (Larkin and Imperiali, 2011, Jarrell et al., 2014, Nothaft and Szymanski, 2010) In Eukarya, N-glycosylation is an essential process and is evolutionarily highly conserved from yeast to humans (Lehle et al., 2006). It is estimated that more than half of all eukaryotic proteins are glycoproteins (Apweiler et al., 1999, Zielinska et al., 2010). N-glycosylation is required for correct protein folding and stability, intra- and extracellular recognition of protein targets and for enzyme activity (Caramelo and Parodi, 2007, Helenius and Aebi, 2004, Varki, 1993). Therefore, the biological functions of protein glycosylation range from relatively minor to crucial for survival of an

62 organism. As such, congenital disorders of glycosylation in humans are often lethal or cause severe
63 phenotypes and diseases ranging from intellectual disability, growth retardation, cardiac anomalies,
64 and early death (Reily et al., 2019). Although it was long thought that N-glycosylation was restricted to
65 Eukarya, N-glycosylation pathways have now been characterized in all three domains of life. Bacterial
66 N-glycosylation seems to be restricted to delta/epsilon proteobacteria (Nothaft and Szymanski, 2013,
67 Whitfield et al., 2017). In contrast, all sequenced Archaea possess the N-glycan
68 oligosaccharyltransferase AglB, with the exceptions of *Aeropyrum pernix* and *Methanopyrus* species
69 (Jarrell et al., 2014, Nikolayev et al., 2020). The common feature of N-glycosylation is the assembly of
70 an oligosaccharide onto a lipid by the action of specific glycosyltransferases (GTs), which use either
71 nucleotide- or lipid-activated sugar donors. The assembled lipid-linked oligosaccharide (LLO) is flipped
72 across the cytoplasmic/ER membrane. The glycan is then either enlarged or directly transferred onto
73 a target protein by an oligosaccharyltransferase (OST). Target proteins contain a specific asparagine
74 residue found within the conserved sequon (N-X-S/T in Eukarya and Archaea or D/E-X_1-N-X_2-S/T in
75 Bacteria, where X is not proline) recognized by the OST.
76 Archaeal N-glycosylation displays a mosaic of bacterial and eukaryal features. Like Bacteria, Archaea
77 use only a single OST subunit, whereas Eukarya require an OST complex of nine subunits, to transfer
78 the LLO onto the target protein. In contrast, Archaea use the shorter version of the recognition sequon
79 (N-X-S/T) as well as dolichol-phosphate (Dol) as lipid, as in Eukarya, whereas Bacteria use
80 undecaprenyl-phosphate. Archaea also display unique N-glycosylation features, *e.g.* the assembly of
81 the N-glycan onto Dol-phosphate in the Euryarchaeota, while the Crenarchaeota use Dol-
82 pyrophosphate as in eukaryotes (Taguchi et al., 2016).
83 Currently, more than 15 different archaeal N-glycan structures have been characterized (Jarrell et al.,
84 2014, Albers et al., 2017). These archaeal N-glycans display unique structures and sugar compositions
85 and based on the limited characterization of archaeal N-glycans, it is plausible to assume that Archaea
86 possess an enormous range of structural and composition diverse N-glycans. However, closely related
87 species *e.g.,* of the thermophilic crenarchaeal Sulfolobales family, retain a conserved N-glycan core
88 structure but differentiated by their terminal sugar residues. The glycosylation pathway of *Sulfolobus*
89 *acidocaldarius*, belonging to Sulfolobales, has been partially characterized. Secreted proteins, such as
90 the cell wall surface-layer (S-layer) protein and motility filament forming protein ArlB, are N-
91 glycosylated with a tribranched hexasaccharide. This N-glycan has the composition: (Glc-β1-4-
92 sulfoquinovoseβ-1-3)(Man α-1-6)(Man α-1-4)GlcNAc-β1-4-GlcNAc-β-Asp (Peyfoon et al., 2010,
93 Zahringer et al., 2000). Interestingly, the N-glycan core structure is composed of a chitobiose (GlcNAc-
94 β1-4-GlcNAc), which is identical to the core structure of all eukaryal N-glycans. Members of the
95 Sulfolobales family are the only Archaea reported to display this eukaryal N-linked chitobiose core
96 structure (van Wolferen et al., 2020, Palmieri et al., 2013). Based on this identical N-glycan core

97 structure, it has been suggested that the *Sulfolobus* N-glycan biosynthesis pathway may be similar to
98 the eukaryal process (Meyer and Albers, 2013). Indeed, a homolog of the yeast Alg7 enzyme, which
99 initiates eukaryal N-glycosylation by transferring a GlcNAc-P from UDP-GlcNAc to Dol-P to generate
100 the precursor Dol-PP-GlcNAc, was identified in *S. acidocaldarius* as AlgH (Meyer et al., 2017). This AglH
101 showed not only a similar topology and highly conserved amino acid motifs to Alg7, but also
102 functionally complemented a conditional yeast Alg7 knockout (Meyer et al., 2017). This is a rare
103 example of cross-domain complementation that further strengthens the hypothesis of a common
104 origin of eukaryotic and archaeal N-glycosylation.
105 Here we report the identification of the β-1,4-N-acetylglucosamine-transferase Agl24, which shows
106 conserved amino acid motifs and apparent structural similarities with the eukaryal Alg13/14 and
107 bacterial MurG functional homologs. Activity assays, along with HPLC, MALDI, and NMR analyses
108 confirmed the function of Agl24 as catalyzing the second N-glycan biosynthesis step by generating the
109 lipid-linked chitobiose core. Our extensive bioinformatics analyses support the hypothesis that the
110 eukaryotic N-glycosylation pathway originates from an archaeal ancestor.
111
112

113 **Results**

114 **Identification of a Dol-PP-GlcNAc UDP-GlcNAc GlcNAc GT candidate in *S. acidocaldarius***

115

116 To identify the glycosyltransferase that conducts this first committed synthetic step in the N-
117 glycosylation process in *S. acidocaldarius*, we searched for similar biosynthesis processes to identify
118 potential homologs. The bacterial enzyme MurG transfers a GlcNAc residue from UDP-GlcNAc to the
119 C4 hydroxyl group of MurNAc-(pentapeptide)-PP-undecaprenol to produce the lipid-linked β-(1,4)
120 disaccharide known as lipid II (Ha et al., 1999, Men et al., 1998). Although MurG does not synthesize
121 chitobiose like *S. acidocaldarius*, the enzyme uses a similar lipid-linked acceptor and an identical sugar
122 donor (UDP-GlcNAc) and therefore presents a valid candidate for evaluation. MurG has been classified
123 by the Carbohydrate Active enZyme database (http://www.cazy.org/) as a family 28 GT(Lombard et
124 al., 2014). Up to now, only members of the archaeal families *Methanobacteriaceae* and
125 *Methanopyraceae*, which synthesize pseudomurein as their cell wall structure (Albers and Meyer,
126 2011, Meyer and Albers, 2020), were found to have a GT28 family homolog in the CAZy database
127 ((Lombard et al., 2014).
128 Another candidate homolog is the eukaryal protein complex formed by Alg13 and Alg14 that transfers
129 GlcNAc from UDP-GlcNAc to the C4 hydroxyl of GlcNAc-PP-dolichol to produce the β-(1-4) chitobiose
130 core of the eukaryal N-glycan. Alg13 and Alg14 are classified as GT 1 family members but are distantly
131 related to GT28 family enzymes (Lombard et al., 2014). Alg13 and Alg14 form the C- and N-terminal
132 domains, respectively, of a functional heterodimeric GT enzyme (Fig 2); Alg14 contains the acceptor-
133 binding site and the membrane association domain, while Alg13 possesses the UDP-GlcNAc-binding
134 pocket. The interaction of these two enzymes is required for the catalytic activity, as the active site lies
135 in the cavity formed by the two proteins (Gao et al., 2005, Bickel et al., 2005, Chantret et al., 2005).
136 Although members of the GT 1 family are most prominent in Archaea, until now no GT1 family member
137 has been classified in Sulfolobales. There is no meaningful shared sequence identity between Alg14-
138 13 and MurG and they are subsequently annotated in different GT families. However, both enzymes
139 use UDP-GlcNAc to GlcNAcylate a lipid-linked sugar acceptor and therefore represent valid candidates
140 for a bioinformatics search in *S. acidocaldarius.*
141 Using the *E. coli* MurG protein sequence (WP_063074721.1; EC 2.4.1.227) as template for a Delta Blast
142 search (Boratyn et al., 2012), various homologs were identified in *S. acidocaldarius* with very low
143 sequence identity of only 10–17% (Table S1). The anchoring subunit Alg14 from *Saccharomyces*
144 *cerevisiae* S288C (NP_009626.1) was used as bait for a homology search but no protein candidate was
145 identified below the threshold of 0.005. However, when either the catalytic subunit Alg13 from
146 *Saccharomyces cerevisiae* S288C (NP_011468.1; EC 2.4.1.141) or the fusion of both Alg14 and Alg13
147 sequences was used as a search model, a single Sulfolobales homolog was identified with 17%

148 sequence identity (Table S1). Thus, the most likely candidate for the β-1,4-N-acetylglucosamine
149 transferase was found to be Saci1262 (Uniprot: Q4J9C3), now named Agl24, a hypothetical 327 amino
150 acid protein. In contrast to Euryarchaeota, where the genes of the N-glycosylation enzymes are
151 clustered with *aglB,* clustering of these genes is uncommon in Crenarchaeota (Kaminski et al., 2013,
152 Nikolayev et al., 2020), which makes the identification of GTs involved in the N-glycosylation process
153 more challenging, as in the Sulfolobales. Interestingly, the gene *saci1262 / agl24* is located only eight
154 genes downstream of the OST *aglB in S. acidocaldarius.* (Fig S1).
155
156
157 **Detailed bioinformatics comparison of Agl24 with Alg13/14 and MurG**
158
159 The overall topology of Agl24 is identical to MurG, which consist of two domain folds separated by a
160 deep cleft corresponding to a GT-B fold structure (Ha et al., 2000, Hu et al., 2003) (Fig 1C). Sequence
161 and domain comparisons of the eukaryal Alg13 and Alg14 to MurG and Agl24 reveal that Alg14
162 corresponds to the N-terminal part of MurG/Agl24, while Agl13 resembles the C-terminal part of
163 MurG/Agl24 (Fig 1A). Both N-terminal (aa1–152) and C-terminal (aa153–327) parts of Agl24 show very
164 low sequence identity to Alg14 and Alg13, with 15% and 16%, respectively. Interestingly, in lower
165 eukaryotes, such as Leishmania and Trypanosoma, Alg14 and Alg13 enzymes are present as a single
166 protein fusion, similar to *S. acidocaldarius* (Averbeck et al., 2007). The question arises whether the
167 ancestral enzyme was fused and split into two domains during evolution. Strikingly, the opposite
168 domain orientation is found *e.g.,* in *Entamoeba histolytica* (Bickel et al., 2005) and *Dictyostelium*
169 *discoideum*. An interesting difference in these proteins from the three domains of life is the presence
170 of an N-terminal transmembrane domain (TMD) in most of eukaryal Alg14 proteins (aa 5–25), which is
171 missing in bacterial MurGs and archaeal Agl24 enzymes. The TMD is predicted to facilitate the
172 association of Agl14 with the ER membrane and to participate in recruitment of Dol-PP-GlcNAc to the
173 membrane (Lu et al., 2012). The introduction of this TMD appears to be derived later in the evolution
174 of eukaryotes, and the interaction with the ER or cytoplasmic membrane is not solely dependent on a
175 TMD, as the presence of hydrophobic patches and protein–protein interactions with membrane
176 proteins are also proposed to facilitate this interaction. Indeed, in lower Eukarya, *e.g.* in *D. discoideum*
177 and *E. histolytica,* having the exchanged domain orientation, lacking an N-terminal TMD (Averbeck et
178 al., 2007).
179
180 The amino acid sequence alignment of the representative bacterial MurG, eukaryal Alg14-13, and
181 archaeal homologs revealed conservation of specific patches and individual residues (Fig 1B and 1D).
182 Despite the low sequence identity, a similar structural topology of Agl24 to Agl14/13 and MurG is not
183 surprising, as GTs in general show only a limited number of structural folds. GTs mainly possess a GT-
184 A or GT-B fold, whereas oligosaccharyltransferases and a number of recently characterized GTs display

185 a GT-C, GT-D, or GT-E folds, respectively (Mestrom et al., 2019). Crystal structures of the bacterial
186 MurG (PDB:3s2U, 1F0K, 1NLM) and the C-terminal half of the eukariotic Alg13 (PDB:2jzc) show typical
187 GT-B structural characteristics (Brown et al., 2013, Ha et al., 2000, Hu et al., 2003, Wang et al., 2008,
188 Raman et al., 2010) (Fig 2). The N-terminal part of Alg13 contains a Rossmann-like fold with a mixed
189 parallel and antiparallel β-sheet rather than the conventional Rossmann fold found in all GT-B
190 enzymes, indicating a unique topology among glycosyltransferases(Wang et al., 2008). The N-terminal
191 helix of Agl14 and MurG is proposed to be the membrane association domain (Chantret et al., 2005,
192 Lu et al., 2011). We have generated an Agl24 structural model to investigate potential features via
193 SWISS-MODEL (Waterhouse et al., 2018) (Fig 2). Previous studies have compared the structures of
194 MurG and Agl13, including a model for Alg14 (Gao et al., 2008). The formation of the Agl14/Alg13
195 complex is mediated by a short C-terminal α-helix of Alg13 in cooperation with the last three amino
196 acids of Alg14 (Gao et al., 2008). In the Alg13 crystal structure this C-terminal helix is associated with
197 itself, as Agl14 is missing. The orientation is predicted to be similar to the linker in MurG and the change
198 is indicated by the black arrow (in Fig 2A) (Gao et al., 2008). The structural model of Agl24 is also
199 predicted to contain a C-terminal helix, which interacts with the N-terminal part of the protein (Fig 2A).
200 We revealed a conserved GGxGGH$_{14}$ motif (Agl24) within the N-terminal sequences of MurG and Alg14
201 (Fig 1D, S2, and 2B). This motif is located in the cavity between the two different structural domains
202 next to the substrate binding pocket. The crystal structure of MurG (PDB: 3s2u) in complex with UDP-
203 GlcNAc shows the close proximity of the sugar donor to the conserved residues (Fig 2B) (Brown et al.,
204 2013). In proximity to the GGxGGH$_{14}$ motif is the conserved glutamic acid (E$_{114}$) (Fig 1D, S2, and 2B).
205 Interestingly, both these conserved residues are absent in the archaeal MurG-like GT28 family
206 orthologs of pseudomurein-producing Archaea, where aspartic acid (D) and leucine (L) are more
207 frequently found replacing the H$_{14}$ and E$_{114}$ (Fig 1D and S2). We propose that these residues might be
208 required to accommodate the different acceptor molecule.

209 **Agl24 is essential in *S. acidocaldarius***

210 Due to the overall low sequence similarity of Agl24 to Alg14-13 and MurG, we proposed to
211 demonstrate the predicted function of Agl24 *in vivo* by generating a deletion mutant of *Agl24* in the
212 genome of *S. acidocaldarius*. This deletion mutant should arrest the N-glycosylation process after the
213 generation of Dol-PP-GlcNAc and should contain either non-glycosylated glycoproteins or N-
214 glycosylated proteins containing only a single GlcNAc residue linked to the asparagine in the conserved
215 N-glycosylation motif. The genomic integration by homologous recombination of the plasmid
216 pSVA1312 via either the up- or downstream region of the *agl24* and the selection *pyrEF* genes was
217 confirmed by PCR (Fig S3A). However, we were not successful in generating the marker-less in-frame
218 *agl24* deletion mutant. An alternative approach was conducted, aimed to delete the *agl24* gene by a

219 single homologous recombination step by the integration of the *pyrEF* selection cassette. All attempts
220 to generate this gene disruption mutant failed. Therefore, this strongly suggests that the *agl24* gene is
221 essential in *S. acidocaldarius*, at least under the conditions tested. To exclude that the other GTs,
222 identified in the bioinformatic homology search (Table S1), are involved in the N-glycosylation, marker-
223 less deletion mutants of *saci1907*, *saci1921*, *saci0807, saci1094*, *saci1201*, *saci1821*, and *Saci1249*
224 were successfully generated. Only the deletion of *saci0807* showed a significant effect on the N-
225 glycosylation of the S-layer proteins, which have been characterized to encode the GT Agl16 that
226 transfers the terminal glucose residue to the N-glycan (Meyer et al., 2013).
227

228 **Agl24 transfers a single GlcNAc residue onto a GlcNAc acceptor**
229

230 In agreement with the observed membrane association in *S. acidocaldarius* (Fig S4), Agl24 was also
231 found in the membrane fraction during its purification from *E. coli*. Recombinant Agl24, produced in *E.*
232 *coli* (Fig S5), was assayed to test the function of Alg24 using the predicted nucleotide sugar donor UDP-
233 GlcNAc and either one of two synthetic acceptor substrates designed to mimic the native lipid-linked
234 acceptor: $C_{13}H_{27}$-PP-GlcNAc (acceptor 1) or phenyl-O-$C_{11}H_{22}$-PP-GlcNAc (acceptor 2) (Fig 3C). The
235 reaction product was purified and characterized. The MALDI–MS spectra obtained confirmed that
236 Agl24 transfers a single GlcNAc to both acceptor substrates when incubated with UDP-GlcNAc (Fig 5B
237 and D) (Fig 3A). The Agl24 assay revealed that the product peaks are shifted by 203 Da to m/z = 811
238 $[M-1H+2Na]^+$ and m/z = 833 $[M-2H+3Na]^+$ (Fig 3B), corresponding to one dehydrated GlcNAc (203 Da).
239 The same shift was observed for acceptor 2 (Fig 3D). We also investigated Agl24 promiscuity towards
240 utilizing UDP-Glucose as the sugar donor, but no mass shift was observed (Fig 3E), indicating that Agl24
241 uses exclusively UDP-GlcNAc.

242

243 **Activity of Agl24**

244 Since *S. acidocaldarius* is a thermophilic microorganism, with an optimal growth temperature of 75°C,
245 the temperature dependency of the Agl24 activity was investigated using our established mass
246 spectrometry assay. The substrates are stable at the conditions tested and MALDI analysis of the
247 negative control lacking the enzyme revealed only the acceptor-1 mass (Fig 4A). At elevated
248 temperatures, the peak intensity from the product increased while the peak intensity of the acceptor
249 molecule was reduced (Fig 4B-F). Highest activity was detected at 70°C, close to the optimal growth
250 temperature of *S. acidocaldarius*. Furthermore, the addition of EDTA did not affect the activity of
251 Agl24, demonstrating that Agl24 is a metal independent GT (Fig S7, 5D). This result is in agreement
252 with the lack of a conserved Asp-X-Asp (DxD) motif within Agl24 sequence, which has been shown to

253 be important to coordinate the metal ions in A-fold GTs, whereas GT-B GTs are metal ion-
254 independent(Lairson et al., 2008, Gloster, 2014).

255 **Conserved His14 is essential for Agl24 function**

256 Two conserved amino acid residues were targeted by mutagenesis to investigate their role in the Agl24
257 enzyme. Alanine substitution of histidine residue H_{14} within the conserved GGH_{14} motif, found in all
258 Alg14 (GSGGH), MurG (GGxGGH), and Agl24 (GGH) homologs, was inactive (Fig 5B). This demonstrated
259 that this highly conserved residue, located next to the nucleotide-binding site, is important for the
260 enzyme function. This GGxGGH motif, and a subsequent second glycine rich motif (GGY), have been
261 proposed to enable MurG to be involved in interaction with the diphosphate group of the lipid
262 acceptor, as these two motifs resemble phosphate-binding loops of nucleotide-binding proteins (Baker
263 et al., 1992, Carugo and Argos, 1997). Alanine substitution of the conserved His residue in MurG
264 resulted in undetectable activity and loss of ability to complement a temperature sensitive MurG
265 mutant (Crouvoisier et al., 2007, Hu et al., 2003). In contrast, the substitution of the second conserved
266 residue E_{114}, opposite the nucleotide-binding site, had no effect on the activity of Agl24 (Fig 5C). In
267 MurG this glutamic acid residue is found in a conserved $H_{124}EQN_{127}$ motif (*E. coli*), proposed to
268 coordinate the lipid acceptor (Crouvoisier et al., 2007, Hu et al., 2003). The MurG $E_{125}A$ variant was able
269 to complement the thermosensitive MurG *E.coli* strain, but the remaining activity was 860-fold lower
270 than that of the wild-type MurG (Crouvoisier et al., 2007).

271
272
273 **Agl24 is an inverting β-1,4-N-acetylglucosamine-transferase**
274
275 For kinetic analysis of Agl24, a HPLC assay was used to monitor conversion of the mono-GlcNAcylated-
276 PP-lipid acceptor-2 substrate to the bi-GlcNAcylated product (Fig S6). A K_m value for acceptor-2 could
277 not be determined using this system due to significant substrate inhibition above 0.5 mM, however, a
278 $K_m^{(app)}$ for UDP-GlcNAc was determined: ${}^{(UDP\text{-}GlcNAc)}K_m$ = 1.37 ± 0.13 mM, V_{max} = 32.5 ± 0.9 pmol min^{-1}.
279 These results are in agreement with the K_m values of other GT enzymes, which typically exhibit K_m
280 affinities for their respective nucleotide sugars in the high micromolar range, reflecting the estimated
281 intracellular concentrations of the nucleotide sugar (Varki et al., 1999). A ^{1}H NMR analysis confirmed
282 that the enzymatic product contained two GlcNAc residues with different linkage types based on the
283 presence of two differently-coupled anomeric protons (Fig 6 and S8). One anomeric proton appeared
284 as a doublet of doublets (5.35 ppm, $J_{H1\text{-}H2}$ = 3.1 Hz, $J_{H1\text{-}P}$ = 7.2 Hz) typical of an α-linked GlcNAc residue
285 (Table 1). Another anomeric signal appeared as a doublet with a large *J* coupling value (4.49 ppm, $J_{H1\text{-}H2}$ = 8.5 Hz), indicative of a β-linked GlcNAc residue. The substrate acceptor 2 contains an α-linked
287 GlcNAc; this strongly suggested the terminal GlcNAc introduced by Agl24 was β-linked. Experiments

288 using 1D total correlation spectroscopy (TOCSY) deciphered the detailed proton signals from each of
289 the two sugar rings (Fig S8), and 2D COSY (Fig S9). In addition, HSQC experiments (Fig S10) were used
290 to assign the identity of each proton and carbon signal. The C4 signal of the α-GlcNAc residue is shifted
291 to 79.6 ppm, strongly suggesting that the terminal β-GlcNAc was linked at this position. This was
292 further supported by a relative increase in shift of the H4 proton of the α-GlcNAc residue compared to
293 the un-modified acceptor (Zorzoli et al., 2019). Relative shifts of other protons reported for the
294 acceptor aligned with our experimental data. In conclusion, a combination of ^{1}H NMR and 1D and 2D
295 TOCSY, COSY and HSQC experiments confirmed that Agl24 is an inverting β-1,4-GlcNAc transferase.

296 **The eukaryotic GTs Alg14 and Alg13 are closely related to Asgard homologs**

297 In order to analyze the phylogenetic relationship between the eukaryotic N-glycosylation GTs Alg13
298 and Alg14 with archaeal and bacterial homologs, an extensive phylogenetic analysis was performed.
299 The eukaryotic Alg13 and Alg14 sequences cluster with homologs from the Asgard phyla
300 Thorarchaeota and Odinarchaeota, as well as with sequences from Verstraetearchaeota,
301 Micrarchaeota, Geothermarchaeota, and an unclassified archaeon (Fig7 A and B). Even though the
302 sister clades of Eykaryotes are Thorarchaeota and Odinarchaeota for Alg13 and Alg14, respectively,
303 the phylogeny is not very well resolved. No closely-related Alg13- and Alg14-like sequences were found
304 in other Asgard phyla, suggesting that the remaining Asgardarchaea, *e.g.* Lokiarchaeota and
305 Heimdallarchaeota, could be using a different enzyme for synthesizing N-glycans.
306 Next to the Alg13- or Alg14-like cluster, is a second distinct cluster of EpsF- or EpsE-like sequences,
307 respectively. The function of EpsF and EpsE have been studied in Lactobacillales, where these enzymes
308 are involved in exopolysaccharide production (De Vuyst et al., 2001, Kolkman et al., 1997). Here, similar
309 to Alg14/13, the combined activity of EpsE and EpsF link either a glucose (Glc) from UDP-Glc (van
310 Kranenburg et al., 1999, Kleerebezem et al., 1999), or a galactose (Gal) from UDP-Gal via a β-1,4 linkage
311 to lipid-linked glucose acceptors (van Kranenburg et al., 1999), creating either a lipid-linked cellobiose
312 or a lactose. The archaeal Eps-like sequences are mainly found in diverse Euryarchaeota, whereas the
313 bacterial sequences are scattered among different phyla, possibly as a result of extensive horizontal
314 gene transfers.
315 The remaining clades in the Alg13 and Alg14 phylogenies consist of sequences from the archaeal TACK
316 superphylum and some Euryarchaeota, mainly methanogens (Fig 7A and B). The Agl24 identified in this
317 study is found among them, within a Cren- and Bathyarchaeota branch. Close homologs of the Agl24
318 are found in all presently available genomes of the Crenarchaeota, *e.g.* in the orders Fervidicoccales,
319 Acidilobales, Desulfurococcales and Sulfolobales, with the exception of any homolog in the members
320 of the order Thermoproteales (Fig S11, Sup Data 1).

321 The remaining phyla of the TACK superphylum (Aig-, Bathy-, Geotherm-, Kor-, Thaum-, Nezha-, Mars-,
322 and Verstraetearchaeota) along with some additional Crenarchaeota, including Geoarchaeota, and an
323 Odinarchaeota sequence (Fig 7A and B), form a separate branch containing Agl24-like sequences. Next
324 to this branch is a clade of methanogens and Altiarchaeales. Due to the large number of very distant
325 Alg13/14 homologs with low similarity, resulting in poor alignments, for the phylogenies in Fig 7 we
326 sought to imitate and expand the clades presented previously (Lombard, 2016). Nevertheless, in the
327 homology searches used to expand the previous set (Lombard, 2016), some of the hits formed well-
328 supported branches in preliminary phylogenies (Sup Data 1). By including some of them in expanded
329 Alg13/EpsF and Alg14/EpsE datasets, none of them branch within the Alg13/EpsF-like or Alg14/EpsE-
330 like clades. Instead, they are fused Agl24-like sequences, mainly from Euryarchaeota and various
331 DPANN phyla. Either way, the Cren-/Bathyarchaeota Agl24 clade along with a small
332 Nanohaloarchaeota clade appear to be the closest relatives of the Alg13/EpsF-like and Alg14/EpsE-like
333 proteins (Sup Data 1).
334 The fact that we are able to trace Agl24-like homologs to the ancestor of TACK and several
335 Euryarchaeota lineages indicates that, despite multiple instances of lateral gene transfer, they are very
336 ancient and diverged from the bacterial MurG at the separation of the two domains, as proposed by
337 Lombard *et al.* (Lombard, 2016). Given that both MurG (Lombard, 2016) and the various Agl24-like
338 homologs are fused genes, a major split event seems to have occurred once at the base of the Alg/Eps-
339 like clade (Fig 8A and B, red arrow), with some independent, more recent, fusions found throughout
340 eukaryotes and Archaea (*e.g.* the Thorarchaeota RLI55694.1). Finally, recovering the monophyly of
341 eukaryotes in our phylogenies corroborates the proposition by Lombard *et al.* (Lombard, 2016) that
342 Alg13 and Alg14 were present in the last eukaryotic common ancestor (LECA). Since the sister clade of
343 eukaryotes is in both cases an Asgard phylum (although not the Heimdallarchaeota), it is possible that
344 these genes were inherited vertically during eukaryogenesis in a two-domain tree of life (Eme et al.,
345 2017).
346
347 **Discussion**
348 In this study we identified the first archaeal and thermostable β-1,4-N-acetylglucosaminyltransferase,
349 named archaeal glycosylation enzyme 24 (Agl24), responsible for the synthesis of the lipid-linked
350 chitobiose on which the N-glycan of *S. acidocaldarius* is assembled. The reported essentiality of Agl24
351 is in agreement with the indispensable properties of the OST AglB and AglH, catalyzing either the last
352 or first step of the N-glycosylation process in *S. acidocaldarius* (Meyer and Albers, 2014, Meyer et al.,
353 2017). The essentiality of the N-glycosylation has also been confirmed by a transposon study in
354 *Sulfolobus islandicus*, a close relative to *S. acidocaldarius*, in which all homologous genes, including
355 *agl24*, lack any transposon insertions (Zhang et al., 2018). However, the fundamental nature of

356 archaeal N-glycosylation cannot be generalized since deletion mutants of the
357 oligosaccharyltransferase AglB can be obtained in Euryarchaeota, *e.g.,* in *Methanococcus maripaludis*
358 and *Haloferax volcanii* (Abu-Qarn et al., 2007, Chaban et al., 2006). The requirement for N-
359 glycosylation might rely on many factors including the N-glycan composition and structure as well as
360 the modification frequency on proteins; this frequency is reported to be higher in thermophilic,
361 compared to mesophilic, Archaea (Meyer and Albers, 2013, Jarrell et al., 2014).
362 The comparison of Agl24 with eukaryotic Alg14/13 and the bacterial MurG enzymes revealed structural
363 and amino acid similarities. Although the sequence identity of Agl24 with the eukaryotic and bacterial
364 enzymes is extremely low (~16%), the structural prediction of Alg24 indicated a MurG–like GT-B fold.
365 Interestingly, the amino acids of the nucleotide sugar donor binding site of MurG (Hu et al., 2003) are
366 not conserved in Agl24 and Alg14, although all enzymes use UDP-GlcNAc as the sugar donor (Fig S2,
367 boxed). Strikingly, a conserved GGH_{14} motif within an N-terminal loop of MurG, Alg14-13, and Agl24
368 (Fig 1 and S2), which protrudes into the cavity formed by the two structural folds, could be identified.
369 In Bacteria this $G_{13}GTGGH_{18}$ loop is extended by a second $G_{103}GY_{105}$ loop. As both loops are suggested
370 to be reminiscent of the phosphate-binding loops of nucleotide-binding proteins, their involvement in
371 the binding of the diphosphate group of the acceptor lipid has been proposed (Crouvoisier et al., 2007).
372 In agreement with our results, the alanine substitutions of the invariant bacterial H_{14} led to an
373 extremely low or undetectable activity of MurG as well as the loss of ability to complement a
374 temperature sensitive MurG mutant (Crouvoisier et al., 2007, Hu et al., 2003). However, up to now, no
375 structure of MurG or Agl14-13 with the bound lipid acceptor molecule has been published and the
376 interaction of the imidazole ring of the histidine residue with the acceptor molecule remained to be
377 shown. Our alignment also reveals a conserved glutamic acid in the bacterial, eukaryotic, and archaeal
378 sequences. In bacteria, this glutamic acid is found in a conserved $H_{124}EQN_{127}$ loop (*E. coli* MurG), which
379 has been proposed to form a key part of the acceptor-binding site (Hu et al., 2003). In agreement with
380 our study, an $E_{125}A$ substitution significantly affected the function of MurG, although a marked
381 decrease in lipid acceptor binding (1760-fold) and nucleotide substrate binding (210-fold) was
382 observed (Hu et al., 2003, Crouvoisier et al., 2007).
383 Although lacking a TMD, Agl24 associates with the membrane, indicating interaction with membrane
384 enzymes, similar to the interaction of MurG or Alg13/14 with their membrane interaction partners
385 MraY and Alg7 respectively. Based on the structural similarities of these complexes, a shared common
386 evolutionary origin of the first enzymes of eukaryotic N-glycosylation and the bacterial peptidoglycan
387 biosynthesis has been hypothesized (Bugg and Brandish, 1994, Burda and Aebi, 1999, Bouhss et al.,
388 2008). As *S. acidocaldarius* possesses a homolog of Alg7, termed AlgH (Meyer et al., 2017), and a
389 bioinformatics study has indicated that Alg7 might have emerged out of the Crenarchaeota (Lombard,
390 2016), AglH presents the most likely candidate for the interaction partner of Agl24.

391 A phylogenetic analysis shed light onto the distribution of the identified Agl24 in Archaea, and
392 determined its relationship with eukaryotic and bacterial homologs. This revealed a very ancient and
393 broad distribution of Agl24-like sequences in Archaea. Importantly, close homologs of Agl24 are only
394 found within the Crenarchaeota and Barthyarchaeota. Indeed, all Crenarchaeota, with the exception
395 of the order Thermoproteales, possess an Agl24 homolog, strongly suggesting that these
396 Crenarchaeota contain a chitobiose N-glycan core structure. In contrast, the N-glycans of members of
397 the Thermoproteales would be assembled using a different enzyme. Indeed, the N-glycan of
398 *Pyrobaculum caldifontis*, belonging to the order Thermoproteales, differs from the N-glycan in
399 Sulfolobales (Fujinami et al., 2017). Here, the N-glycan core is composed of two di-N-acetylated β-1,4-
400 linked GlcNAc residues with the second sugar carboxylated at C6 ($GlcA(NAc)_2$-β-1,4-$(Glc(NAc)_2$. A di-N-
401 acetylated glucuronic acid sugar is also found in the N-glycan from *Methanococcus voltae,* which is
402 transferred by the GT AglC onto the Dol-P linked GlcNAc (Chaban et al., 2009). Thus, a homolog of AglC
403 is likely to be involved in the N-glycan biosynthesis in *P. caldifontis*.
404 Our phylogenetic analysis revealed that the closest related enzymes of the eukaryotic Alg14 and Alg13
405 are found within the Asgard superphylum. Since the discovery of the Archaea as the third domain of
406 life next to Bacteria and Eukarya (Woese and Fox, 1977), the comparison of archaeal molecular
407 processes has gradually revealed a strong resemblance to those found in eukaryotes (Zillig et al., 1989,
408 Huet et al., 1983). Furthermore, recent phylogenetic analysis indicated that eukaryotes originated
409 close to the TACK superphylum (Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota)
410 and the Asgard superphylum (Spang et al., 2015, Zaremba-Niedzwiedzka et al., 2017, Petitjean et al.,
411 2015, Koonin, 2015). The current hypothesis for the genesis of the first eukaryotic cell proposes a
412 membrane protrusion of an ancestral archaeon that engulfed a bacterium to increase the interspecies
413 interaction surface (Baum and Baum, 2014, Imachi et al., 2020). The eukaryotic N-glycosylation would
414 therefore be inherited from this ancient archaeal ancestor, which is in agreement with the high
415 conservation of the initial N-glycosylation biosynthesis steps in all eukaryotes (Helenius and Aebi, 2004,
416 Samuelson et al., 2005, Schwarz and Aebi, 2011). In addition, the AglB in members of the Asgard and
417 TACK superphyla shows considerable similarity to its eukaryotic homologue Stt3, and is clearly distinct
418 from the AglB of the DPANN superphylum and the Euryarchaeota (Nikolayev et al., 2020). Most
419 eukaryotic Stt3 contain a double sequon motif for N-glycosylation near the substrate-binding site,
420 which is hypothesized to contribute to the LLO (Shrimal and Gilmore, 2019, Wild et al., 2018).
421 Interestingly, this double sequon is also present in the AglB of the members of the Asgard and TACK
422 superphyla but absent in the AglB of Euryarchaeota (Shrimal and Gilmore, 2019). In addition, homologs
423 of non-catalytic subunits of the eukaryotic Stt3 complex, *e.g.,* ribophorin, Ost3/6, Ost5, and Wbp1,
424 have been detected in some Arsgardarchaeota (Zaremba-Niedzwiedzka et al., 2017). However, their
425 specific function in the N-glycosylation has to be shown, especially as ribophorin is only found in higher

426 mammals (Wilson and High, 2007) and therefore unlikely to exist in an early eukaryotes. Nevertheless,
427 the assembly of the N-glycan linkage in the TACK superphylum is conducted via a pyrophosphate
428 linkage to the Dol, identical to that in Eukarya. This is in contrast to Euryarchaeota, which use a mono-
429 phosphate linkage. A summary of the similarities and differences of the archaeal N-glycosylation to
430 Eukarya and Bacteria are given in Table S4. All these observations strengthen the hypothesis that the
431 eukaryotic N-glycosylation has emerged from an ancient archaeon.
432 In the future, we believe that the detailed characterization of the N-glycosylation process in members
433 of the TACK and Asgard superphyla will lead to the elucidation of further molecular similarities and
434 unique properties to the eukaryotic N-glycosylation process and will provide further support for the
435 origin of the eukaryotic N-glycosylation.

Materials and methods

Strains and growth conditions

436 The strain *Sulfolobus acidocaldarius* MW001 (Δ*pyrE*) (Wagner et al., 2012) and all derived modified
437 strains (Table S3) were grown in Brock medium at 75°C, pH 3, adjusted using sulfuric acid. The medium
438 was supplemented with 0.1% w/v NZ-amine and 0.1% w/v dextrin as carbon and energy source (Brock
439 et al., 1972). Selection gelrite (0.6%) plates were supplemented with the same nutrients with the
440 addition of 10 mM $MgCl_2$ and 3 mM $CaCl_2$. For second selection plates 10 mg ml^{-1} uracil and 100 mg
441 ml^{-1} 5-fluoroorotic acid (5-FOA) were added. For the growth of the uracil auxotrophic mutants, 10 mg
442 ml^{-1} uracil was added to the medium. Cell growth was monitored by measuring the optical density at
443 600 nm. Protein expression in *S. acidocaldarius* was conducted in medium supplemented with 0.1%
444 w/v NZ-amine and 0.1% w/v L-arabinose as carbon and energy source. All *E. coli* strains DH5α, BL21,
445 or ER1821 were grown in LB media at 37°C in a shaking incubator at 200 rpm. According to the
446 antibiotic resistance in the transformed vector(s), media were supplemented with the antibiotics
447 carbenicillin (amp) at 100 µg ml^{-1} and/or kanamycin (kan) at 50 µg ml^{-1}.

Construction of deletion plasmids

The predicted function of *Agl24* was verified via the generation of the lipid-linked chitobiose core of the *N*-glycan. Therefore, a marker-less deletion mutant of this gene was constructed in the *S. acidocaldarius* MW001, as previously described (Wagner et al., 2012). Briefly, the strain MW001, auxotrophic for uracil biosynthesis, was transformed with the plasmid pSVA1312. Therefore, two ~1000 bp DNA fragments, one from the upstream and one from the downstream regions of *Agl24* (*saci1262*) gene, were PCR amplified. Restriction sites *Apa*I and *Bam*HI were introduced at the 5´ ends of the upstream forward primer (4168) and of the downstream reverse primer (4165), respectively. The upstream reverse primer (4163) and the downstream forward primer (4164) were each designed to incorporate 15 bp of the reverse complement strand of the other primer, resulting in a 30 bp overlap stretch. The overlapping PCR fragments were purified, digested with *Apa*I and *Bam*HI, and ligated in the digested plasmid pSVA407, containing *pyrEF* (Wagner et al., 2012).

Generation of a linear *Agl24*$_{up}$-*pyrEF*-*Agl24*$_{down}$ fragment for the direct *Agl24*::*pyr*EF replacement

To further underline the essential role of *Agl24* in *S. acidocaldarius*, a disruption of the *Agl24* gene by direct homologous integration of the *pryEF* cassette was performed. For this approach, 387 bp of the *Agl24* upstream region, the full 1525 bp of the *pyrEF* cassette, and 1011 bp of the *Agl24* downstream region were PCR amplified using the primers: 4168+6336, 4115+4116, and 6337+6338, respectively.

At the 5´ ends of the upstream forward primer (4168) and of the downstream reverse primer (6338) restriction sites *Apa*I and *Bam*HI were introduced, respectively. The *pyrEF* forward primer (4115) was designed to incorporate 40 bp of the upstream reverse primer (6336) resulting in a 40 bp overlapping stretch. The *pyrEF* reverse primer (4116) and the downstream forward primer (6337) weres designed to incorporate the reverse complement strand of the other primer, resulting in a 46 bp overlapping stretch. The upstream, *pyrEF*, and downstream fragments were fused by an overlapping PCR, using the 3' ends of each fragment as primers. The 2904 bp overlap PCR fragment gained, was amplified using the outer primers (4168 and 6338), digested with *Apa*I and BamHI and ligated into the p407 vector predigested with the same restriction enzymes. The resulting plasmid, pSVA3338 (*Agl24*$_{up}$-*pyrEF*-*Agl24*$_{down}$), was transformed into *E. coli* DH5α and selected on LB-plates containing 50 mg ml^{-1} ampicillin. The accuracy of the plasmid was verified by sequencing. Before transformation in *S. acidocaldarius,* the plasmid was digested with *Apa*I and *Bam*HI to create the linear *Agl24*$_{up}$-*pyrEF*-*Agl24*$_{down}$ fragment.

448 **Generation of Agl24 Expression plasmids for the production of Agl24 in *S. acidocaldarius* and *E. coli***

449 Cellular localization studies were carried out using N- and C-terminal fusion proteins of Agl24. The
450 *Agl24* gene was amplified from genomic DNA of *S. acidocaldarius* introducing *NcoI* and *Pst*I restriction
451 sites at the 5'ends of the primer pair 4176 and 4177, respectively. The PCR fragment was cloned into
452 pSVA1481, yielding vector pSVA1336 containing an inducible arabinose promoter and *Agl24*-Strep-
453 His$_{10}$. The plasmid was digested with *Nco*I and *Eag*I and the insert *Agl24*-Strep-His$_{10}$ was ligated into
454 pSVA1481 containing an inducible maltose promoter, yielding pSVA1337. The N-terminal Strep-His$_{10}$-
455 TEV-Agl24 was cloned using the primer pair 4180 and 4181, which incorporated the restriction sites
456 *NcoI* and *NotI*, respectively. The Insert was ligated into pSVA2301 containing an inducible maltose
457 promoter yielding pSVA1339.
458 For heterologous expression of Agl24 in *E. coli,* the plasmid pHD0499 was generated. The GFP-His$_8$-
459 tagged *Agl24* gene was constructed by amplifying the full-length *Agl24* gene from *S. acidocaldarius*
460 genomic DNA with the primer pair A596 and A597. The PCR fragment was cloned in-frame into the
461 vector pWaldo (Waldo et al., 1999) using *Xho*I and *Kpn*I restriction sites, generating a C-terminal fusion
462 with a TEV cleavage site, GFP, and His$_8$ tag. For the generation of Agl24 mutants of the conserved
463 mutants H_{15} and E_{114}, a quick-change mutagenesis PCR was applied using the primers: A597 and A684
464 or A685 and A686, respectively.
465

466 **Transformation and selection of the deletion mutant in *S. acidocaldarius***

467 Generation of competent cells was performed based on the protocol of Kurosawa and Grogan
468 (Kurosawa and Grogan, 2005). Briefly, *S. acidocaldarius* strain MW001 was grown to an OD_{600} between
469 0.1 and 0.3 in Brock medium supplemented with 0.1% w/v NZ-amine and 0.1% dextrin. Cooled cells
470 were harvested by centrifugation (2000 x g at 4°C for 20 min). The cell pellet was washed three times
471 successively in 50 ml, 10 ml and 1 ml of ice-cold 20 mM sucrose (dissolved in demineralized water)
472 after mild centrifugation steps (2000 x g at 4°C for 20 min). The final cell pellet was resuspended in 20
473 mM sucrose at an OD_{600} of 10.0 and stored in 50 µl aliquots at -80°C. 400–600 ng of methylated
474 pSVA1312 or the linearized *Agl24*$_{up}$-*pryEF-Agl24*$_{down}$ fragment, was added to a 50 µl aliquot of
475 competent MW001 cells and incubated for 5 min on ice before transformation in a 1 mm gap
476 electroporation cuvette at 1250 V, 1000 Ω, 25 mF using a Bio-Rad gene pulser II (Bio-Rad, X, USA).
477 Directly after transformation 50 µl of a 2x concentrated recovery solution (1% sucrose, 20 mM b-
478 alanine, 20 mM malate buffer pH 4.5, 10 mM $MgSO_4$) was added to the sample and incubated at 75°C
479 for 30 min under mild shaking conditions (150 rpm). Before plating, the sample was mixed with 100 µl
480 of heated 2x concentrated recovery solution and twice 100 µl were spread onto gelrite plates
481 containing Brock medium supplemented with 0.1% NZ-amine and 0.1% dextrin. After incubation for
482 5–7 days at 75°C, large brownish colonies were used to inoculate 50 ml of Brock medium containing
483 0.1% NZ-amine and 0.1% dextrin, which were incubated for 3 days at 78°C. Cultures confirmed by PCR
484 to contain the genomically integrated plasmid were grown in Brock medium supplemented with 0.1%
485 NZ-amine and 0.1% dextrin to an OD of 0.4. Aliquots of 40 µl were spread on second selection plates,
486 supplemented with 0.1% NZ-amine and 0.1% dextrin and 10 mg ml^{-1} uracil, were incubated for 5–7
487 days at 78°C. Newly formed colonies were streaked on fresh second selection plates to ensure single
488 colony selection before each colony was screened by PCR for the genomic absence, presence, or
489 modification, of the *Agl24* gene.

490 **Expression of Agl24 in *S. acidocaldarius***

491 Plasmid transformation into *S. acidocaldarius* cells was performed as described above for the deletion
492 mutants. Cells were spread onto gelrite plates containing Brock medium supplemented with 0.1% NZ-
493 amine and 0.1% dextrin. After incubation for 5–7 days at 75°C large brownish colonies were used to
494 inoculate 50 ml of Brock medium containing 0.1% NZ-amine and 0.1% dextrin and incubated for 3 days
495 of 78°C. Presence of the expression plasmid was confirmed by PCR. For induction of expression the
496 strains were grown either in Brock medium supplemented with either 0.1% NZ-amine and 0.1% dextrin
497 ($MAL_{promotor}$) or with 0.1% NZ-amine and 0.1% L-arabinose ($ARA_{promotor}$) to an OD_{600} of 1.0.
498

499 **Expression and purification of recombinant Agl24 protein from *E. coli***

500 For heterologous Agl24 protein expression, 6 x 1 Liter LB medium was inoculated with 10 ml from an
501 overnight BL21 culture previously transformed with pHD0499. Cells were grown at 16˚C overnight in
502 auto-induction medium (Studier, 2014) containing 30 µg ml^{-1} kanamycin. The cells were harvested by
503 centrifugation at 4,200 x g for 25 minutes and used to isolate the membrane fractions. All subsequent
504 purification steps were carried out at 4˚C. Cells were fractionated by passing four times through an
505 Avestin C3 High Pressure Homogeniser (Biopharma, UK), followed by a 20 min low speed spin at 4,000
506 x g. The resulting supernatant was centrifuged at 200,000 x g for 2 h to obtain the membrane fraction
507 and 2–3 g of membranes were routinely used for isolation of Agl24-GFP-His_8 proteins. Samples were
508 solubilized in 18 ml Buffer 1 (500 mM NaCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4 2.7 mM KCl, pH 7.4, 20
509 mM imidazole, 4 mM TCEP) with the addition of 1% n-dodecyl-β-maltoside (DDM) for 2 h at 4˚C. The
510 sample was two-fold diluted with Buffer 1 and centrifuged at 200,000 x g for 2 h. The supernatant was
511 loaded onto a Ni-Sepharose 6 Fast Flow (GE Healthcare) column with 1 ml of prewashed Ni-NTA-beads.
512 The column was washed with 20 ml of wash buffer (500 mM NaCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4
513 2.7 mM KCl, pH 7.4, 20 mM imidazole, 0.4 mM TCEP, 0.03% DDM) and eluted with 5 x 1ml elution
514 buffer (500 mM NaCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO4 2.7 mM KCl, pH 7.4, 250 mM imidazole, 0.4
515 mM TCEP, 0.03% DDM). Elution fractions were combined and imidazole removed using a HiPrep 26/10
516 desalting column (GE Healthcare) equilibrated with Buffer (1x PBS, 0.03% DDM, 0.4 mM TCEP).

517 Protein concentration was determined using a Bradford reaction (Bio-Rad) and purity was confirmed
518 by SDS–PAGE analysis. The concentrated fractions were separated by SDS–PAGE and stained with
519 Coomassie blue. Protein identity was confirmed by tryptic peptide mass fingerprinting, and the level
520 of purity and molecular weight of the recombinant protein was determined by matrix-assisted laser
521 desorption/ionization time-of-flight mass spectrometry (MALDI–TOF). The analysis was provided by
522 the University of Dundee 'Fingerprints' Proteomics Facility.

523 **Agl24 activity**

524 Activity was measured in a 100 µl reaction volume containing 1 mM UDP-D-GlcNAc, 1 mM acceptor
525 molecule, 5 mM $MgCl_2$, and 5 µg of purified Agl24 in TBS Buffer (150 mM NaCl, 50 mM Tris-HCl, pH
526 7.6); reaction was performed in a PCR cycler at 60°C for 12 h.

527 **MS analysis**

528 MALDI–TOF was used to analyze the acceptors and products of the Alg24 *in vitro* assay. The 100ul
529 reaction samples were purified over 100 mg Sep-Pak C18 cartridges (Waters) pre-equilibrated with 5%
530 EtOH. The bound samples were washed with 800 µl of H_2O and 800 µl of 15% EtOH, eluted in two
531 fractions with (a) 800 µl of 30% EtOH and (b) 800 µl of 60% EtOH. The two elution fractions were dried
532 in a SpeedVac vacuum concentrator and resuspended in 20 µl of 50% MeOH. A 1 µl sample was mixed

533 with 1 µl of 2,5-dihydroxybenzoic acid matrix (15 mg ml^{-1} in 30:70 acetonitrile, 0.1% TFA), and 1 µl was
534 added to the MALDI grid. Samples were analyzed by MALDI in an Autoflex Speed mass spectrometer
535 set up in reflection positive ion mode (Bruker, Germany).

536 **HPLC analysis**

537 With the exception of kinetics reactions, which are detailed below, Agl24 reactions were analyzed
538 using a HPLC assay with, typically, 50 µl samples containing 2.5 mM UDP-GlcNAc, 1.5 mM lipid
539 acceptor, and 1.8 µg of purified Agl24 in a TBS buffer containing 5 mM $MgCl_2$ (150 mM NaCl, 50 mM
540 Tris-HCl, pH 7.5). Reactions were left at 60˚C (or alternative temperatures) for the desired time period
541 and quenched with one equivalent of acetonitrile to precipitate Agl24. Following filtration to remove
542 precipitate, reactions were injected onto an XBridge BEH Amide OBD Prep column (130 Å, 5 µM, 10 x
543 250 mm) at a flow rate of 4 ml min^{-1} using a Dionex UltiMate 3000 system (Thermo Scientific) fitted
544 with a UV detector optimized to detect the *O*-phenyl functional group of the acceptor molecule at 270
545 nm. Each run was 35 minutes using running Buffer A (95% acetonitrile, 10 mM ammonium acetate, pH
546 8) and Buffer B (50% acetonitrile, 10 mM ammonium acetate, pH 8). A linear gradient from 20–80%
547 buffer B was performed over 20 min, followed by an immediate drop back to 20% buffer B for the
548 remaining 15 minutes of the run to re-equilibrate the column to starting conditions. The Alg24 reaction
549 substrates and products typically eluted 8–11 minutes into a run. For kinetic analyses, assays were
550 performed in triplicate (with the exception of 1 mM conc. which was only performed in duplicate) and
551 altered to contain a fixed concentration of lipid-acceptor (0.5 mM) whilst varying the concentration of
552 UDP-GlcNAc (0.5, 0.75, 1, 1.5, 3, 6, 9, 12 mM). Reactions were run for 10 hours before being quenched
553 and, after completion, areas of substrate and product peaks were calculated to determine the reaction
554 conversion. Conversion over 10 hours was converted to pmol per minute, and the resulting data were
555 analyzed using GraphPad Prism v8 to generate a Michaelis–Menten curve and resulting kinetic data.

556 **NMR Analysis**

557 The HPLC-purified Agl24 products (0.5mg–2mg) were dried using a Christ RVC 2-25 speed vacuum
558 fitted with a Christ CT 02-50 cold-trap to remove excess acetonitrile, then freeze dried (Alpha 1-2
559 LDplus, Christ) to remove residual water. Products were subsequently dissolved in 600 µl of D_2O and
560 NMR spectra were recorded at 293K. The spectra were acquired on a Bruker AVANCE III HD 500 MHz
561 NMR Spectrometer equipped with a 5 mm QCPI cryoprobe. For 1D TOCSY experiments, H1' was
562 irradiated at 5.35 ppm, H2' was irradiated at 3.90 ppm, and H1" was irradiated at 4.49 ppm (Fig S8). A
563 combination of ^{1}H ^{1}H COSY, 1D TOCSY, and ^{1}H ^{13}C HSQC experiments were used to fully assign the ^{1}H
564 and ^{13}C signals for the Agl24 reaction product. Full ^{1}H and ^{13}C chemical shift assignments can be found
565 in Table 1 and are recorded with respect to the residual HDO signal at 4.7 ppm.

566 **Phylogenetic analysis**

567 The eukaryotic Alg13-Alg14 and the bacterial EpsF-EpsE sequences were taken from the datasets in
568 (Lombard, 2016), since our own databanks (25118 bacterial and 1611 eukaryotic genomes) contained
569 too many and too divergent hits that could possibly render downstream analyses unfeasible. For
570 Archaea, HMM searches were impossible, since they recovered too many and too divergent possible
571 homologs to handle, while DIAMOND searches were not sensitive enough. Ultimately, we searched for
572 homologs with DIAMOND (--more-sensitive -k 0 & default e-value cutoff) (Buchfink et al., 2015) using
573 multiple seeds from (Lombard, 2016) to cover the different clades of archaeal homologs: 1) the *S.*
574 *acidocaldarius* Alg24 from this study; 2) the Alg13/EpsF-like and Alg14/EpsE-like homologs from
575 *Methanothrix soehngenii* (NCBI: AEB67939.1 and AEB67938.1 respectively); 3) AIC14646.1 from
576 *Nitrososphaera viennensis* (TACK clade); 4) ABX11947.1 from *Nitrosopumilus maritimus* (additional
577 Alg13/EpsF-like Thaumarchaeota clade); 5) AAB99267.1 from *Methanocaldococcus jannaschii*
578 (methanogen clade). For the first two, we repeated the homology search with the same parameters
579 using all the hits from the previous round as seeds to ensure we did not miss any sequences from our
580 primary clades of interest. The hits from each search were aligned with MAFFT L-INS-I v7.453 (Katoh
581 and Standley, 2013) and preliminary phylogenies were reconstructed with IQ-TREE 2.0.5 (Minh et al.,
582 2020) (-m TEST -alrt 1000). From these phylogenies we isolated the sequences forming well-supported
583 branches corresponding to the ones in (Lombard, 2016), albeit with increased diversity plus some
584 additional clades of potential interest. The individual datasets (archaeal, bacterial, eukaryotic) were
585 then fused into single Alg13/EpsF-like and Alg14/EpsE-like datasets that were aligned as above and
586 trimmed with ClipKIT (kpigappy) (Steenwyk et al., 2020). Then phylogenies were inferred with IQ-TREE
587 2.0.5 (Minh et al., 2020) under the model selected by ModelFinder (Kalyaanamoorthy et al., 2017), and
588 ultrafast bootstrap (Hoang et al., 2018), aLRT SH-like (Guindon et al., 2010), and approximate Bayesian
589 (Anisimova et al., 2011) branch support tests (-m MFP -bb 1000 -alrt 1000 -abayes). The trees were
590 visualized in iTol (Letunic and Bork, 2019).

591

592 **Author contributions:**

593 B. H. M., S. V. A., and H. C. D. conceptualization; B. H. M. data curation, initial BLAST searches,
594 deletion analysis, protein alignments, protein expression, protein purification, alanine
595 substitutions, activity analyses, MALDI analyses, visualization; B. A. W. HPLC activity analyses and
596 NMR analyses; P. S. A. phylogenetic analyses, B. H. M.., B. A. W., P. S. A. and H. C. D. investigation;
597 B. H. M. and H. C. D. validation; B. H. M. and H. C. D. methodology; B. H. M. writing-original draft;
598 B. H. M., P. S. A., B. A. W. , S. V. A., and H. C. D. writing-review and editing; H. C. D. supervision; H.
599 C. D. and S. V. A. funding acquisition

Acknowledgments

We thank Vladimir I. Torgov, Vladimir V. Veselovsky, and Leonid L. Danilov from the N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Moscow, Russia for sharing the synthetic acceptor molecules. We thank Prof. Alexander J. Probst from the University Duisburg-Essen for fruitful discussions and Mr. Tom Snelling for assay assistance.

Funding

The HCD laboratory (BHM, BAW and HCD) is supported by The Wellcome Trust and Royal Society Grant 109357/Z/15/Z. SVA and BHM have been supported by the SFG grant SFB 1381 (Deutsche Forschungsgemeinschaft (German Research Foundation) under project no. 403222702-SFB 1381)

For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

References

ABU-QARN, M., YURIST-DOUTSCH, S., GIORDANO, A., TRAUNER, A., MORRIS, H. R., 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. *J Mol Biol,* 374**,** 1224-36.

ALBERS, S., EICHLER, J. & AEBI, M. 2017. Archaea. *In:* VARKI, A., CUMMINGS, R. D., ESKO, J. D., STANLEY, P., HART, G. W., AEBI, M., DARVILL, A. G., KINOSHITA, T., PACKER, N. H., PRESTEGARD, J. H., SCHNAAR, R. L. & SEEBERGER, P. H. (eds.) *Essentials of Glycobiology.* Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press

ALBERS, S. V. & MEYER, B. H. 2011. The archaeal cell envelope. *Nat Rev Microbiol,* 9**,** 414-426.

ANISIMOVA, M., GIL, M., DUFAYARD, J. F., DESSIMOZ, C. & GASCUEL, O. 2011. Survey of Branch Support Methods Demonstrates Accuracy, Power, and Robustness of Fast Likelihood-based Approximation Schemes. *Syst Biol,* 60**,** 685-699.

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.

AVERBECK, N., KEPPLER-ROSS, S. & DEAN, N. 2007. Membrane topology of the Alg14 endoplasmic reticulum UDP-GlcNAc transferase subunit. *J Biol Chem,* 282**,** 29081-8.

BAKER, P. J., BRITTON, K. L., RICE, D. W., ROB, A. & STILLMAN, T. J. 1992. Structural consequences of sequence patterns in the fingerprint region of the nucleotide binding fold. Implications for nucleotide specificity. *J Mol Biol,* 228**,** 662-71.

BAUM, D. A. & BAUM, B. 2014. An inside-out origin for the eukaryotic cell. *BMC Biol,* 12**,** 76.

631 BICKEL, T., LEHLE, L., SCHWARZ, M., AEBI, M. & JAKOB, C. A. 2005. Biosynthesis of lipid-linked
632 oligosaccharides in *Saccharomyces cerevisiae*: Alg13p and Alg14p form a complex required for
633 the formation of GlcNAc(2)-PP-dolichol. *J Biol Chem,* 280**,** 34500-6.
634 BORATYN, G. M., SCHAFFER, A. A., AGARWALA, R., ALTSCHUL, S. F., LIPMAN, D. J. & MADDEN, T. L.
635 2012. Domain enhanced lookup time accelerated BLAST. *Biol Direct,* 7**,** 12.
636 BOUHSS, A., TRUNKFIELD, A. E., BUGG, T. D. & MENGIN-LECREULX, D. 2008. The biosynthesis of
637 peptidoglycan lipid-linked intermediates. *FEMS Microbiol Rev,* 32**,** 208-33.
638 BROCK, T. D., BROCK, K. M., BELLY, R. T. & WEISS, R. L. 1972. *Sulfolobus* - new genus of sulfur-oxidizing
639 bacteria living at low pH and high temperature. *Arch Microbiol,* 84**,** 54-&.
640 BROWN, K., VIAL, S. C., DEDI, N., WESTCOTT, J., SCALLY, S., BUGG, T. D., CHARLTON, P. A. & CHEETHAM,
641 G. M. 2013. Crystal structure of the *Pseudomonas aeruginosa* MurG: UDP-GlcNAc substrate
642 complex. *Protein Pept Lett,* 20**,** 1002-8.
643 BUCHFINK, B., XIE, C. & HUSON, D. H. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat*
644 *Methods,* 12**,** 59-60.
645 BUGG, T. D. & BRANDISH, P. E. 1994. From peptidoglycan to glycoproteins: common features of lipid-
646 linked oligosaccharide biosynthesis. *FEMS Microbiol Lett,* 119**,** 255-62.
647 BURDA, P. & AEBI, M. 1999. The dolichol pathway of *N*-linked glycosylation. *Biochim Biophys Acta,*
648 1426**,** 239-57.
649 CARAMELO, J. J. & PARODI, A. J. 2007. How sugars convey information on protein conformation in the
650 endoplasmic reticulum. *Semin Cell Dev Biol,* 18**,** 732-42.
651 CARUGO, O. & ARGOS, P. 1997. NADP-dependent enzymes. I: Conserved stereochemistry of cofactor
652 binding. *Proteins,* 28**,** 10-28.
653 CHABAN, B., LOGAN, S. M., KELLY, J. F. & JARRELL, K. F. 2009. AglC and AglK are involved in biosynthesis
654 and attachment of diacetylated glucuronic acid to the *N*-glycan in *Methanococcus voltae*. *J*
655 *Bacteriol,* 191**,** 187-95.
656 CHABAN, B., VOISIN, S., KELLY, J., LOGAN, S. M. & JARRELL, K. F. 2006. Identification of genes involved
657 in the biosynthesis and attachment of *Methanococcus voltae N*-linked glycans: insight into *N*-
658 linked glycosylation pathways in Archaea. *Mol Microbiol,* 61**,** 259-68.
659 CHANTRET, I., DANCOURT, J., BARBAT, A. & MOORE, S. E. 2005. Two proteins homologous to the N-
660 and C-terminal domains of the bacterial glycosyltransferase MurG are required for the second
661 step of dolichyl-linked oligosaccharide synthesis in *Saccharomyces cerevisiae*. *J Biol Chem,* 280**,**
662 9236-42.
663 CROUVOISIER, M., AUGER, G., BLANOT, D. & MENGIN-LECREULX, D. 2007. Role of the amino acid
664 invariants in the active site of MurG as evaluated by site-directed mutagenesis. *Biochimie,* 89**,**
665 1498-508.

666 DE VUYST, L., DE VIN, F., VANINGELGEM, F. & DEGEEST, B. 2001. Recent developments in the
667 biosynthesis and applications of heteropolysaccharides from lactic acid bacteria. *Int Dairy J,*
668 11**,** 687-707.
669 EME, L., SPANG, A., LOMBARD, J., STAIRS, C. W. & ETTEMA, T. J. G. 2017. Archaea and the origin of
670 eukaryotes. *Nat Rev Microbiol,* 15**,** 711-723.
671 FUJINAMI, D., TAGUCHI, Y. & KOHDA, D. 2017. Asn-linked oligosaccharide chain of a crenarchaeon,
672 *Pyrobaculum calidifontis,* is reminiscent of the eukaryotic high-mannose-type glycan.
673 *Glycobiology*.
674 GAO, X. D., MORIYAMA, S., MIURA, N., DEAN, N. & NISHIMURA, S. 2008. Interaction between the C
675 termini of Alg13 and Alg14 mediates formation of the active UDP-N-acetylglucosamine
676 transferase complex. *J Biol Chem,* 283**,** 32534-41.
677 GAO, X. D., TACHIKAWA, H., SATO, T., JIGAMI, Y. & DEAN, N. 2005. Alg14 recruits Alg13 to the
678 cytoplasmic face of the endoplasmic reticulum to form a novel bipartite UDP-N-
679 acetylglucosamine transferase required for the second step of N-linked glycosylation. *J Biol*
680 *Chem,* 280**,** 36254-62.
681 GLOSTER, T. M. 2014. Advances in understanding glycosyltransferases from a structural perspective.
682 *Curr Opin Struct Biol,* 28**,** 131-41.
683 GUINDON, S., DUFAYARD, J. F., LEFORT, V., ANISIMOVA, M., HORDIJK, W. & GASCUEL, O. 2010. New
684 Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the
685 Performance of PhyML 3.0. *Syst Biol,* 59**,** 307-321.
686 HA, S., CHANG, E., LO, M. C., MEN, H., PARK, P., GE, M. & WALKER, S. 1999. The kinetic characterization
687 of *Escherichia coli* MurG using synthetic substrate analogues. *J Am Chem Soc,* 121**,** 8415-8426.
688 HA, S., WALKER, D., SHI, Y. G. & WALKER, S. 2000. The 1.9 angstrom crystal structure of *Escherichia coli*
689 MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis.
690 *Protein Sci,* 9**,** 1045-1052.
691 HELENIUS, A. & AEBI, M. 2004. Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev*
692 *Biochem,* 73**,** 1019-49.
693 HOANG, D. T., CHERNOMOR, O., VON HAESELER, A., MINH, B. Q. & VINH, L. S. 2018. UFBoot2:
694 Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol,* 35**,** 518-522.
695 HU, Y., CHEN, L., HA, S., GROSS, B., FALCONE, B., WALKER, D., MOKHTARZADEH, M. & WALKER, S. 2003.
696 Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a
697 superfamily of glycosyltransferases. *Proc Natl Acad Sci U S A,* 100**,** 845-9.
698 HUET, J., SCHNABEL, R., SENTENAC, A. & ZILLIG, W. 1983. Archaebacteria and eukaryotes possess DNA-
699 dependent RNA polymerases of a common type. *EMBO J,* 2**,** 1291-4.

700 IMACHI, H., NOBU, M. K., NAKAHARA, N., MORONO, Y., OGAWARA, M., TAKAKI, Y., TAKANO, Y.,
701 UEMATSU, K., IKUTA, T., ITO, M., MATSUI, Y., MIYAZAKI, M., MURATA, K., SAITO, Y., SAKAI, S.,
702 SONG, C., TASUMI, E., YAMANAKA, Y., YAMAGUCHI, T., KAMAGATA, Y., TAMAKI, H. & TAKAI,
703 K. 2020. Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature,* 577**,** 519-525.
704 JARRELL, K. F., DING, Y., MEYER, B. H., ALBERS, S. V., KAMINSKI, L. & EICHLER, J. 2014. N-linked
705 glycosylation in Archaea: a structural, functional, and genetic analysis. *Microbiol Mol Biol Rev,*
706 78**,** 304-341.
707 KALYAANAMOORTHY, S., MINH, B. Q., WONG, T. K. F., VON HAESELER, A. & JERMIIN, L. S. 2017.
708 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods,* 14**,** 587-
709 589.
710 KAMINSKI, L., LURIE-WEINBERGER, M. N., ALLERS, T., GOPHNA, U. & EICHLER, J. 2013. Phylogenetic-
711 and genome-derived insight into the evolution of N-glycosylation in Archaea. *Mol Phylogenet*
712 *Evol,* 68**,** 327-335.
713 KATOH, K. & STANDLEY, D. M. 2013. MAFFT Multiple Sequence Alignment Software Version 7:
714 Improvements in Performance and Usability. *Mol Biol Evol,* 30**,** 772-780.
715 KLEEREBEZEM, M., VAN KRANENBURG, R., TUINIER, R., BOELS, I. C., ZOON, P., LOOIJESTEIJN, E.,
716 HUGENHOLTZ, J. & DE VOS, W. M. 1999. Exopolysaccharides produced by *Lactococcus lactis*:
717 from genetic engineering to improved rheological properties? *Antonie Van Leeuwenhoek* 76**,**
718 357-365.
719 KOLKMAN, M. A. B., VANDERZEIJST, B. A. M. & NUIJTEN, P. J. M. 1997. Functional analysis of
720 glycosyltransferases encoded by the capsular polysaccharide biosynthesis locus of
721 *Streptococcus pneumoniae* serotype 14. *J Biol Chem,* 272**,** 19502-19508.
722 KOONIN, E. V. 2015. Archaeal ancestors of eukaryotes: not so elusive any more. *BMC Biol,* 13.
723 KUROSAWA, N. & GROGAN, D. W. 2005. Homologous recombination of exogenous DNA with the
724 *Sulfolobus acidocaldarius* genome: properties and uses. *FEMS Microbiol Lett,* 253**,** 141-9.
725 LAIRSON, L. L., HENRISSAT, B., DAVIES, G. J. & WITHERS, S. G. 2008. Glycosyltransferases: structures,
726 functions, and mechanisms. *Annu Rev Biochem,* 77**,** 521-55.
727 LARKIN, A. & IMPERIALI, B. 2011. The expanding horizons of asparagine-linked glycosylation.
728 *Biochemistry,* 50**,** 4411-26.
729 LEHLE, L., STRAHL, S. & TANNER, W. 2006. Protein glycosylation, conserved from yeast to man: a model
730 organism helps elucidate congenital human diseases. *Angew Chem Int Ed Engl,* 45**,** 6802-18.
731 LETUNIC, I. & BORK, P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments.
732 *Nucleic Acids Res,* 47**,** W256-W259.
733 LOMBARD, J. 2016. The multiple evolutionary origins of the eukaryotic N-glycosylation pathway. *Biol*
734 *Direct,* 11**,** 36.

735 LOMBARD, V., GOLACONDA RAMULU, H., DRULA, E., COUTINHO, P. M. & HENRISSAT, B. 2014. The
736 carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res,* 42**,** D490-5.
737 LU, J., TAKAHASHI, T., OHOKA, A., NAKAJIMA, K., HASHIMOTO, R., MIURA, N., TACHIKAWA, H. & GAO,
738 X. D. 2012. Alg14 organizes the formation of a multiglycosyltransferase complex involved in
739 initiation of lipid-linked oligosaccharide biosynthesis. *Glycobiology,* 22**,** 504-16.
740 LU, J., TAKAHASHI, T., OHOKA, A., NAKAJIMA, K. I., HASHIMOTO, R., MIURA, N., TACHIKAWA, H. & GAO,
741 X. D. 2011. Alg14 organizes the formation of a multi-glycosyltransferase complex involved in
742 initiation of lipid-linked oligosaccharide biosynthesis. *Glycobiology*.
743 MEN, H. B., PARK, P., GE, M. & WALKER, S. 1998. Substrate synthesis and activity assay for MurG. *J Am*
744 *Chem Soc,* 120**,** 2484-2485.
745 MESTROM, L., PRZYPIS, M., KOWALCZYKIEWICZ, D., POLLENDER, A., KUMPF, A., MARSDEN, S. R.,
746 BENTO, I., JARZEBSKI, A. B., SZYMANSKA, K., CHRUSCIEL, A., TISCHLER, D., SCHOEVAART, R.,
747 HANEFELD, U. & HAGEDOORN, P. L. 2019. Leloir Glycosyltransferases in Applied Biocatalysis:
748 A Multidisciplinary Approach. *Int J Mol Sci,* 20.
749 MEYER, B. H. & ALBERS, S.-V. 2020. Archaeal Cell Walls. *In eLS , John Wiley & Sons, Ltd (Ed.).* .
750 MEYER, B. H. & ALBERS, S. V. 2013. Hot and sweet: protein glycosylation in Crenarchaeota. *Biochem*
751 *Soc Trans,* 41**,** 384-392.
752 MEYER, B. H. & ALBERS, S. V. 2014. AglB, catalyzing the oligosaccharyl transferase step of the archaeal
753 N-glycosylation process, is essential in the thermoacidophilic crenarchaeon *Sulfolobus*
754 *acidocaldarius*. *Microbiologyopen,* 3**,** 531-43.
755 MEYER, B. H., PEYFOON, E., DIETRICH, C., HITCHEN, P., PANICO, M., MORRIS, H. R., DELL, A. & ALBERS,
756 S. V. 2013. Agl16, a thermophilic glycosyltransferase mediating the last step of N-glycan
757 biosynthesis in the thermoacidophilic crenarchaeon *Sulfolobus acidocaldarius*. *J Bacteriol,* 195**,**
758 2177-2186.
759 MEYER, B. H., SHAMS-ELDIN, H. & ALBERS, S. V. 2017. AglH, a thermophilic UDP-N-acetylglucosamine-
760 1-phosphate:dolichyl phosphate GlcNAc-1-phosphotransferase initiating protein N-
761 glycosylation pathway in *Sulfolobus acidocaldarius*, is capable of complementing the eukaryal
762 Alg7. *Extremophiles,* 21**,** 121-134.
763 MINH, B. Q., SCHMIDT, H. A., CHERNOMOR, O., SCHREMPF, D., WOODHAMS, M. D., VON HAESELER,
764 A. & LANFEAR, R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
765 Inference in the Genomic Era. *Mol Biol Evol,* 37**,** 1530-1534.
766 NIKOLAYEV, S., COHEN-ROSENZWEIG, C. & EICHLER, J. 2020. Evolutionary considerations of the
767 oligosaccharyltransferase AglB and other aspects of N-glycosylation across Archaea. *Mol*
768 *Phylogenet Evol,* 153**,** 106951.

769 NOTHAFT, H. & SZYMANSKI, C. M. 2010. Protein glycosylation in bacteria: sweeter than ever. *Nat Rev*
770 *Microbiol,* 8**,** 765-78.
771 NOTHAFT, H. & SZYMANSKI, C. M. 2013. Bacterial protein N-glycosylation: new perspectives and
772 applications. *J Biol Chem,* 288**,** 6912-20.
773 PALMIERI, G., BALESTRIERI, M., PETER-KATALINIC, J., POHLENTZ, G., ROSSI, M., FIUME, I. & POCSFALVI,
774 G. 2013. Surface-exposed glycoproteins of hyperthermophilic *Sulfolobus solfataricus* P2 show
775 a common glycosylation profile. *J Proteome Res*.
776 PETITJEAN, C., DESCHAMPS, P., LOPEZ-GARCIA, P. & MOREIRA, D. 2015. Rooting the Domain Archaea
777 by Phylogenomic Analysis Supports the Foundation of the New Kingdom Proteoarchaeota.
778 *Genome Biol Evol,* 7**,** 191-204.
779 PEYFOON, E., MEYER, B., HITCHEN, P. G., PANICO, M., MORRIS, H. R., HASLAM, S. M., ALBERS, S. V. &
780 DELL, A. 2010. The S-layer glycoprotein of the crenarchaeote *Sulfolobus acidocaldarius* is
781 glycosylated at multiple sites with chitobiose-linked *N*-glycans. *Archaea,* 2010.
782 RAMAN, S., LANGE, O. F., ROSSI, P., TYKA, M., WANG, X., ARAMINI, J., LIU, G., RAMELOT, T. A., ELETSKY,
783 A., SZYPERSKI, T., KENNEDY, M. A., PRESTEGARD, J., MONTELIONE, G. T. & BAKER, D. 2010.
784 NMR structure determination for larger proteins using backbone-only data. *Science,* 327**,**
785 1014-8.
786 REILY, C., STEWART, T. J., RENFROW, M. B. & NOVAK, J. 2019. Glycosylation in health and disease. *Nat*
787 *Rev Nephrol,* 15**,** 346-366.
788 SAMUELSON, J., BANERJEE, S., MAGNELLI, P., CUI, J., KELLEHER, D. J., GILMORE, R. & ROBBINS, P. W.
789 2005. The diversity of dolichol-linked precursors to Asn-linked glycans likely results from
790 secondary loss of sets of glycosyltransferases. *Proc Natl Acad Sci U S A,* 102**,** 1548-53.
791 SCHWARZ, F. & AEBI, M. 2011. Mechanisms and principles of N-linked protein glycosylation. *Curr Opin*
792 *Struct Biol,* 21**,** 576-82.
793 SHRIMAL, S. & GILMORE, R. 2019. Oligosaccharyltransferase structures provide novel insight into the
794 mechanism of asparagine-linked glycosylation in prokaryotic and eukaryotic cells.
795 *Glycobiology,* 29**,** 288-297.
796 SPANG, A., SAW, J. H., JORGENSEN, S. L., ZAREMBA-NIEDZWIEDZKA, K., MARTIJN, J., LIND, A. E., VAN
797 EIJK, R., SCHLEPER, C., GUY, L. & ETTEMA, T. J. G. 2015. Complex archaea that bridge the gap
798 between prokaryotes and eukaryotes. *Nature,* 521**,** 173-9.
799 STEENWYK, J. L., BUIDA, T. J., LI, Y., SHEN, X.-X. & ROKAS, A. 2020. ClipKIT: a multiple sequence
800 alignment-trimming algorithm for accurate phylogenomic inference. 2020.06.08.140384.
801 STUDIER, F. W. 2014. Stable expression clones and auto-induction for protein production in *E. coli*.
802 *Methods Mol Biol,* 1091**,** 17-32.

803 TAGUCHI, Y., FUJINAMI, D. & KOHDA, D. 2016. Comparative analysis of archaeal lipid-linked
804 oligosaccharides that serve as oligosaccharide donors for Asn glycosylation. *J Biol Chem,* 291**,**
805 11042-54.
806 VAN KRANENBURG, R., VOS, H. R., VAN SWAM, I. I., KLEEREBEZEM, M. & DE VOS, W. M. 1999.
807 Functional analysis of glycosyltransferase genes from *Lactococcus lactis* and other gram-
808 positive cocci: Complementation, expression, and diversity. *J Bacteriol,* 181**,** 6347-6353.
809 VAN WOLFEREN, M., SHAJAHAN, A., HEINRICH, K., BRENZINGER, S., BLACK, I. M., WAGNER, A., BRIEGEL,
810 A., AZADI, P. & ALBERS, S.-V. 2020. Species-Specific Recognition of *Sulfolobales* Mediated by
811 UV-Inducible Pili and S-Layer Glycosylation Patterns. *mBio,* 11**,** e03014-19.
812 VARKI, A. 1993. Biological roles of oligosaccharides - all of the theories are correct. *Glycobiology,* 3**,** 97-
813 130.
814 VARKI, A., CUMMINGS, R., ESKO, J., FREEZE, H., HART, G., MARTH, J. & EDITORS 1999. Chapter 17,
815 Glycosyltransferases. *Essentials of Glycobiolog.* Cold Spring Harbor (NY): Cold Spring Harbor
816 Laboratory Press.
817 WAGNER, M., VAN WOLFEREN, M., WAGNER, A., LASSAK, K., MEYER, B. H., REIMANN, J. & ALBERS, S.
818 V. 2012. Versatile genetic tool Box for the crenarchaeote *Sulfolobus acidocaldarius*. *Front*
819 *Microbiol,* 3**,** 214.
820 WALDO, G. S., STANDISH, B. M., BERENDZEN, J. & TERWILLIGER, T. C. 1999. Rapid protein-folding assay
821 using green fluorescent protein. *Nat Biotechnol,* 17**,** 691-5.
822 WANG, X., WELDEGHIORGHIS, T., ZHANG, G., IMPERIALI, B. & PRESTEGARD, J. H. 2008. Solution
823 structure of Alg13: the sugar donor subunit of a yeast N-acetylglucosamine transferase.
824 *Structure,* 16**,** 965-75.
825 WATERHOUSE, A., BERTONI, M., BIENERT, S., STUDER, G., TAURIELLO, G., GUMIENNY, R., HEER, F. T.,
826 DE BEER, T. A. P., REMPFER, C., BORDOLI, L., LEPORE, R. & SCHWEDE, T. 2018. SWISS-MODEL:
827 homology modelling of protein structures and complexes. *Nucleic Acids Res,* 46**,** W296-W303.
828 WHITFIELD, C., SZYMANSKI, C. M. & AEBI, M. 2017. Eubacteria. *In:* VARKI, A., CUMMINGS, R. D., ESKO,
829 J. D., STANLEY, P., HART, G. W., AEBI, M., DARVILL, A. G., KINOSHITA, T., PACKER, N. H.,
830 PRESTEGARD, J. H., SCHNAAR, R. L., SEEBERGER, P. H. & EDITORS (eds.) *Essentials of*
831 *Glycobiology.* Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press
832 WILD, R., KOWAL, J., EYRING, J., NGWA, E. M., AEBI, M. & LOCHER, K. P. 2018. Structure of the yeast
833 oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science,* 359**,**
834 545-550.
835 WILSON, C. M. & HIGH, S. 2007. Ribophorin I acts as a substrate-specific facilitator of N-glycosylation.
836 *J Cell Sci,* 120**,** 648-57.

837 WOESE, C. R. & FOX, G. E. 1977. Phylogenetic structure of the prokaryotic domain: the primary
838 kingdoms. *Proc Natl Acad Sci U S A,* 74**,** 5088-5090.
839 ZAHRINGER, U., MOLL, H., HETTMANN, T., KNIREL, V. A. & SCHAFER, G. 2000. Cytochrome $b_{558/566}$ from
840 the archaeon *Sulfolobus acidocaldarius* has a unique Asn-linked highly branched
841 hexasaccharide chain containing 6-sulfoquinovose. *Eur J Biochem,* 267**,** 4144-4149.
842 ZAREMBA-NIEDZWIEDZKA, K., CACERES, E. F., SAW, J. H., BACKSTROM, D., JUZOKAITE, L., VANCAESTER,
843 E., SEITZ, K. W., ANANTHARAMAN, K., STARNAWSKI, P., KJELDSEN, K. U., STOTT, M. B.,
844 NUNOURA, T., BANFIELD, J. F., SCHRAMM, A., BAKER, B. J., SPANG, A. & ETTEMA, T. J. 2017.
845 Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature,* 541**,** 353-358.
846 ZHANG, C., PHILLIPS, A. P. R., WIPFLER, R. L., OLSEN, G. J. & WHITAKER, R. J. 2018. The essential genome
847 of the crenarchaeal model *Sulfolobus islandicus*. *Nat Commun,* 9**,** 4908.
848 ZIELINSKA, D. F., GNAD, F., WISNIEWSKI, J. R. & MANN, M. 2010. Precision mapping of an *in vivo N*-
849 glycoproteome reveals rigid topological and sequence constraints. *Cell,* 141**,** 897-907.
850 ZILLIG, W., KLENK, H. P., PALM, P., PUHLER, G., GROPP, F., GARRETT, R. A. & LEFFERS, H. 1989. The
851 phylogenetic relations of DNA-dependent RNA polymerases of archaebacteria, eukaryotes,
852 and eubacteria. *Can J Microbiol,* 35**,** 73-80.
853 ZORZOLI, A., MEYER, B. H., ADAIR, E., TORGOV, V. I., VESELOVSKY, V. V., DANILOV, L. L., UHRIN, D. &
854 DORFMUELLER, H. C. 2019. Group A, B, C, and G Streptococcus Lancefield antigen biosynthesis
855 is initiated by a conserved alpha-d-GlcNAc-beta-1,4-l-rhamnosyltransferase. *J Biol Chem,* 294**,**
856 15237-15256.

Figures

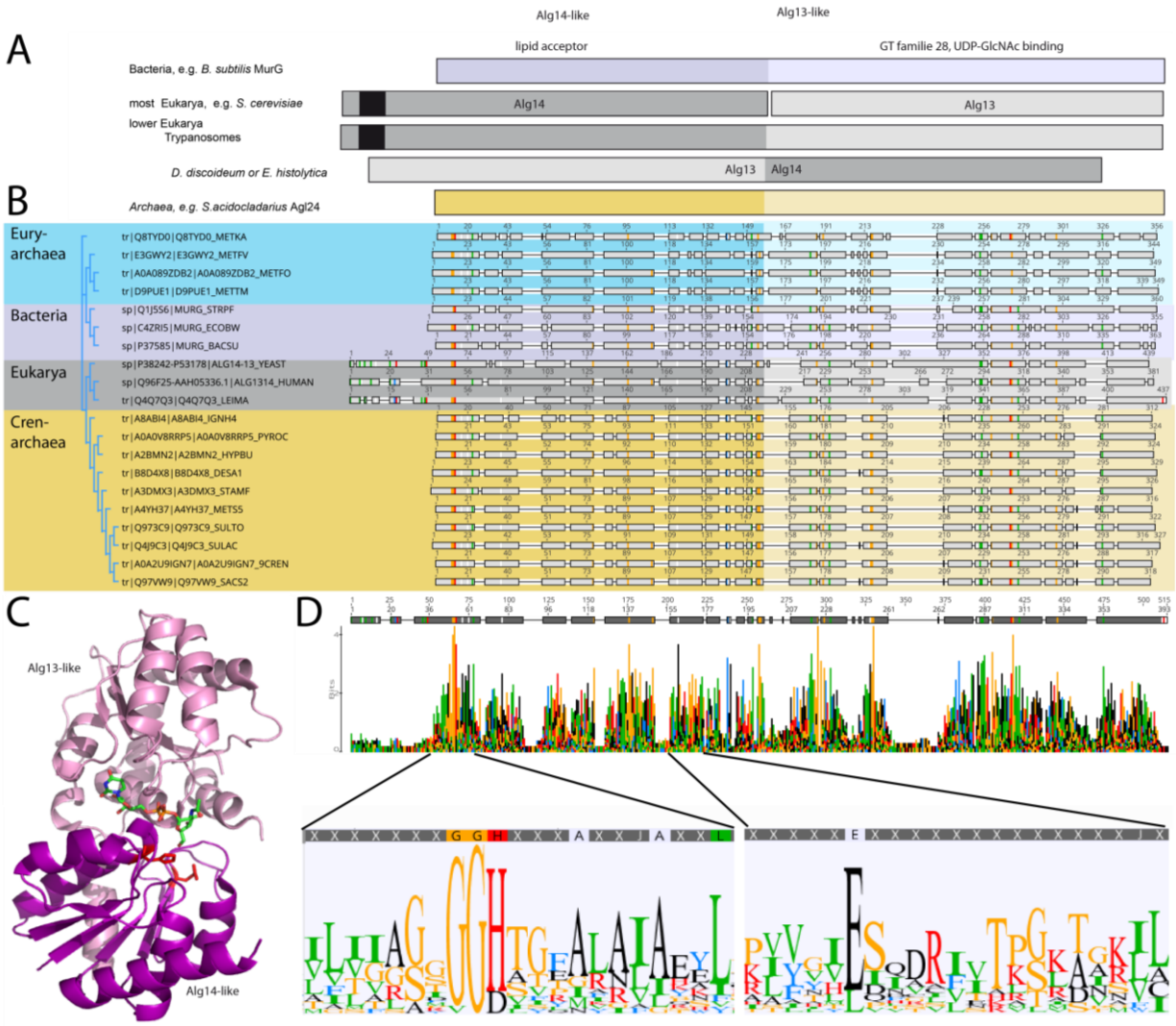


Fig. 1 Comparison of the archaeal Agl24, bacterial MurG, and eukaryal Alg14-13

A) Simplified comparison of the Agl14-like (darker) or Alg13-like (light color) domain in eukaryal Alg14-13, archaeal Agl24, and bacterial MurG ortologs. Transmembrane domain is depicted in black.

B) Protein sequences were aligned with the Clustal Omega (Sievers et al., 2018) (for the full alignment see Figure S2). Sequences derived from i) three bacterial MurG: *E.coli* (C4ZRI5), *Streptococcus pyogenes* (Q1J5S6), and *Bacillus subtilis* (P37585); ii) eukaryal Alg14-13: *Leishmania major* (Q4Q7Q3), an artificial Alg14-13 fusion from *Saccharomyces cerevisiae* (P38242-P53178) and *Homo sapiens* (Q96F25-Q9NP73-2); iii) the crenarchaeal Agl24 homologs: *Ignicoccus hospitalis (*A8ABI4), *Pyrodictium occultum* (A0A0V8RRP5), *Hyperthermus butylicus* (A2BMN2), *Desulfurococcus amylolyticus* (B8D4X8), *Staphylothermus marinus* (A3DMX3), *Metallosphaera sedula* (A4YH37), *Sulfurisphaera tokodaii* (Q973C9), *Sulfolobus acidocaldarius* (Q4J9C3), *Acidianus brierleyi* (A0A2U9IGN7), and *Saccharolobus solfataricus (*Q97VW9). Selected sequences from pseudomurein producing Euryarchaea with higher sequence similarity to the MurG are aligned: *Methanopyrus kandleri* (Q8TYD0), *Methanothermus fervidus (*E3GWY2), *Methanobacterium formicicum* (A0A089ZDB2), *Methanothermobacter marburgensis* (D9PUE1). Conserved amino acids (65% threshold) are highlighted in color corresponding to their chemical properties of the amino acids: green polar amino acids (G, S, T, Y, C, Q, N), blue basic (K, R, H), red acidic (D, E), and black hydrophobic amino acids (A, V, L, I, P, W, F, M). End of the Agl14-like domain and start of Alg13-like domains are indicated by the change from dark to light background color.

C) Ribbon presentation of MurG (PDB: 3s2u) bound to UDP-GlcNAc (sticks, green carbon, blue nitrogen, red oxygen) and orientation towards the cell membrane (bottom). Agl13-like domain (turquoise) and Agl14-like domain (violet) are labelled. Conserved His and Glu residues in Agl14 are shown as red sticks.

D) Strict consensus (65%) and Weblogo from the sequence alignment shown in B). Height of each bar correspond to the observed frequency, highly conserved motifs or amino acids are enlarged below, including conserved motifs G(x)GGH$_{15}$ (Loop I) and E$_{114}$ (full Weblogo see Figure S2).

A B

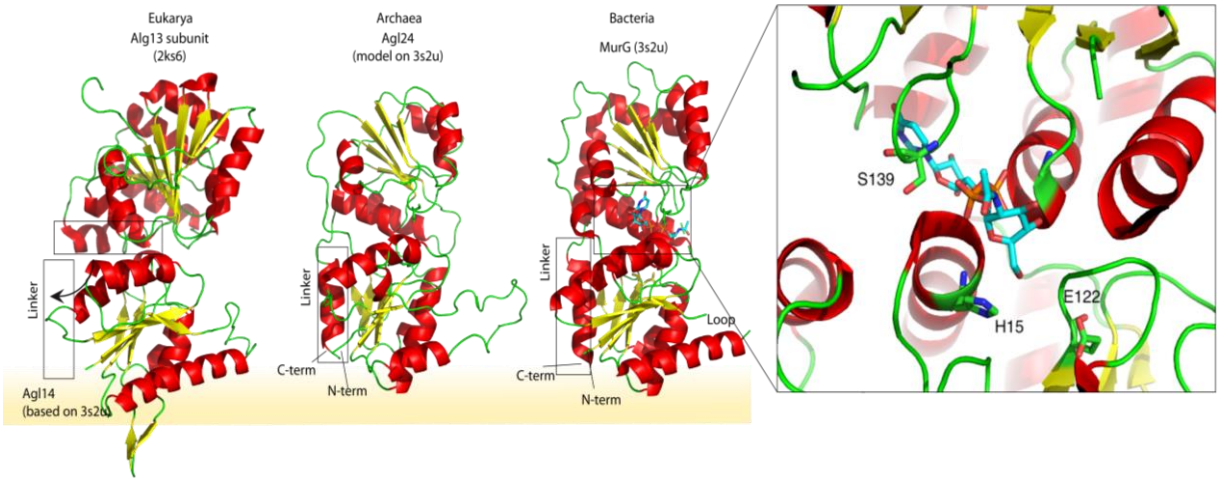


Figure 2: A) Structural comparison of the eukaryal Agl13 (PDB:2ks6) with a structural model of Agl14, the bacterial MurG (PDB: 3s2u), and the structural model of the crenarchaeal Agl24. Structural models were built with SWISS-MODEL (Waterhouse et al., 2018). Detailed results are listed in Supplementary Table S2. The interactions of the N-terminal helices with the membrane are depicted with a yellow background.

B) Magnified and 90˚ rotated view into the active site of MurG (PDB: 3s2u) in complex with UDP-GlcNAc (sticks, turquoise carbon atoms). The catalytic site is located in the cleft of both domains, with conserved His_{15} and $Glu_{(122)}$ residues are highlighted in red.

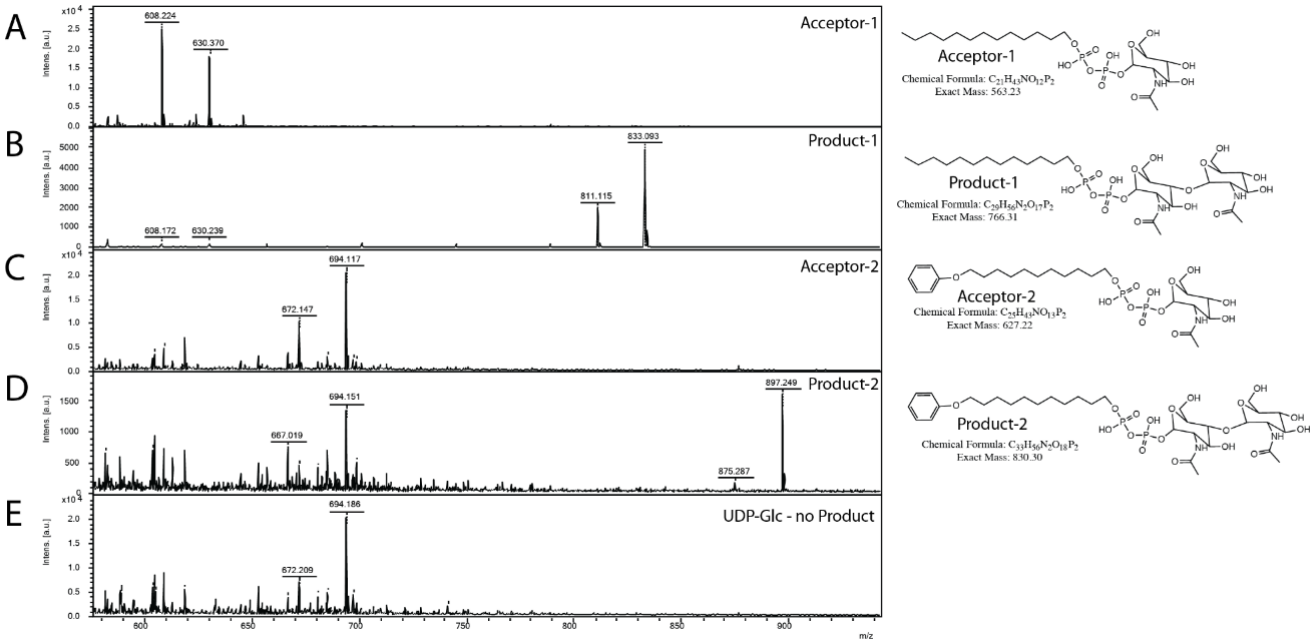


Figure 3: MALDI-MS spectra of the *in vitro* Agl24 reaction assessing the N-acetylglucosamine transferase activity. Spectra obtained from the enzymatic reactions containing Agl24-GFP and (**A**) acceptor-1, **(B)** acceptor-1 and UDP-GlcNAc, and Agl24-GFP, **(C)** acceptor-2, **(D)** acceptor-2, and UDP-GlcNAc, **(E)** acceptor-2 and UDP-Glc. The conversion from acceptor 1 (608 m/z $[M-1H+2Na]^+$ and 630 m/z $[M-2H+3Na]^+$) or acceptor 2 (672 m/z $[M-1H+2Na]^+$ and 694 m/z $[M-2H+3Na]^+$) to the product (811 m/z $[M-1H+2Na]^+$ and 833 m/z $[M-2H+3Na]^+$) or (875 m/z $[M-1H+2Na]^+$ and 897 m/z $[M-2H+3Na]^+$) was observed only when UDP-GlcNAc was used as nucleotide sugar donor.

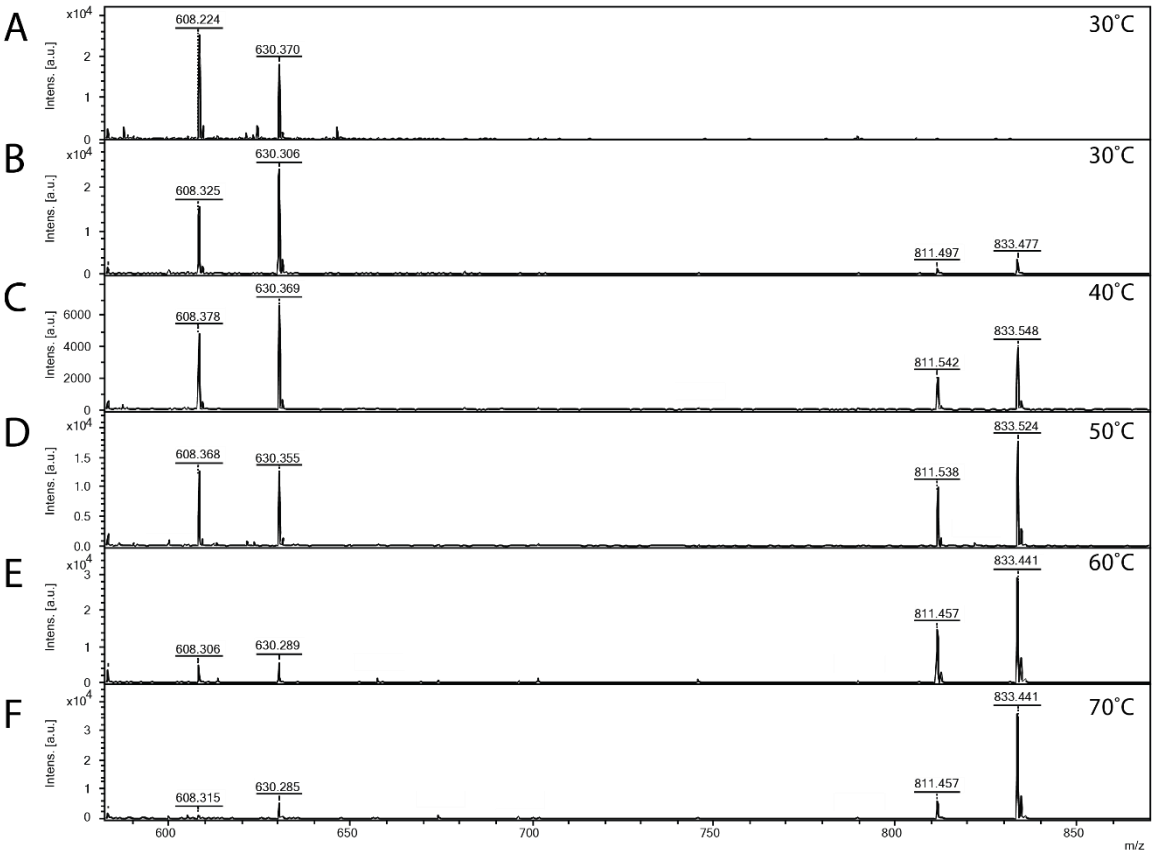


Figure4: MALDI-MS spectra of the *in vitro* Agl24 reaction at different temperatures. Spectra obtained from the purified enzymatic reaction mix with only the acceptor-1 (**A**), acceptor-1, UDP-GlcNAc, and Agl24-GFP WT at 30°C (**B**), 40°C (**C**), 50°C (D), 60°C (**E**), and 70°C (F). Conversion of the acceptor-1 (608 m/z [M-1H+2Na] and 630 m/z [M-2H+3Na]) towards the product (811 m/z [M-1H+2Na] and 833 m/z [M-2H+3Na]) is significantly dependent on the applied temperature. Above 50°C the depletion of the acceptor towards the gain of product is clearly visible.

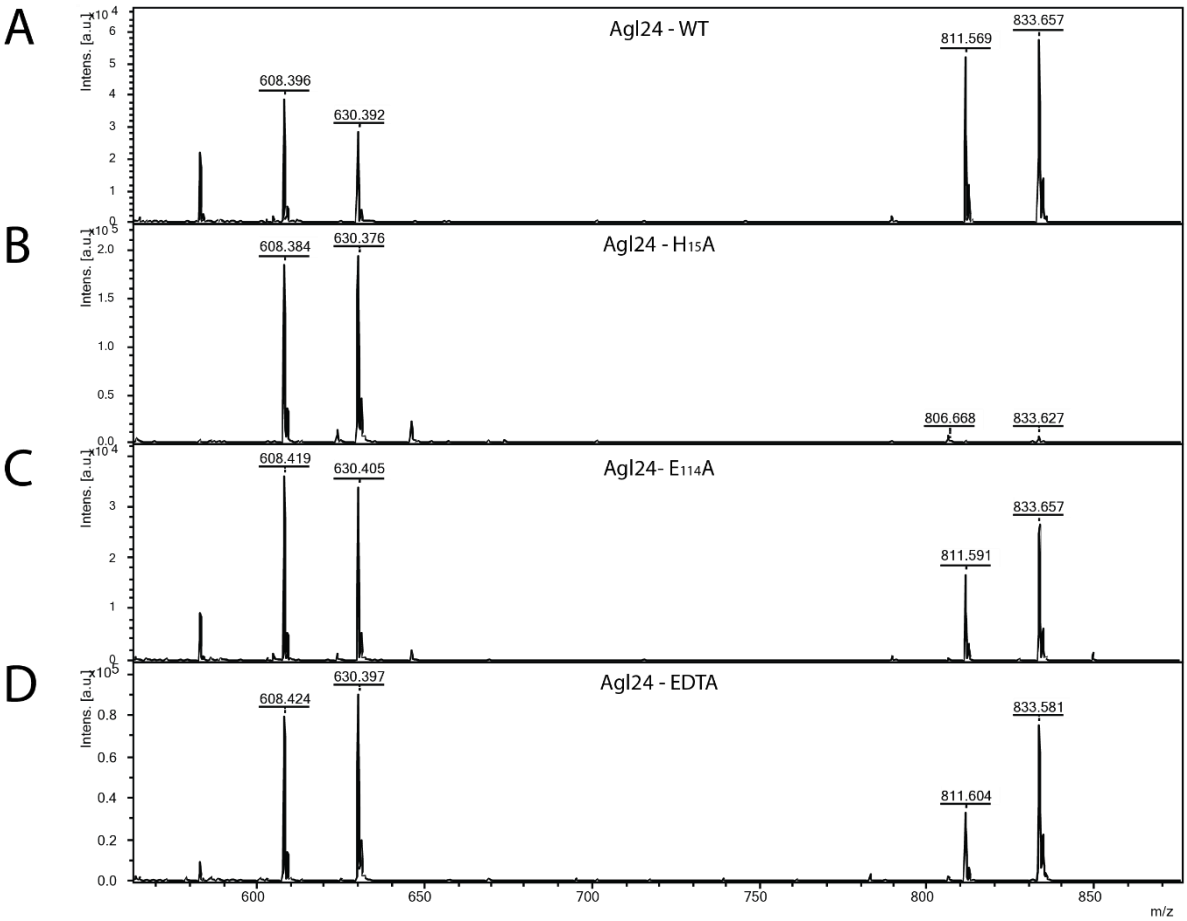


Figure 5: MALDI-MS spectra of the *in vitro* Agl24 reaction, the generated point mutations H15A and E114A and EDTA control. Spectra obtained from the purified enzymatic reaction mix with acceptor-1(608 m/z [M-1H+2Na] and 630 m/z [M-2H+3Na]), UDP-GlcNAc and (**A**) Agl24-GFP WT, (**B**) Agl24-H_{15}A-GFP, (**C**) Agl24-E_{114}A-GFP, and (**D**) Agl24-GFP WT with addition of 10 mM EDTA. The activity was significantly reduced in the H_{15}A mutants, while a similar amount of product (811 m/z [M-1H+2Na] and 833 m/z [M-2H+3Na]) was obtained in the E_{114}A mutant.

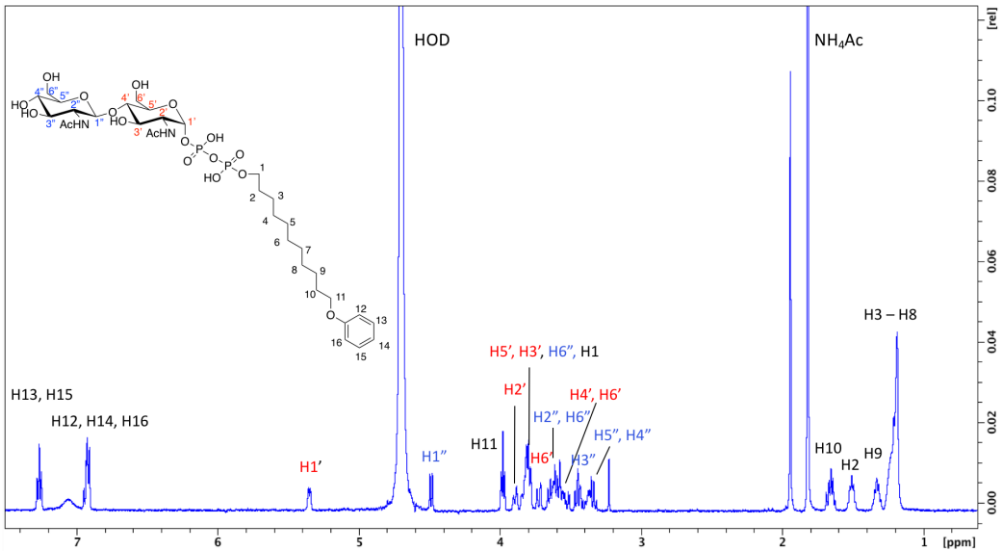


Figure 6: ^{1}H NMR spectra of purified Agl24 enzymatic product. Spectra were acquired in D_2O at 293K on a Bruker AVANCE III HD 500 MHz NMR Spectrometer fitted with a 5-mm QCPI cryoprobe. Chemical shifts are reported with respect to the residual HDO signal at δ_H 4.70 ppm.

Table 1: ^{1}H and ^{13}C chemical shifts of the *in vitro* Agl24 reaction product.

Lipid		**1**	**2**	**3 to 8**	**9**	**10**	**11**	**12, 14**	**13, 15**	**16**
	^{1}H NMR	3.79	1.51	1.16-1.26	1.33	1.66	3.98	6.92	7.26	6.92
	^{13}C NMR	66.8	30.3	28.8	28.7	28.7	68.8	115.6	130.1	121.8
α-GlcNAc		**1'**	**2'**	**3'**	**4'**	**5'**	**6'a**	**6'b**		
	^{1}H NMR	5.35	3.896	3.8	3.58	3.83	3.73	3.57		
	^{13}C NMR	94.25	53.65	70	79.65	71.69	60.15			
β-GlcNAc		**1"**	**2"**	**3"**	**4"**	**5"**	**6"a**	**6"b**		
	^{1}H NMR	4.49	3.64	3.45	3.34	3.38	3.8	3.6		
	^{13}C NMR	101.55	56.14	74.01	70.38	76.27	61.15			

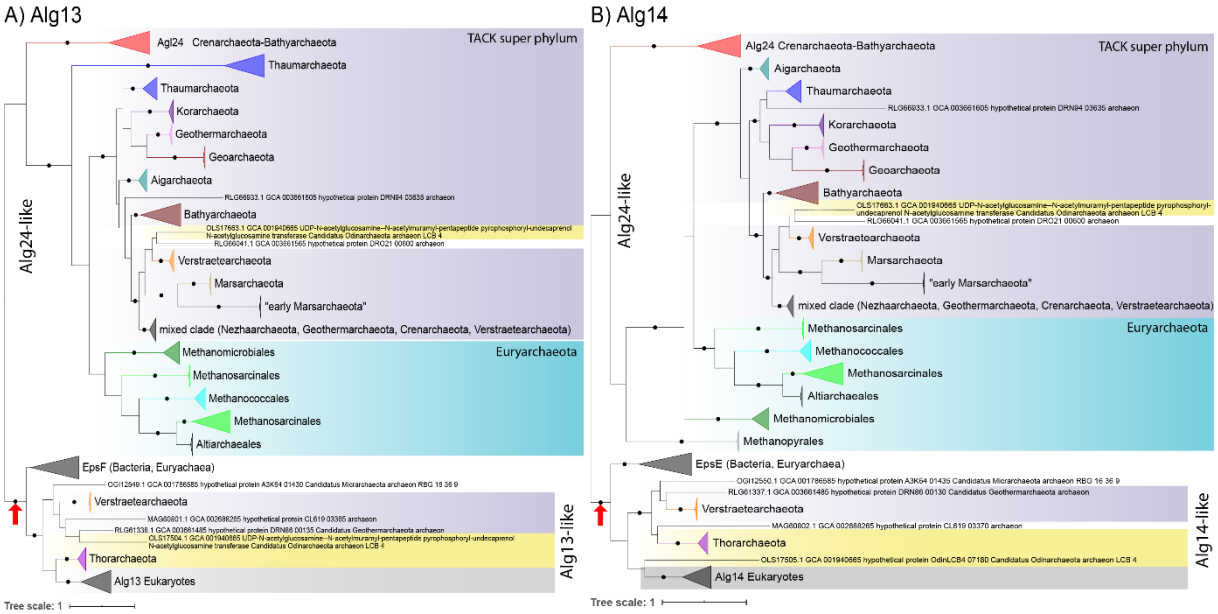


Figure 7: Single gene phylogenies of Alg13/EpsF and Alg14/EpsE. Dots indicate strongly supported branches (ultrafast bootstrap >=95 and aLRT SH-like support >=80). Higher-level classifications are highlighted in specific background colors: purple: archaeal TACK superphylum, turquoise: Euryarchaeota, yellow: Asagradarchaeota, grey: Eukaryotes. Agl24 of *S. acidocaldarius* is found within the red colored clade at the top of the phylogenetic tree. Red arrow: indicates major split event to form the Alg13/EpsF-like and Alg14/EpsE-like enzymes.

Supplementary Data

Table S1 Results of the Domain Enhanced Lookup Time Accelerated BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). DELTA BLAST was performed to identify homologs of MurG (WP_063074721.1), Alg13 (NP_011468.1), Alg14 (NP_009626.1), and Alg14-13 fusion in *Sulfolobus acidocaldarius*.

Description	Max score	Total score	Query cover	E value	Per. Ident	Accession	Saci gene	KO	Annotation
MurG WP_063074721.1 undecaprenyldiphospho-muramoylpentapeptide beta-N-acetylglucosaminyltransferase [*Escherichia coli*]									
conserved Archaeal protein [Sulfolobus acidocaldarius DSM 639]	77.0	77.0	76%	6e-17	14.52%	AAY81210.1	*saci_1907*	MW061 BM-A045-B027	RFaB GT1_YqgM_like
conserved protein [Sulfolobus acidocaldarius DSM 639]	71.2	71.2	98%	5e-15	16.76%	AAY80607.1	*saci_1262*	*essential*	MurG
hypothetical protein Saci_1921 [Sulfolobus acidocaldarius DSM 639]	71.2	71.2	72%	6e-15	13.33%	AAY81224.1	*saci_1921*	MW069 BM-A055-B003	RfaB
RecName: Full=Archaeal glycosylation protein 16 [Sulfolobus acidocaldarius DSM 639]	65.9	65.9	75%	3e-13	9.90%	Q4JAK2.1	*saci_0807*	MW043 BM-A110-B008	Archaeal glycosylation protein 16
partially conserved Archaeal protein [Sulfolobus acidocaldarius DSM 639]	61.2	61.2	96%	1e-11	12.76%	AAY81207.1	*saci_1904*	MW077 BM-A050-B002	Cell wall/membrane/envelope biogenesis
glycosyl transferase [Sulfolobus acidocaldarius DSM 639]	59.7	98.2	74%	5e-11	14.02%	AAY80548.1	*saci_1201*	MW053 BM-A150-B1	Glycogen synthase
glycosyl transferase group 1 [Sulfolobus acidocaldarius DSM 639]	47.8	47.8	77%	3e-07	12.38%	AAY81133.1	*saci_1827*	MW042 BM-A280-B001	GT1_Trehalose_phosphorylase
gylcosyl transferase group 1 protein [Sulfolobus acidocaldarius DSM 639]	45.8	45.8	49%	1e-06	16.59%	AAY80595.1	*saci_1249*	MW054 BM-A155-B001	GT1_YqgM_like
conserved protein [Sulfolobus acidocaldarius DSM 639]	42.4	42.4	34%	1e-05	18.11%	AAY80235.1	*saci_0869*	*essential*	Glycosyltransferase involved in cell wall biosynthesis
conserved protein [Sulfolobus acidocaldarius DSM 639]	42.4	42.4	58%	2e-05	9.91%	AAY81219.1	*saci_1916*	MW066 BM-A024-B001	
RecName: Full=UDP-sulfoquinovose synthase [Sulfolobus acidocaldarius DSM 639]	40.4	40.4	26%	7e-05	19.39%	Q4JBJ3.1	*saci_0423*	MW039 BM-A083-B008	Agl3 UDP-sulfoquinovose synthase
hypothetical protein Saci_1923 [Sulfolobus acidocaldarius DSM 639]	39.3	39.3	34%	1e-04	11.72%	AAY81226.1	*saci_1923*	n.d.	
conserved protein [Sulfolobus acidocaldarius DSM 639]	38.5	38.5	78%	3e-04	11.85%	AAY81225.1	*saci_1922*	MW070 BM-A060-B002	
ALG13 NP_011468.1 N-acetylglucosaminyldiphosphodolichol N-acetylglucosaminyltransferase [*Saccharomyces cerevisiae* S288C]									
conserved protein [Sulfolobus acidocaldarius DSM 639]	43.9	43.9	58%	1e-06	17.36%	AAY80607.1	*saci_1262*		MurG
ALG14 NP_009626.1 N-acetylglucosaminyldiphosphodolichol N-acetylglucosaminyltransferase [*Saccharomyces cerevisiae* S288C]									
	No homolog above the standard Delta-Blast e-value threshold of 0.05								
Alg14-13 fusion (NP_009626.1+ NP_011468.1)									
conserved protein [Sulfolobus acidocaldarius DSM 639]	59.7	59.7	73%	4e-11	16.92%	AAY80607.1	*saci_1262*		

Table S2: Structural alignment modeling of Agl24 revealed a conservation of fold to the available structures of MurG and Alg13. Structural modelling was performed by SWISS-MODEL (Waterhouse et al., 2018), using either the full length Agl24 sequence or only the c-terminal Alg13-like part. Identified PDB numbers as well as the SWISS_MODLE results, *e.g.* Global Model Quality Estimation (GMQE), quaternary structure quality estimate, and sequence identity are shown.

PDB	GMQE:	QSQE:	Identity	X-ray	Reference
Full Agl24 sequence					
3s2U MurG	0.51	-	17.2	2.2Å	(Brown et al., 2013)
1f0k MurG	0.55	0,31	17.36	1.9Å	(Ha et al., 2000);
1nlm MurG	0.55	0,20	17.36	2.5Å	(Hu et al., 2003)
c-terminal Agl13-like part of Agl24					
2ks6 Alg13	0.56		20.53	NMR	(Raman et al., 2010)
2jzc Alg13	0.56	-	19.87	NMR	(Wang et al., 2008)
1f0k MurG	0.46	0.17	19.70	1.9Å	(Ha et al., 2000);
1nlm MurG	0.46	-	19.70	2.5Å	(Hu et al., 2003)

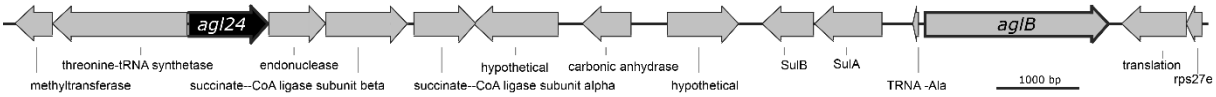


Figure S1: Physical map of the gene region adjacent to *agl24 and aglB* of *S. acidocaldarius*. Illustrated are the genes *Saci1260* until *Saci1276*. The gene *agl24* (*saci1262,* SACI_RS06030), displayed in black, which is annotated to encode a polysaccharide biosynthesis protein. The gene *aglB* (*saci1274,* SACI_RS06085), displayed with bold border, encodes the N-oligosaccharyltransferase, catalysing the transfer of the lipid-linked N-glycan onto the target protein.

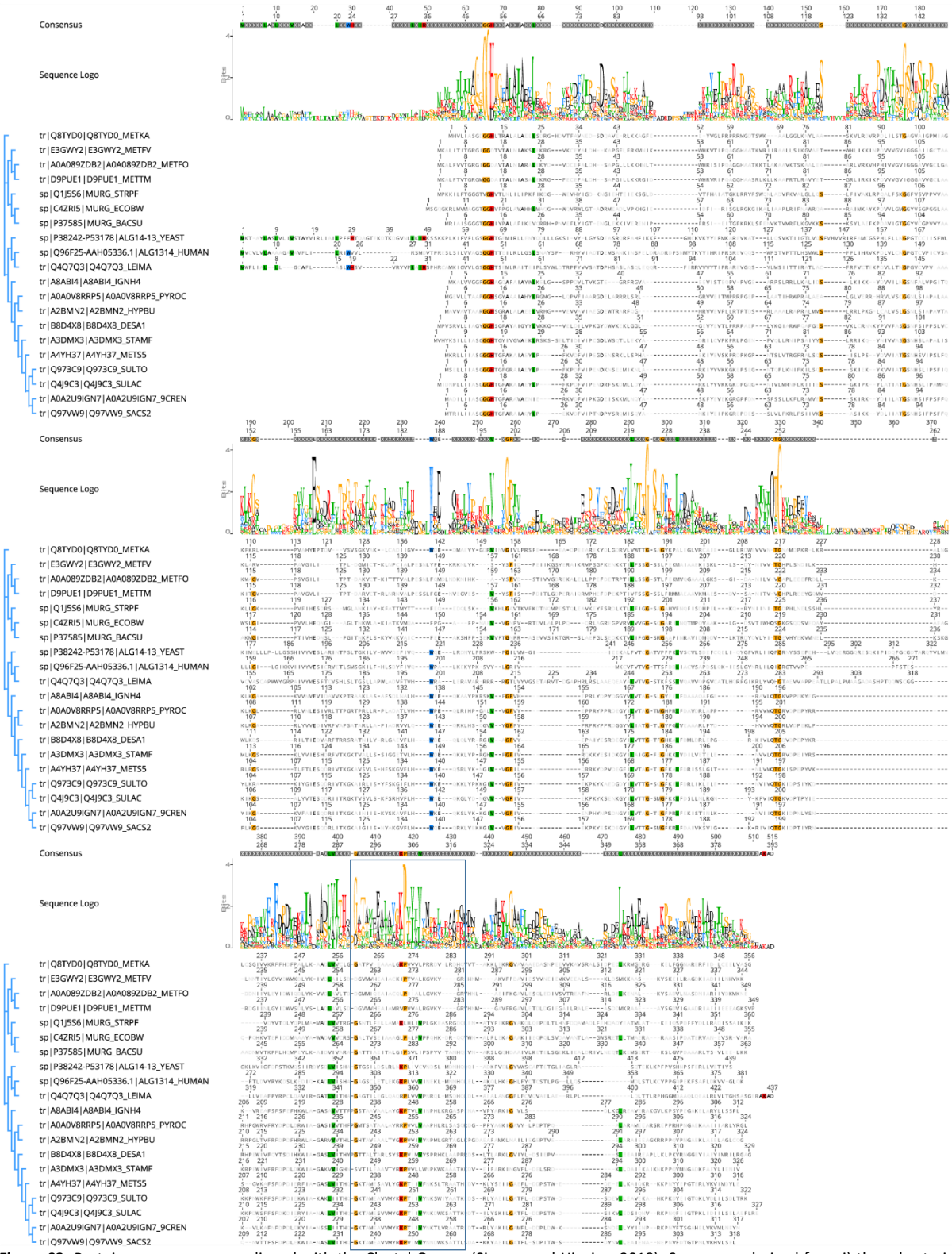


Figure S2: Protein sequences were aligned with the Clustal Omega (Sievers and Higgins, 2018). Sequences derived from i) three bacterial MurG: *E.coli* (C4ZRI5), *Streptococcus pyogenes* (Q1J5S6), and *Bacillus subtilis* (P37585); ii) eukaryal Alg14-13: *Leishmania major* (Q4Q7Q3), an artificial Alg14-13 fusion from *Saccharomyces cerevisiae* (P38242-P53178) and *Homo sapiens* (Q96F25-Q9NP73-2); iii) the crenarchaeal Agl24 homologs: *Ignicoccus hospitalis (*A8ABI4), *Pyrodictium occultum* (A0A0V8RRP5), *Hyperthermus butylicus* (A2BMN2), *Desulfurococcus amylolyticus* (B8D4X8), *Staphylothermus marinus* (A3DMX3), *Metallosphaera sedula* (A4YH37), *Sulfurisphaera tokodaii* (Q973C9), *Sulfolobus acidocaldarius* (Q4J9C3), *Acidianus brierleyi* (A0A2U9IGN7), and *Saccharolobus solfataricus (*Q97VW9). Selected sequences from pseudomurein producing Euryarchaea with higher sequence similarity to the MurG are aligned: *Methanopyrus kandleri* (Q8TYD0), *Methanothermus fervidus (*E3GWY2), *Methanobacterium formicicum* (A0A089ZDB2), *Methanothermobacter marburgensis* (D9PUE1). Conserved amino acids (65% threshold) are highlighted in color corresponding to their chemical properties of the amino acids: green polar amino acids (G, S, T, Y, C, Q, N), blue basic (K, R, H), red acidic (D, E), and black hydrophobic amino acids (A, V, L, I, P, W, F, M). End of the Agl14-like domain and start of Alg13-like domains are indicated by the change from dark to light background color. Conserved amino acids (65% threshold) are highlighted in color. Weblogo from the sequence aliment shown above. Height of each bar correspond to the observed frequency. **Boxed**: area of the sugar donor binding site, in *E. coli* MurG the amino acids A_{264}, L_{265}, T_{266}, E_{269}, Q_{288}, and Q_{289} have been to interact with UDP-GlcNAc.

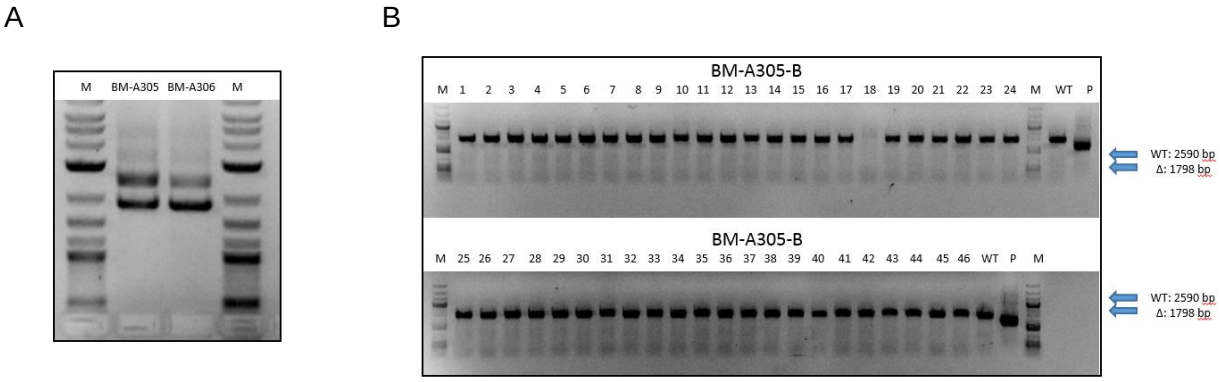
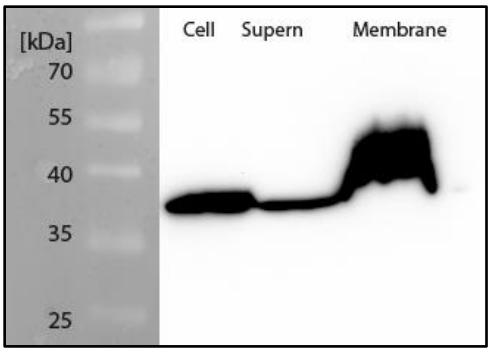
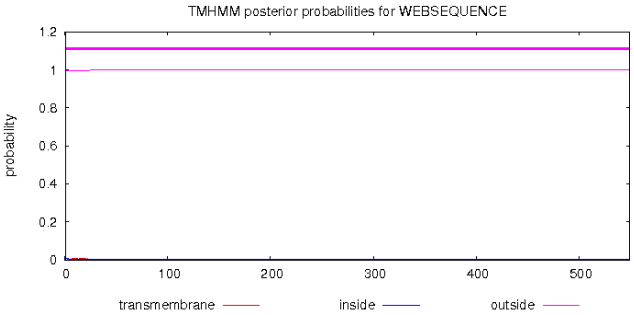


Figure S3: Confirmation of the integration and segregation of the *ag23* deletion plasmid pSVA1312 in *S. acidocaldarius* MW001 A) The integration of the *Agl24* deletion plasmid pSVA1312 in *S. acidocaldarius* MW001 (first selection) was monitored by PCR using the out primers of the upstream and downstream region of *Agl24* and the genomic DNA from two first selection colonies incorporate the plasmid (BM-A305 and BM-A-306). DNA from the background strain MW001 and the plasmid pSVA1312 were used as control, showing a PCR fragment corresponding to the flanking region including or excluding the *Agl24* gene, respectively. **B)** The segregation of pSVA1312 (second selection) was confirmed by PCR using the outer primers of the flanking region of *Agl24* and the genomic DNA from second selection colonies. All PCR fragments gained from genomic DNA of second selection colonies correspond to the full length *Agl24* gene (2590 bp), while a deletion would result in a 1798 bp PCR fragment.

A



B



C

```
MIDNPLLIIASGGGHTGFARAIAEYLPFKPDFVIPENDRFSKDMLLDYARKLYYVKKGKDPGQGNIVLMRNFLKIIIESGKIPKYLATIATGSN
----EEEEEEE-----HHHHHHHHHH----EEEE----HHHHHH------EEEE--------HHHHHHHHHHHHHHHHHHH---EEEE----
HSLIPAMFQKIKGSALYVTESQDRIITRGKTVSVLSKFSRHVFLHWNEQKGLYDNGVVVGPIVEKPKYKSENKGYILVTTGSMGFKKLFDSLLN
--HHHHHHHH----EEEEE---EEEE---HHHHH-----EEE------HHH----------------------EEEEEE----HHHHHHHHH
LRGNYKFVIQTGKVDPTPYIEKKPDWSFFSFDKDIERYIANAELVITHQGKTAMESVVMYGKPTIIVYNKDWKSATTKQDTILYSKILGATFLD
-----EEEEEE-----HHHHH----EEEEE----HHHHHHH-HEEE---HHHHHHHHHH---EEEEE--------HHHHHHHHHHH--EEEE
DPSTWDSIKVLEDSIQNVRKPNQFEIGTPKLIDIILSELAEFLRE
------HHHHHHHHHHHH---HHHHHHHHHHHHHHHHHHHH--
```

Figure S4: Localization of Alg24 in *S. acidocaldarius*. A. Anti-His immunoblot of samples from *S. acidocaldarius* BM-A724 expressing Agl24 with C-terminal His- and Strep-Tag reveal that the majority of Agl24 is found in the membrane fraction. Cell: total cell pellet, **B.** Prediction of transmembrane helices in Agl24, by the TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) confirmed the lack of any transmembrane domain. **C** Secondary protein structure prediction by the Jpred Server (Drozdetskiy et al., 2015).

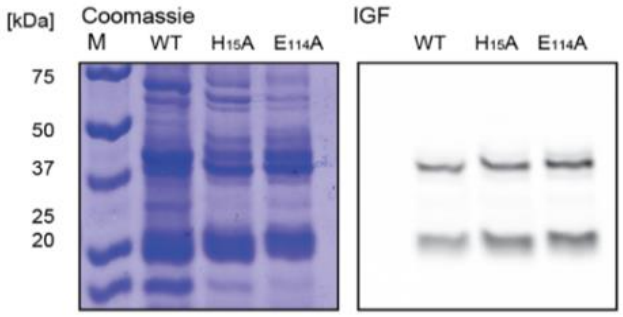
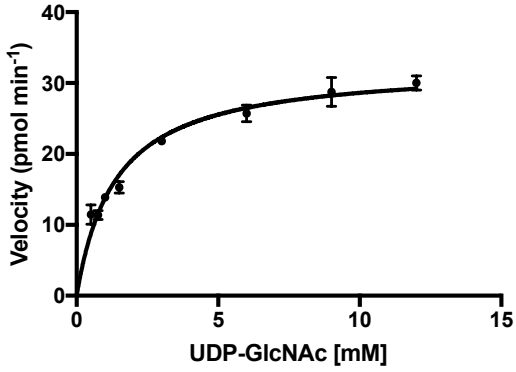


Figure S5: SDS-PAGE of the purified Agl24-WT-GFP, mutants Agl24-H_{15}A-GFP, and Agl24-E_{114}A-GFP from *E.coli*, stained with Coomassie Brilliant Blue or visualized by in gel fluorescents (IGF).



Michaelis-Menten	
Best-fit values	
Vmax	32.53
Km	1.366
Std. Error	
Vmax	0.8961
Km	0.1282
95% CI (profile likelihood)	
Vmax	30.68 to 34.53
Km	1.112 to 1.675
Goodness of Fit	
Degrees of Freedom	21
R square	0.9548
Absolute Sum of Squares	56.05
Sy.x	1.634
Constraints	
Km	Km > 0
Number of points	
# of X values	24
# Y values analyzed	23

Figure S6: Agl24 K_m^{app} for UDP-GlcNAc determination

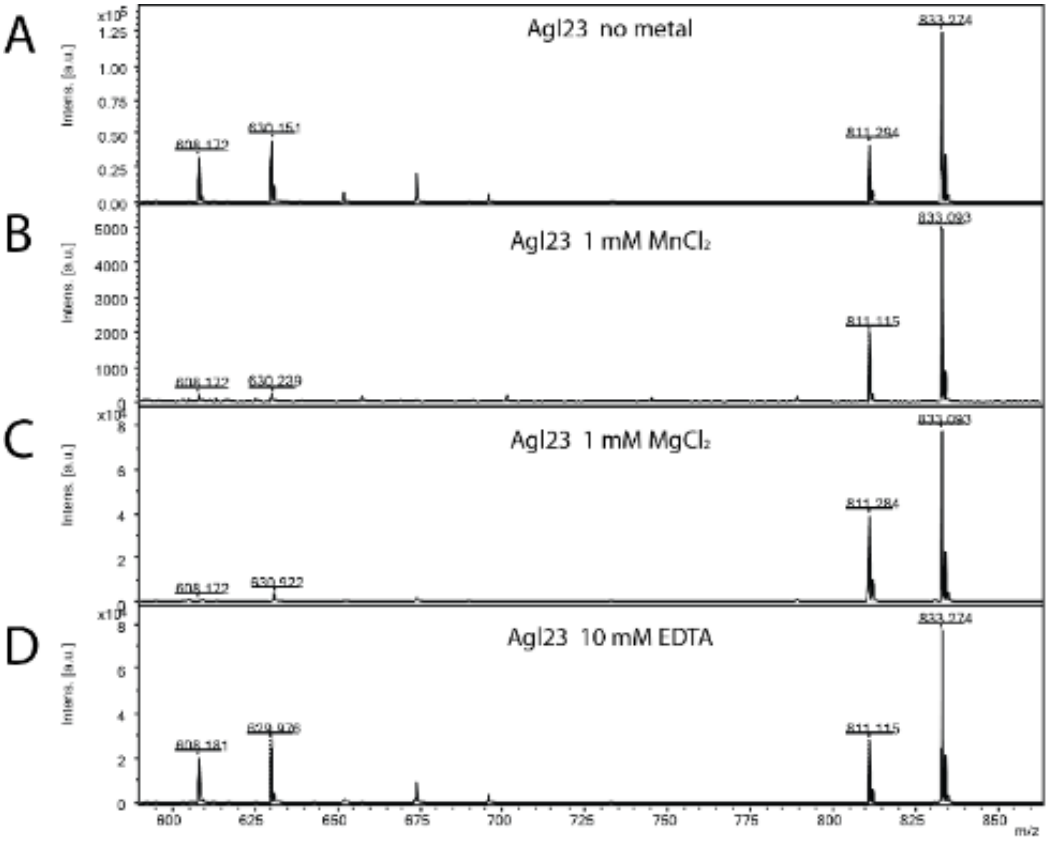


Figure S7: MALDI-MS spectra of the *in vitro* Agl24 reaction to assess the metal dependency for the specificity activity. Spectra obtained from the purified enzymatic reaction mix with acceptor-1, UDP-GlcNAc and Agl24-GFP without addition of metal cations (**A**), with the addition of 1 mM $MnCl_2$ (**B**), 1 mM $MgCl_2$ (**C**), or 10 mM EDTA (**D**). In all cases the acceptor-1 (608 m/z [M-1H+2Na] and 630 m/z [M-2H+3Na]) was converted to the product (811 m/z [M-1H+2Na] and 833 m/z [M-2H+3Na]), indicating the metal cations are not required for enzymatic activity.

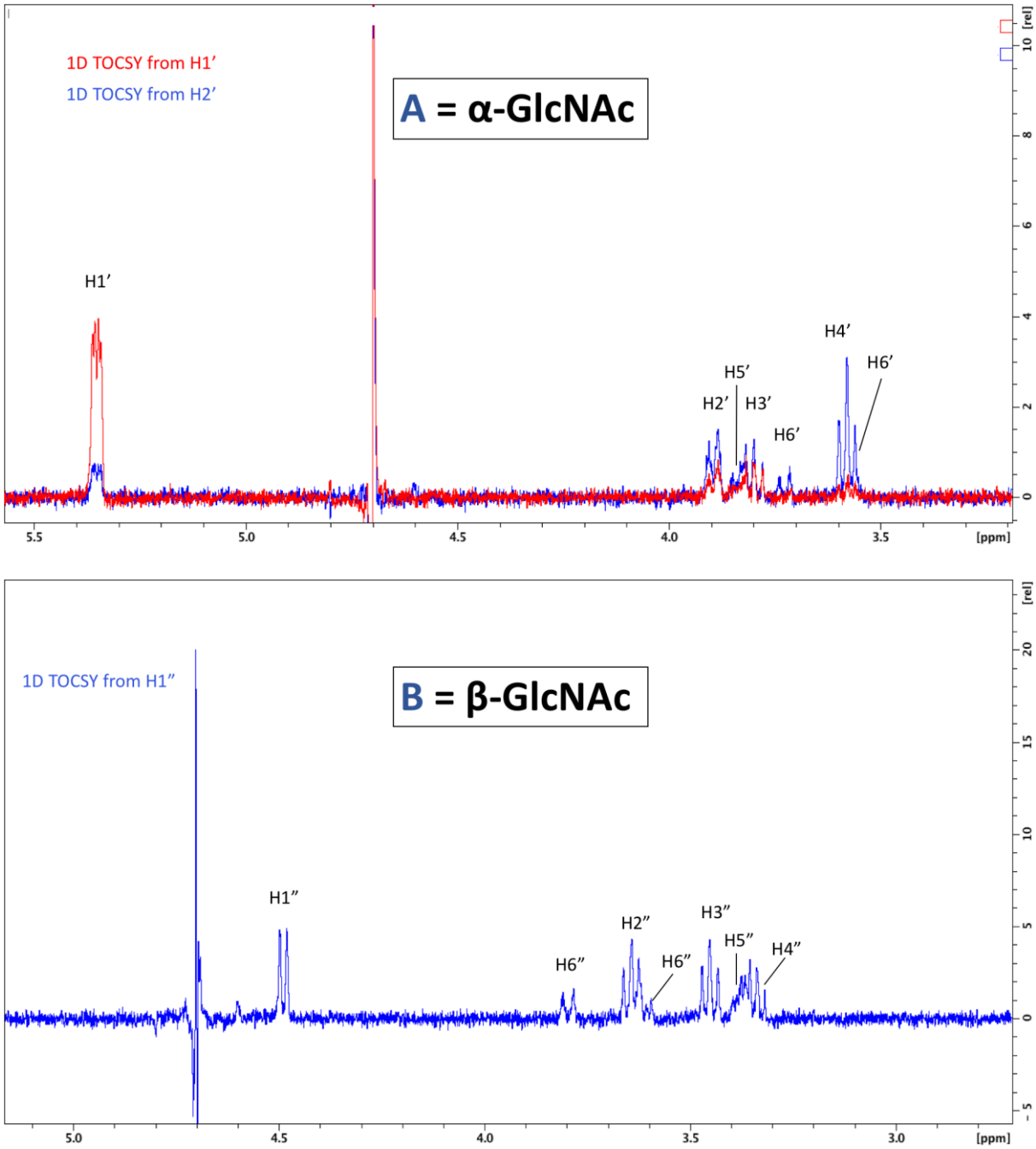


Figure S8: 1D TOCSY correlation spectra. To visualize all protons of the α-linked GlcNAc residue, irradiation was set for the shift of H1' and H2' (Top). To visualize protons of the β-linked GlcNAc, irradiation of H1" was sufficient to see all protons (Bottom).

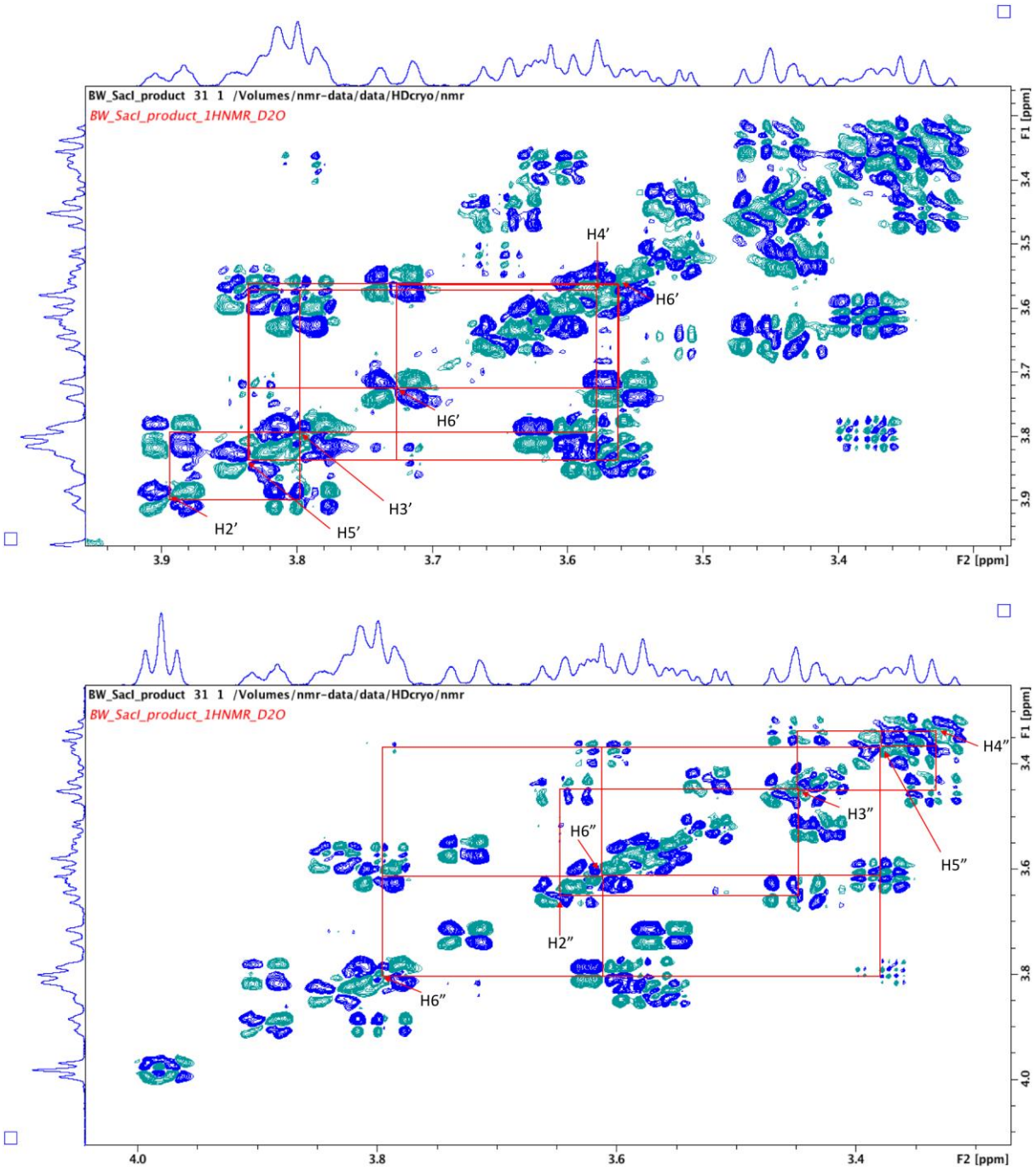


Figure S9: 2D COSY spectra used to assign proton signals arising from α-GlcNAc (Top) and β-GlcNAc (Bottom).

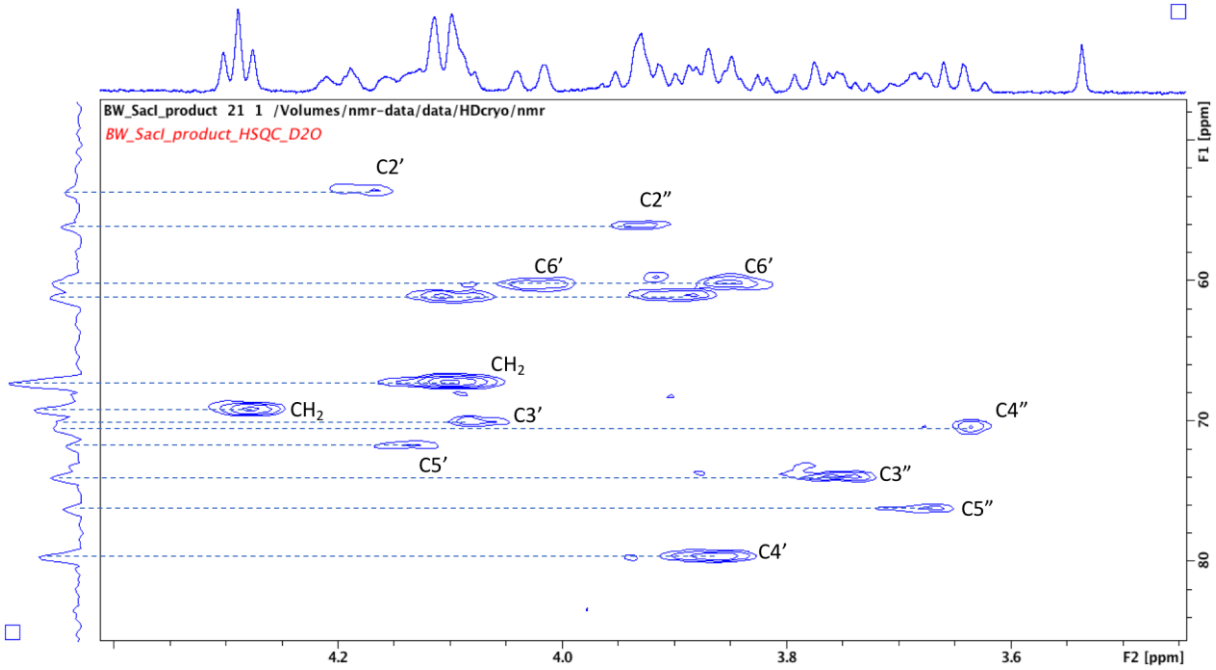


Figure S10: 2D HSQC spectra used to assign carbon signals arising from both α- and β-linked GlcNAc residues. Note the signal for C4' is significantly shifted (79.6 ppm) due to the presence of the glycosidic linkage at this position.

Figure S11: Universal distribution of Agl24 homologs in Crenarchaeota, white the exception of the Order Thermoproteaces. The BLAST analyses (https://blast.ncbi.nlm.nih.gov) of Agl24 (Q4J9C3) with the restriction to Crenarchaeota (Taxid:28889; 126 genomes), revealed 100 sequences (30-100% sequence identity), lacking any homology within the 27 genomes of the Order Thermoproteaces. The orders of Crenarchaeaota are background colored: Fervidicoccales (green), Acidilobales (blue), Desulfurococcales (orange), Sulfolobales (yellow), Thermoproteales (red).

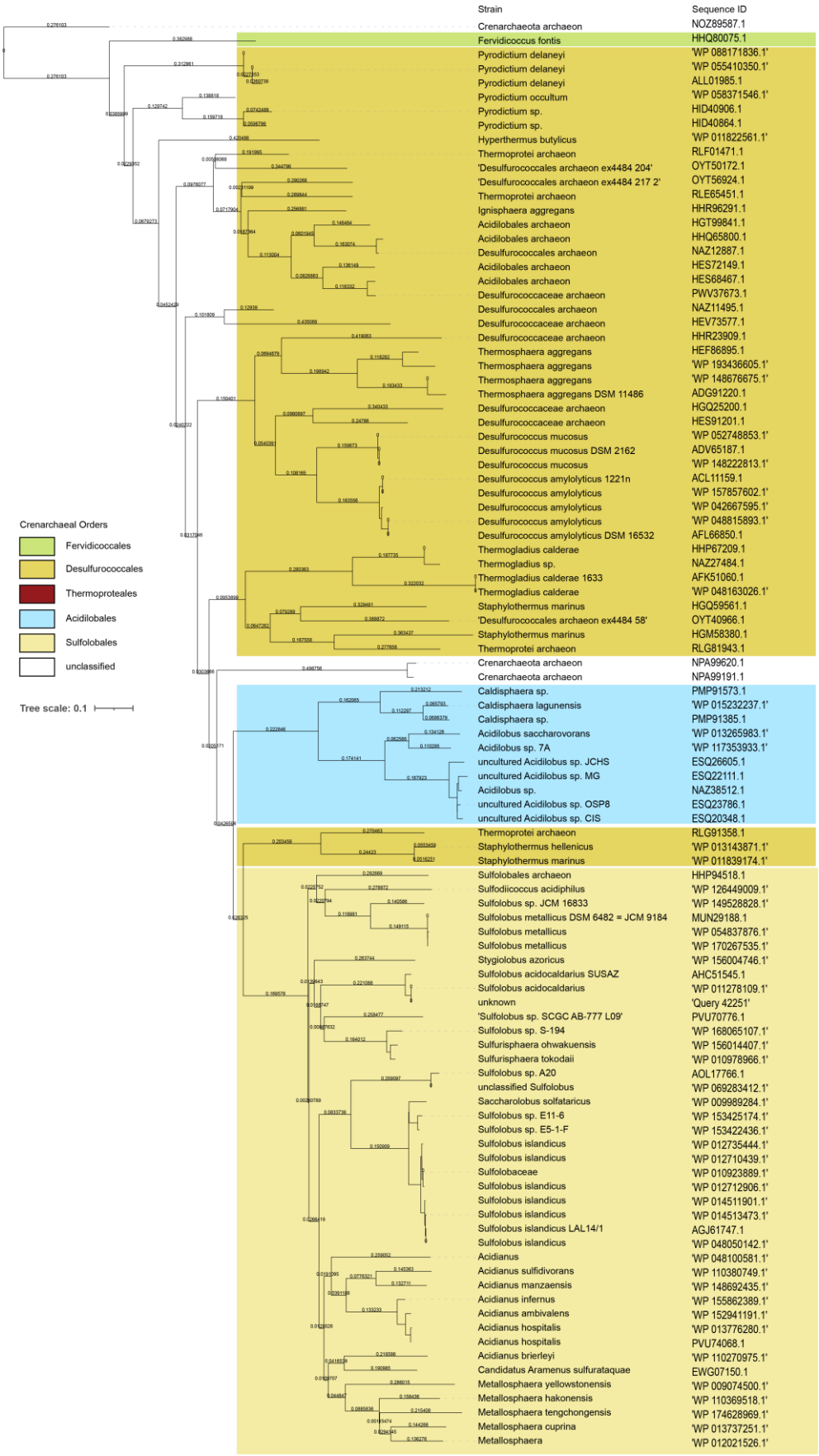


Table S3: Strains and plasmids used in this study.

Strains	Genotype	Reference
E. coli		
DH5α	F– ϕ80lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17(rK–, mK+) phoA supE44 λ– thi-1 gyrA96 relA1	(Taylor et al., 1993)
BL21 (DE3)	F^- *ompT hsdS*$_B$ (r_B^-, m_B^-) *gal dcm* (DE3)	(Studier and Moffatt, 1986)
ER1821	F- glnV44 e14-(McrA-) rfbD1? relA1? endA1 spoT1? thi-1 Δ(mcrC-mrr)114::IS10, MM294 background	(Meselson and Yuan, 1968)
Sulfolobus acidocladarius		
MW001	*S. acidocaldarius* DSM 639, Δ*pyr*E	(Wagner et al., 2012)
BM-A305	*S. acidocaldarius MW001* with integrated deletion plasmid pSVA1312 incl. *pyr*EF	*This study*
BM-A306	*S. acidocaldarius MW001* with integrated deletion plasmid pSVA1312 incl. *pyr*EF	*This study*
BM-A719	*S. acidocaldarius MW001* with pSVA1337 ($MAL_{promotor}$ *saci1262,* C-terminal Strep- and His-Tag)	*This study*
BM-A720	*S. acidocaldarius MW001* with pSVA1339 ($MAL_{promotor}$ *saci1262*, N-terminal Strep- and His-Tag)	*This study*
BM-A724	*S. acidocaldarius MW001* with pBM-0078 ($ARA_{promotor}$ *saci1262* C-terminal Strep- and His-Tag)	*This study*
Plasmids for generating deletion mutants		
pSAV0407	Gene targeting plasmid, pGEM-T Easy backbone, *pyrEF* cassette of *S. solfataricus*	(Wagner et al., 2012)
pSVA1312	In-frame deletion of *Agl24* (*saci1262*) cloned into pSVA407 with *Apa*I and *Bam*HI	*This study*
pSVA3338	Linear *saci1262*($_{short}$)-*pyrEF-saci1262* cut with *BamH*I and *Apa*I into pSVA407	*This study*
Expression plasmids for *S. acidocaldarius*		
pSVA1481	*E. coli* vector with Ara-promoter and C-terminal Strep- and His-tag based on pGEM-T Easy backbone and pMZ1 cassette, pMZ: Amp^r, cloning vector containing replicon ColE1 (pBR322) and a 10X His-Strep tag sequence at the 3' end of the MC	Wagner & Albers
pSVA1518	Expression vector with $ARA_{promotor}$ and c-terminal Strep-His_{10} tag	Wagner & Albers
pSVA1431	Expression vector with $MAL_{promotor}$ and c-terminal Strep-His_{10} tag	Wagner & Albers
pSVA2301	Expression vector with $MAL_{promotor}$ and N-terminal His_{10}- Strep- tag	Wagner & Albers
pSVA1336	*Agl24* cloned into pSVA1481 ($ARA_{promotor}$)with a C-terminal Strep- and His-tag with *Nco*I and *Pst*I	*This study*
pSVA1337	*Agl24* with c-terminal Strep- and His-Tag derived from pSAV1336 subcloned into pSVA1431 ($MAL_{promotor}$) with *NcoI* and *EagI*	*This study*
pSVA1339	Agl24 cloned into pSVA2301 with N-terminal His_{10} Strep - TEV into pSVA2301 ($MAL_{promotor}$) with *NcoI* and *NotI* (Primer 4180 and 4181*)*	*This study*
pBM-0078	Ara promotor saci1262 with C-terminal Strep- and His-tag, cloned with *NcoI ApaI*	*This study*
Expression plasmids for *E. coli*		
*p*Waldo	T7 expression vector containing TEV site-GFP-His_8	*(Waldo et al., 1999)*
*pHD*0499	pWaldo containg *Agl24*-GFP-His_8 cloned with *Xho*I and *Kpn*I	*This study*
pHD0554	pWaldo containg *Agl24*-H_{14}A-GFP-His_8 cloned with *Xho*I and *Kpn*I	*This study*
pHD0586	pWaldo containg *Agl24*-E_{114}A-GFP-His_8 cloned with *Xho*I and *Kpn*I	*This study*

Table S4: Primers used in this study.

Primer	Sequence (5´-3´)	Restriction Site
ΔAgl24		
4168	ACTA GGGCCC GGTCTTGTGCTTAAATTCACTCTGACATC	*Apa*I
4163	GTC TAA AAA TGT GGC TCC TCC ACT GGC GAT AAT CAG TAA TGG GTT GTC	
4164	ATC GCC AGT GGA GGA GCC ACA TTT TTA GAC GAT CCC TCG ACC TGG GAC 4165	
4165	CGCCGA GGATCC CAC AAA TAT ATT CTC CCA AGA GTC TGG C	*BamH*I
pSVA3338 Linear fragment *Agl24$_{up}$-pyrEF-Agl24$_{down}$*		
4168	ACTA GGGCCC GGTCTTGTGCTTAAATTCACTCTGACATC	*ApaI*
6336	GTA ACA TAT CTT TGC TAA ATC GGT CAT TTT CGG GTA TTA C	
4115	GTAATACCCGAAAATGACCGATTTAGCAAAGATATGTTACTCGATTACGCTAGAAAA TTTGAGCAGTTCTAGTACTTGGCTCAAAGAATG	
4116	GGT GTA CCT ATT TCG AAT TGG TTA GGT TTT CTA ACA TTT TGG ATA CTA TCT CAC TAA TTT CAT TTT TTC CTA AAA ATT GCT CCT TTA CAT TTC	
6337	GTATCCAAAATGTTAGAAAACCTAACCAATTCGAAATAGGTACACC	
6338	CGCCGA GGATCC GAA TGC CCA ATT TTT TCA TTT CTG CAA TAG TCT GTT C	*BamHI*
***S. acidocaldarius* Expression**		
pSVA1336		
4176	CTGCA CCATGGG T ATCGACAACCCATTACTGATTATCGCCAGTGG	*NcoI*
4177	CCGTT CTGCAG C TCT AAG AAA TTC TGC TAA TTC ACT TAA AAT TAT ATC	*PstII*
pSVA1339		
4180	CTGCA CCATGGG TCGACAACCCATTACTGATTATCGCCAGTGG	*NcoI*
4181	CCGTT GCGGCCGC TTA TCT AAG AAA TTC TGC TAA TTC ACT TAA AAT TAT ATC	*NotI*
E. coli* Expression of *agl24		
A596	*saci1262*-fwd-*XhoI* AGGAGA CTCGAG ATGATCGACAACCCATTACTGATTATCGCC	*XhoI*
A597	*saci1262*-rev-*KpnI* GATCCA GGTACC TCTCTAAGAAATTCTGCTAATTCACTTAAAATTATATC	*KpnI*
A684	*saci1262*_H_{14}A-*XhoI*-fwd AGGAGACTCGAGATGATCGACAACCCATTACTGATTATCGCCAGTGGAGGAGGGGCTACTGGCTTCGCTAGAGCTATTGC	*XhoI*
A685	*saci1262*_E_{114}A-fwd GTGCACTTTATGTAACAGCTAGCCAAGACAGAATTATTAC	

A686	*saci1262*_E_{114}A-rev GTAATAATTCTGTCTTGGCTAGCTGTTACATAAAGTGCAC	

Table S4: Overview of the properties of the archaeal, eukaryal and bacterial N-glycosylation. Highlighted in grey background color are the mammalian orthologs of the non-catalytic OST subunits found in *Saccharomyces cerevisiae.* *based on sequence analyses of metagenome-assembled genomes of Asgardarchaeota (Zaremba-Niedzwiedzka et al., 2017), **not present in STT3 proteins from organisms that express single subunit OSTs. ***chitobiose is not present in Thermoproteales, but showing an modified version: (GlcA(NAc)2-β-1,4-(Glc(NAc)2

Domain / Component	**Eukarya**	**Archaea**			**Bacteria**
		Asgardarchaeota	Crenarchaeota	Euryarchaeota	
Lipid carrier	Dol	Dol	Dol	Dol	Und
Linkage between N-glycan and lipid carrier	PP	?	PP	P	PP
Nucleotide activated sugar donor	yes	?	yes	yes	yes
Lipid activated sugar donor	yes	?	yes	yes	no
Oligosaccharyltransferase	Stt3/ STT3A/STT3B	AglB (Stt3-like)	AglB (Stt3-like)	AglB (PglB-like)	PglB
DK motif (DXXKXXX(M/I)	yes	yes	yes	yes	
DKi motif (D/E< >KXXXM/I/P)				yes	
MI motif (MXXIXXX(I/V/W)				yes	yes
double sequon DNXTZNXS/T	yes**	yes	yes		
Non-catalytic OST subunits	Ost1/RPN1 Ost5/TMEM258 Swp1/RPN2 Wbp1/DDOST Ost2/DAD1 Ost4 / Ost4 Ost6/TUSC3 or Ost3/MAGT1	RPN1(ribophorin-1)* Ost5* Wbp1* Ost6-like/Ost3-like*	n.d.	n.d.	n.d.
First enzymes in the glycosylation	Alg7 Agl14/13	?	AglH Agl24	diverse	PglC PglA
N-glycan linking sugar(s)	Chitobiose (GlcNAc- β 1,4- GlcNAc -β1-)	?	Chitobiose***	diverse	GalNAc-α1,3-*Bac*-β1
N-glycan structure	tri-branched	?	di- and tri-branched	linear, di-, tri-branched	linear

Supplementary Data 1:

Seed sequences, homology searches, preliminary phylogenies, final datasets (individual homologs), final datasets (full and trimmed alignments), and final phylogenies for the phylogenetic analysis

Supplementary References

BROWN, K., VIAL, S. C., DEDI, N., WESTCOTT, J., SCALLY, S., BUGG, T. D., CHARLTON, P. A. & CHEETHAM, G. M. 2013. Crystal structure of the *Pseudomonas aeruginosa* MurG: UDP-GlcNAc substrate complex. *Protein Pept Lett,* 20**,** 1002-8.

DROZDETSKIY, A., COLE, C., PROCTER, J. & BARTON, G. J. 2015. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res,* 43**,** W389-94.

HA, S., WALKER, D., SHI, Y. G. & WALKER, S. 2000. The 1.9 angstrom crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Science,* 9**,** 1045-1052.

HU, Y., CHEN, L., HA, S., GROSS, B., FALCONE, B., WALKER, D., MOKHTARZADEH, M. & WALKER, S. 2003. Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proc Natl Acad Sci U S A,* 100**,** 845-9.

MESELSON, M. & YUAN, R. 1968. DNA Restriction Enzyme from E. coli. *Nature,* 217**,** 1110-1114.

RAMAN, S., LANGE, O. F., ROSSI, P., TYKA, M., WANG, X., ARAMINI, J., LIU, G., RAMELOT, T. A., ELETSKY, A., SZYPERSKI, T., KENNEDY, M. A., PRESTEGARD, J., MONTELIONE, G. T. & BAKER, D. 2010. NMR structure determination for larger proteins using backbone-only data. *Science,* 327**,** 1014-8.

SIEVERS, F. & HIGGINS, D. G. 2018. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci,* 27**,** 135-145.

STUDIER, F. W. & MOFFATT, B. A. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol,* 189**,** 113-30.

TAYLOR, R. G., WALKER, D. C. & MCLNNES, R. R. 1993. *E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Research,* 21**,** 1677-1678.

WAGNER, M., VAN WOLFEREN, M., WAGNER, A., LASSAK, K., MEYER, B. H., REIMANN, J. & ALBERS, S. V. 2012. Versatile genetic tool Box for the crenarchaeote *Sulfolobus acidocaldarius*. *Front Microbiol,* 3**,** 214.

WALDO, G. S., STANDISH, B. M., BERENDZEN, J. & TERWILLIGER, T. C. 1999. Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol,* 17**,** 691-5.

WANG, X., WELDEGHIORGHIS, T., ZHANG, G., IMPERIALI, B. & PRESTEGARD, J. H. 2008. Solution structure of Alg13: the sugar donor subunit of a yeast N-acetylglucosamine transferase. *Structure,* 16**,** 965-75.

WATERHOUSE, A., BERTONI, M., BIENERT, S., STUDER, G., TAURIELLO, G., GUMIENNY, R., HEER, F. T., DE BEER, T. A. P., REMPFER, C., BORDOLI, L., LEPORE, R. & SCHWEDE, T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res,* 46**,** W296-W303.

ZAREMBA-NIEDZWIEDZKA, K., CACERES, E. F., SAW, J. H., BACKSTROM, D., JUZOKAITE, L., VANCAESTER, E., SEITZ, K. W., ANANTHARAMAN, K., STARNAWSKI, P., KJELDSEN, K. U., STOTT, M. B., NUNOURA, T., BANFIELD, J. F., SCHRAMM, A., BAKER, B. J., SPANG, A. & ETTEMA, T. J. 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature,* 541**,** 353-358.