RESEARCH ARTICLE

Figures and figure supplements

Agl24 is an ancient archaeal homolog of the eukaryotic N-glycan chitobiose synthesis enzymes

Benjamin H Meyer *et al*

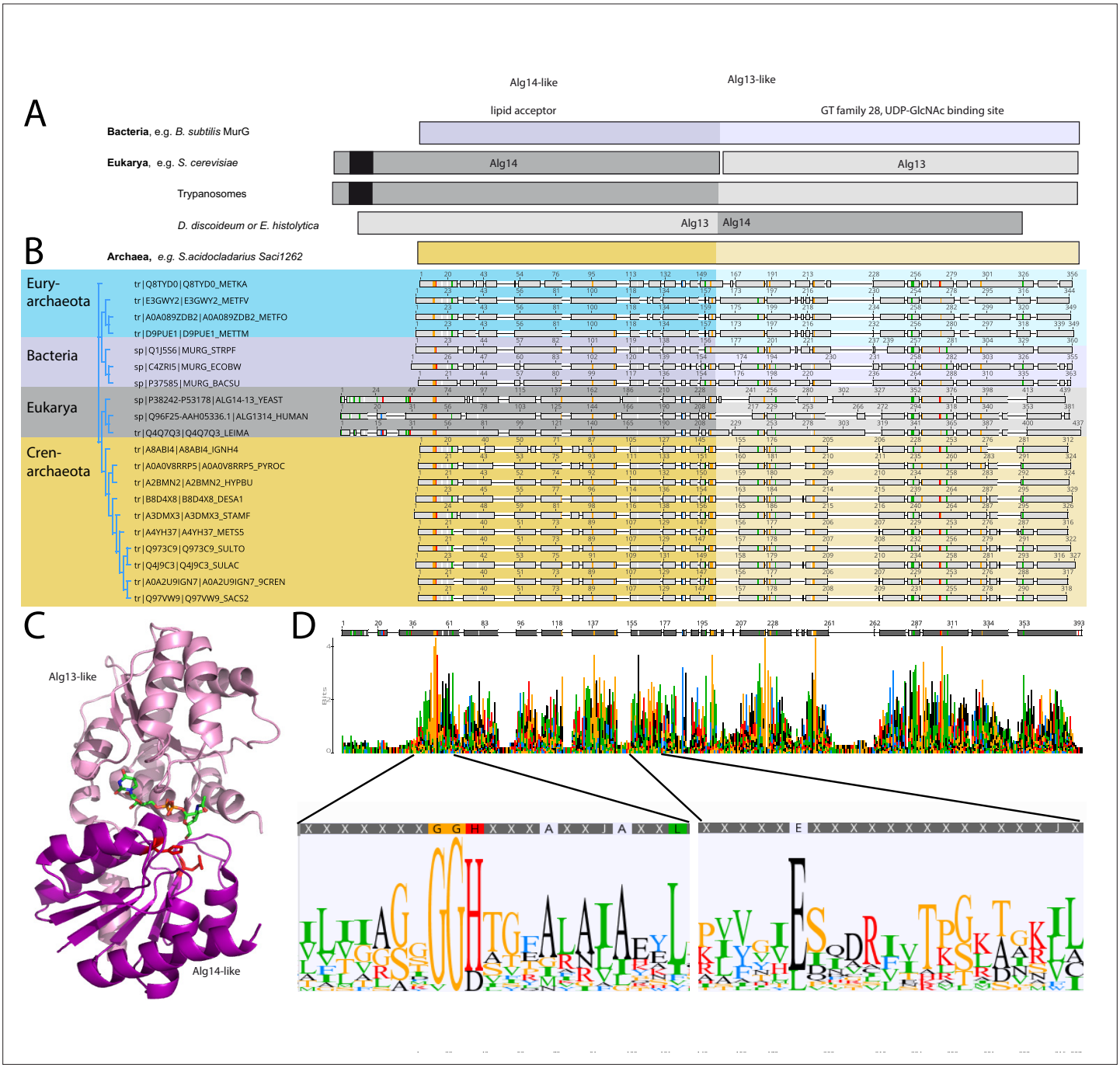

Figure 1. Comparison of the archaeal Saci1262, bacterial MurG, and eukaryotic Alg14-13. (**A**) Simplified comparison of the Alg14-like (darker) or Alg13-like (light color) domain in bacterial MurG (violet), eukaryotic Alg14-13 (gray), and archaeal Saci1262 (orange) homologs. The transmembrane domain is depicted in black. (**B**) Protein sequences were aligned with Clustal Omega (***Sievers and Higgins, 2018***) (for the full alignment, see ***Figure 1—figure supplement 2***). Sequences derived from (i) three bacterial (violet background) MurGs: *Escherichia coli* (C4ZRI5), *Streptococcus pyogenes* (Q1J5S6), and *Bacillus subtilis* (P37585); (ii) eukaryotic (gray background) Alg14-13: *Leishmania major* (Q4Q7Q3), an artificial Alg14-13 fusion from *Saccharomyces cerevisiae* (P38242-P53178) and *Homo sapiens* (Q96F25-Q9NP73-2); (iii) the crenarchaeal (orange background) Saci1264 homologs: *Ignicoccus hospitalis* (A8ABI4), *Pyrodictium occultum* (A0A0V8RRP5), *Hyperthermus butylicus* (A2BMN2), *Desulfurococcus amylolyticus* (B8D4 × 8), *Staphylothermus marinus* (A3DMX3), *Metallosphaera sedula* (A4YH37), *Sulfurisphaera tokodaii* (Q973C9), *Sulfolobus acidocaldarius* (Q4J9C3), *Acidianus brierleyi* (A0A2U9IGN7), and *Saccharolobus solfataricus* (Q97VW9). Selected sequences from pseudomurein producing Euryarchaeota with higher sequence similarity to the MurG are aligned (turquoise background): *Methanopyrus kandleri* (Q8TYD0), *Methanothermus fervidus* (E3GWY2), *Methanobacterium formicicum* (A0A089ZDB2), *Methanothermobacter marburgensis* (D9PUE1). Conserved amino acids (65% threshold) are highlighted in color corresponding to the amino acids: G, S, T, P (orange), K, R, H (red), F, W,Y (blue), A, D, E, C, Q, N (black), and V, L, I, M (green). The end of the Alg14-like domain and start of

Figure 1 continued on next page

Figure 1 continued

Alg13-like domains are indicated by the change from dark to light background color. (**C**) MurG (PDB: 3s2u; shown as ribbon) bound to UDP-GlcNAc (shown as sticks, green carbon, blue nitrogen, red oxygen) and orientation toward the cell membrane (bottom). The Alg13-like domain (light violet) and Alg14-like domain (dark violet) are labelled. Conserved His and Glu residues in Alg14 are shown as red sticks. (**D**) Strict consensus (65%) and Weblogo from the sequence alignment shown in (B). Bar heights 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$_{114}$ (for the full Weblogo, see ***Figure 1—figure supplement 2***).

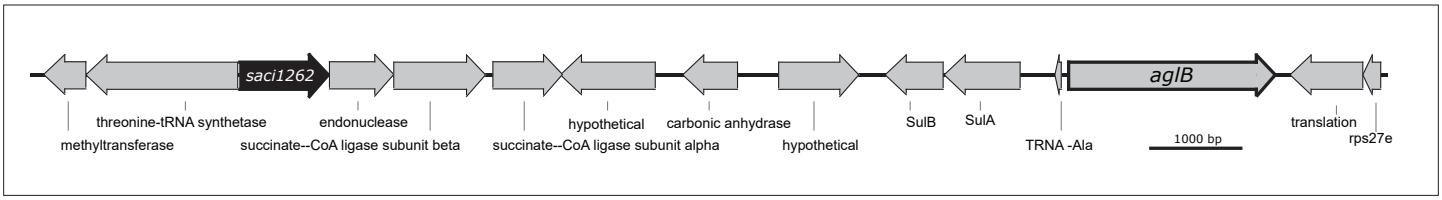

Figure 1—figure supplement 1. Physical map of the gene region adjacent to *saci1262 and aglB* of *Sulfolobus acidocaldarius*. Illustrated are the genes *Saci1260* until *Saci1276*. The gene *saci1262* (SACI_RS06030), displayed in black, is annotated to encode a polysaccharide biosynthesis protein. The gene *aglB* (*saci1274,* SACI_RS06085), displayed with a bold border, encodes the *N*-oligosaccharyltransferase, catalyzing the transfer of the lipid-linked N-glycan onto the target protein.

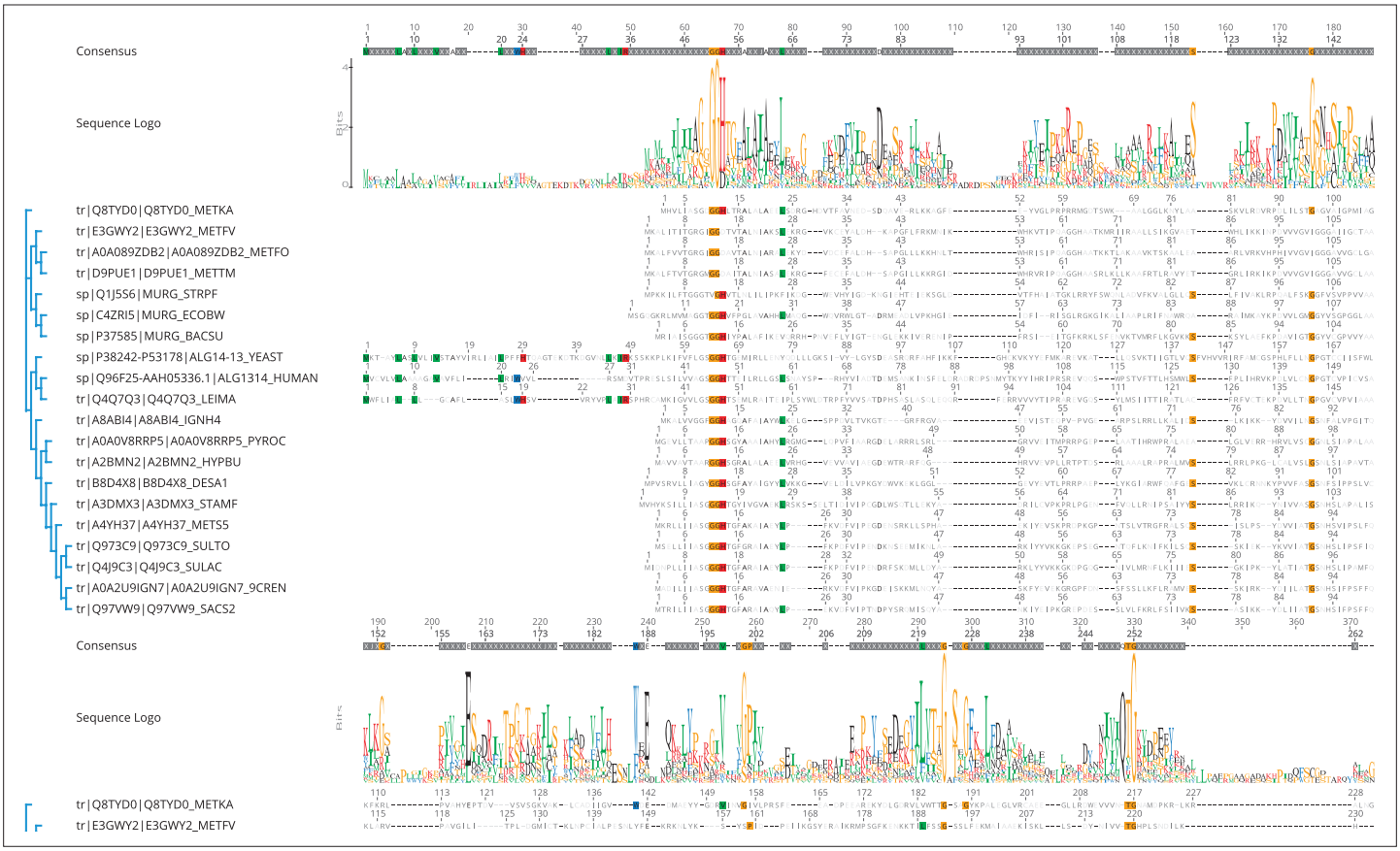

Figure 1—figure supplement 2. Protein sequences were aligned with Clustal Omega (***Sievers and Higgins, 2018***). Sequences derived from (i) three bacterial (violet background) MurGs: *Escherichia coli* (C4ZRI5), *Streptococcus pyogenes* (Q1J5S6), and *Bacillus subtilis* (P37585); (ii) eukaryotic (gray background) Alg14- 13: *Leishmania major* (Q4Q7Q3), an artificial Alg14-13 fusion from *Saccharomyces cerevisiae* (P38242-P53178) and *Homo sapiens* (Q96F25- Q9NP73-2); (iii) the crenarchaeal (orange background) Saci1264 homologs: *Ignicoccus hospitalis* (A8ABI4), *Pyrodictium occultum* (A0A0V8RRP5), *Hyperthermus butylicus* (A2BMN2), *Desulfurococcus amylolyticus* (B8D4 × 8), *Staphylothermus marinus* (A3DMX3), *Metallosphaera sedula* (A4YH37), *Sulfurisphaera tokodaii* (Q973C9), *Sulfolobus acidocaldarius* (Q4J9C3), *Acidianus brierleyi* (A0A2U9IGN7), and *Saccharolobus solfataricus* (Q97VW9). Selected sequences from pseudomurein producing Euryarchaeota with higher sequence similarity to the MurG are aligned (turquoise background): *Methanopyrus kandleri* (Q8TYD0), *Methanothermus fervidus* (E3GWY2), *Methanobacterium formicicum* (A0A089ZDB2), *Methanothermobacter marburgensis* (D9PUE1). Conserved amino acids (65% threshold) are highlighted in color corresponding to the amino acids: G, S, T, P (orange), K, R, H (red), F, W, Y (blue), A, D, E, C, Q, N (black), and V, L, I, M (green). The end of the Alg14-like domain and start of Alg13-like domains are indicated by the change from dark to light background color. Boxed: area of the sugar donor-binding site, in *E. coli* MurG the amino acids A264, L265, T266, E269, Q288, and Q289 have been shown to interact with UDP-GlcNAc.

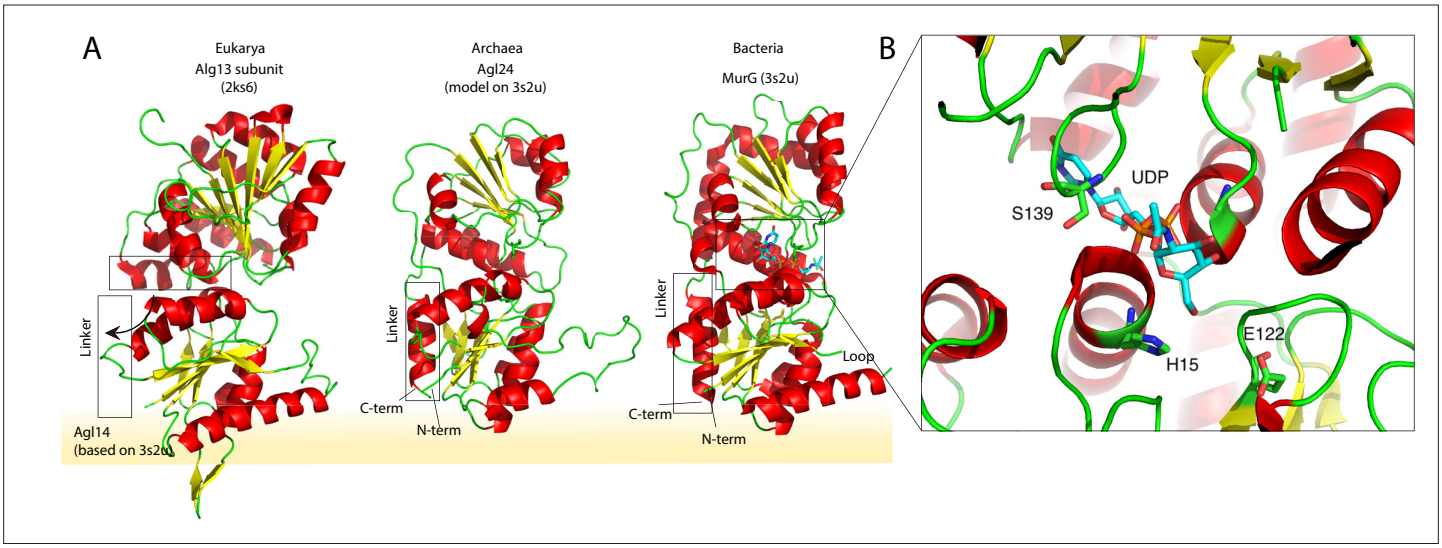

Figure 2. Structural comparison of the eukaryotic and bacterial homologs of the archaeal Saci1262. (**A**) Structural comparison of the eukaryotic Alg13 (PDB:2ks6) with a structural model of Alg14, the bacterial MurG (PDB: 3s2u), and the structural model of the crenarchaeal Saci1262 (archaeal glycosylation enzyme 24 [Agl24]). Structural models were built with SWISS-MODEL (***Waterhouse et al., 2018***). Detailed results are listed in ***Supplementary file 2***. 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). The catalytic site is located in the cleft of both domains, with conserved H15 and E122 residues shown in sticks with green carbon, red oxygen, and blue nitrogen atoms.

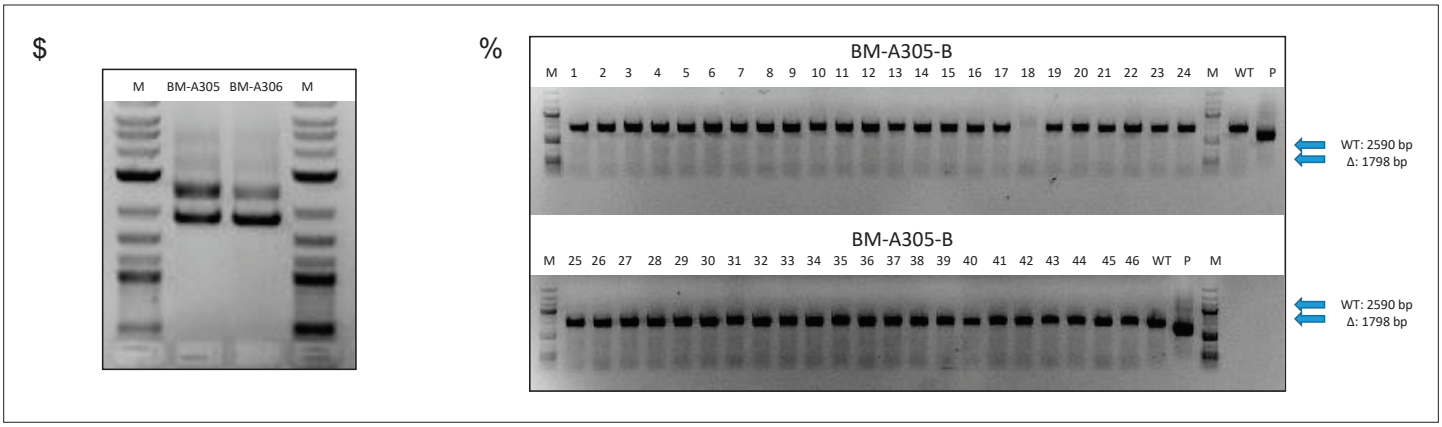

Figure 2—figure supplement 1. Confirmation of the integration and segregation of the *saci1262* deletion plasmid pSVA1312 in *Sulfolobus acidocaldarius* MW001. (**A**) The integration of the *saci1262* deletion plasmid pSVA1312 in *S. acidocaldarius* MW001 (first selection) was monitored by PCR using the out primers of the upstream and downstream region of *saci1262* 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 *saci1262* gene, respectively. (**B**) The segregation of pSVA1312 (second selection) was confirmed by PCR using the outer primers of the flanking region of *saci1262* and the genomic DNA from second selection colonies. All PCR fragments gained from genomic DNA of second selection colonies correspond to the full-length *saci1262* gene (2590 bp), while a deletion would result in a 1798 bp PCR fragment.

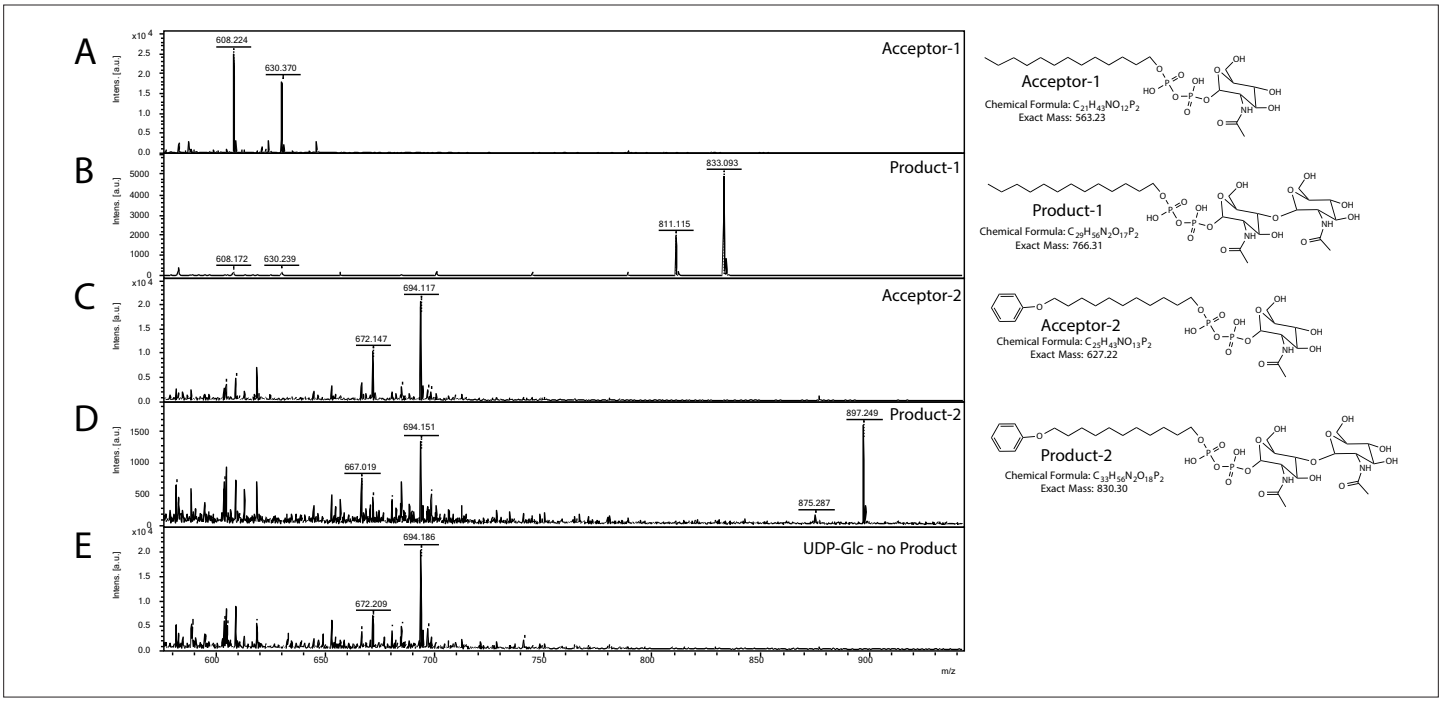

Figure 3. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) spectra of the in vitro Saci1262 reaction assessing the *N*-acetylglucosaminetransferase activity. Spectra obtained from the enzymatic reactions containing Saci1262-GFP and (**A**) acceptor-1, (**B**) acceptor-1 and UDP-GlcNAc, and Saci1262-GFP, (**C**) acceptor-2, (**D**) acceptor-2, and UDP-GlcNAc, (**E**) acceptor-2 and UDP-Glc. The conversion from acceptor-1 (608 m/z $[M\text{-}1H + 2Na]^+$ and 630 m/z $[M\text{-}2H + 3Na]^+$) or acceptor-2 (672 m/z $[M\text{-}1H + 2Na]^+$ and 694 m/z $[M\text{-}2H + 3Na]^+$) to the product (811 m/z $[M\text{-}1H + 2Na]^+$ and 833 m/z $[M\text{-}2H + 3Na]^+$) or (875 m/z $[M\text{-}1H + 2Na]^+$ and 897 m/z $[M\text{-}2H + 3Na]^+$) was observed only when UDP-GlcNAc was used as nucleotide sugar donor.

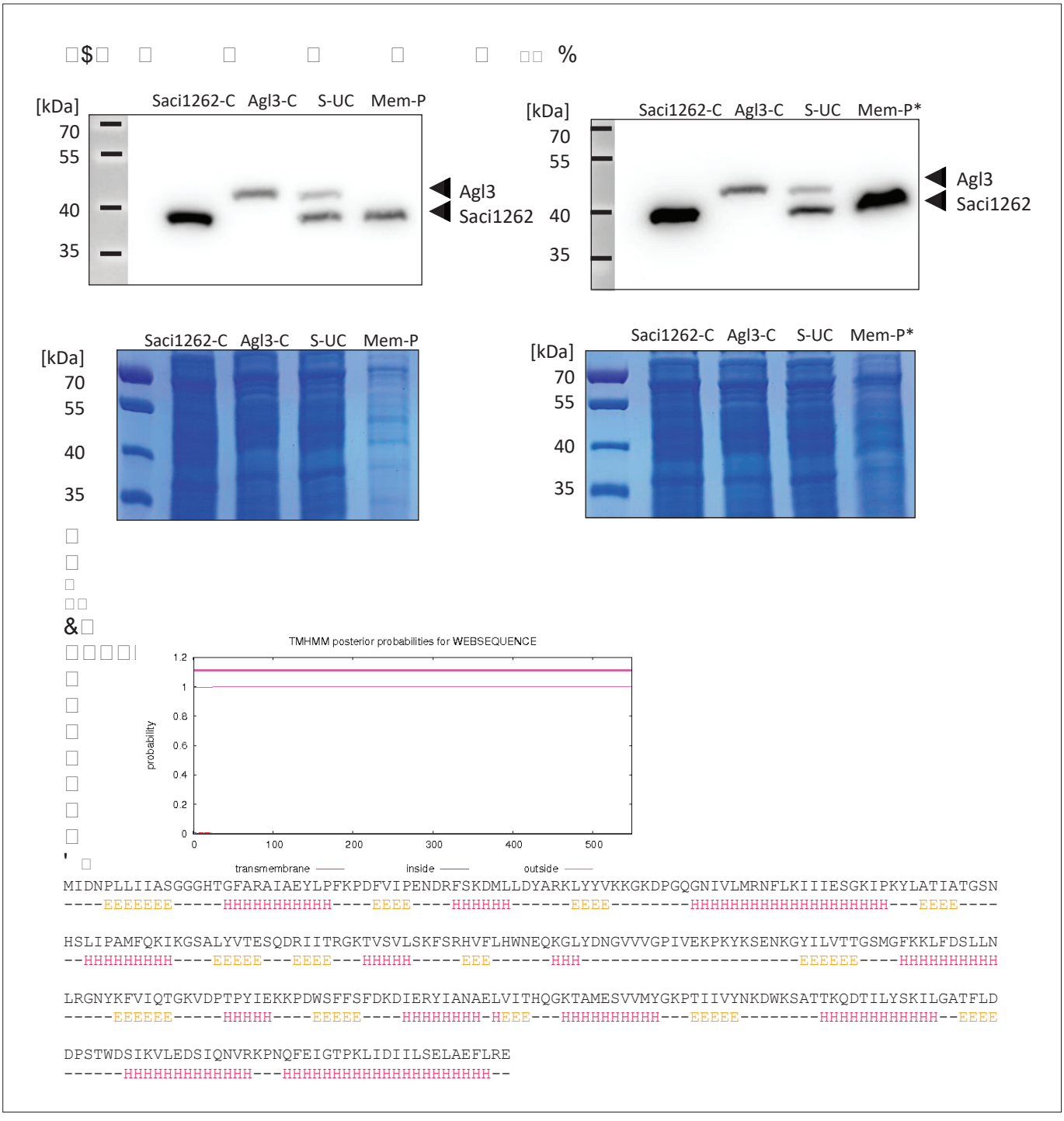

Figure 3—figure supplement 1. Anti-His Immunoblot indicates membrane association of Saci1262, while Alg3 (sulfoquinovose synthase) is only found in the soluble fraction. (**A**) Different fractions of *Sulfolobus acidocaldarius* cells expressing the His-tagged version of Saci1262 (39.7 kDa) and Agl3 (48.1 kDa) or the mixture of the two strains (1:1 ratio) were loaded onto a 7.5% SDS-PAGE. The loaded fractions correspond to the resuspended cells expressing *saci1262* (Saci1262-C) or *agl3* (Agl3-C), and after mixture (in a 1:1 ratio) and cell disruption of the two strains, the supernatant after ultracentrifugation (S–UC), and membrane pellet (Mem-P). One SDS-PAGE was transferred onto PVDF membrane and probed with Anti-His Immunoblot (upper panel), while the second SDS-PAGE was stained with Coomassie brilliant blue as a loading control (lower panel). (**B**) As the expression of *saci1262* was slightly higher than *agl3* (based on optimized Ara-promotor), the amount of the membrane fraction was increased threefold (Mem-P*) to exclude a missing signal of Agl3 in A. (**C**) Prediction of transmembrane helices in archaeal glycosylation enzyme 24 (Agl24), by the TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) confirmed the lack of any transmembrane domain. (**D**) Secondary protein structure prediction by the Jpred Server. The top line corresponds to the amino acid sequence of Saci1262/Agl24, the second line represents extended (**E**), helical (**H**), and other (-) types of secondary structure, respectively (***Drozdetskiy et al., 2015***).

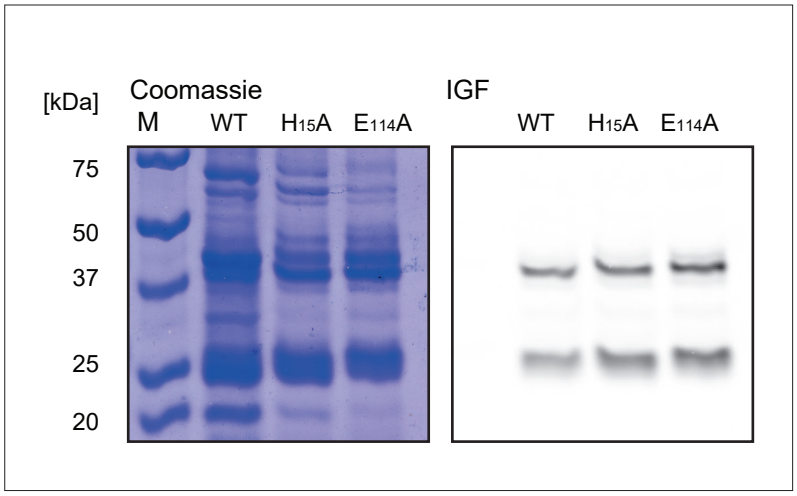

Figure 3—figure supplement 2. SDS-PAGE of the purified Agl24-WT-GFP, mutants Agl24-H15A-GFP, and Agl24-E114A-GFP from *Escherichia coli*, stained with Coomassie Brilliant Blue or visualized by in-gel fluorescence (IGF).

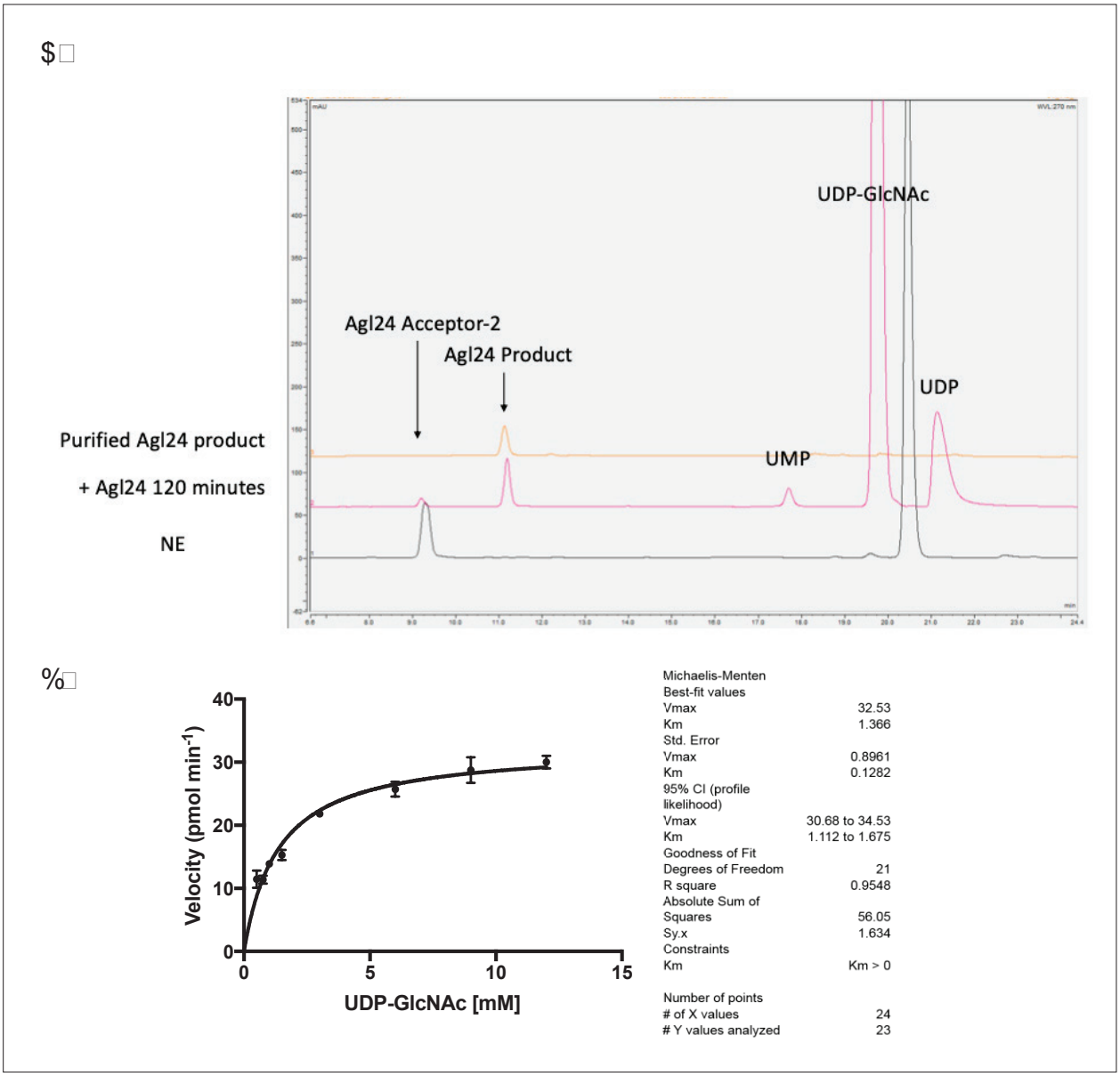

Figure 3—figure supplement 3. HPLC chromatogram and kinetic parameters of the archaeal glycosylation enzyme 24 (Agl24) enzymatic reaction. (**A**) Representative HPLC chromatogram of the archaeal glycosylation enzyme 24 (Agl24) enzymatic reaction product formation. NE = no enzyme control; UMP = uridine monophosphate; UDP = uridine diphosphate; UDP-GlcNAc = uridine diphosphate *N*-acetylglucosamine. The top trace shows a reinjection of the purified Agl24 product used for subsequent NMR analysis. (**B**) Agl24 Kmapp for UDP-GlcNAc determination.

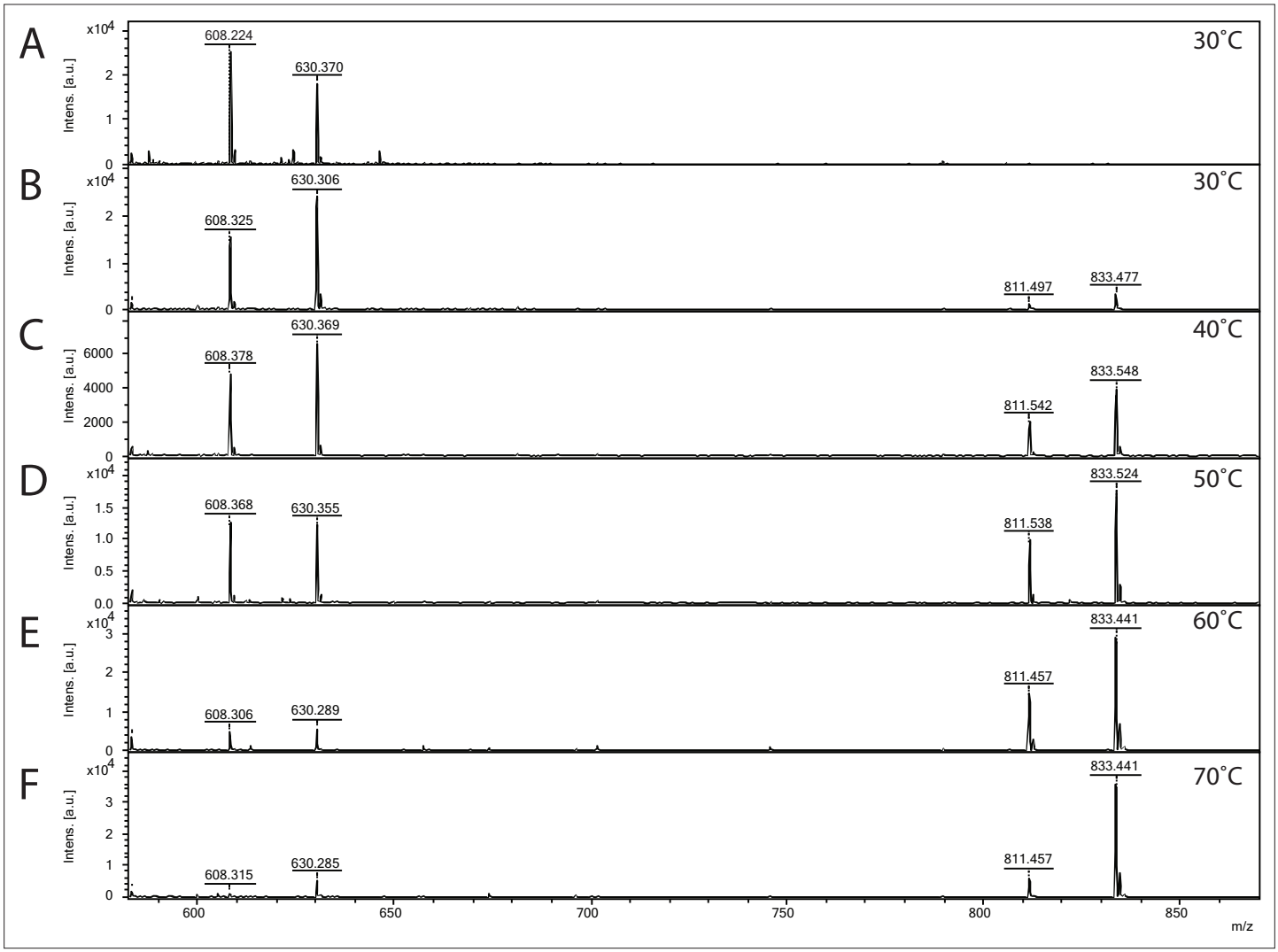

Figure 4. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) spectra of the in vitro archaeal glycosylation enzyme 24 (Agl24) reaction at different temperatures. Spectra obtained from the purified enzymatic reaction mix with only acceptor-1 (**A**), acceptor-1, UDP-GlcNAc, and Agl24-GFP wild-type (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]) toward the product (811 m/z [M-1H + 2Na] and 833 m/z [M-2H + 3Na]) is dependent on the applied temperature. Above 50°C, the increase in the level of the product at the expense of the level of the acceptor is clearly visible.

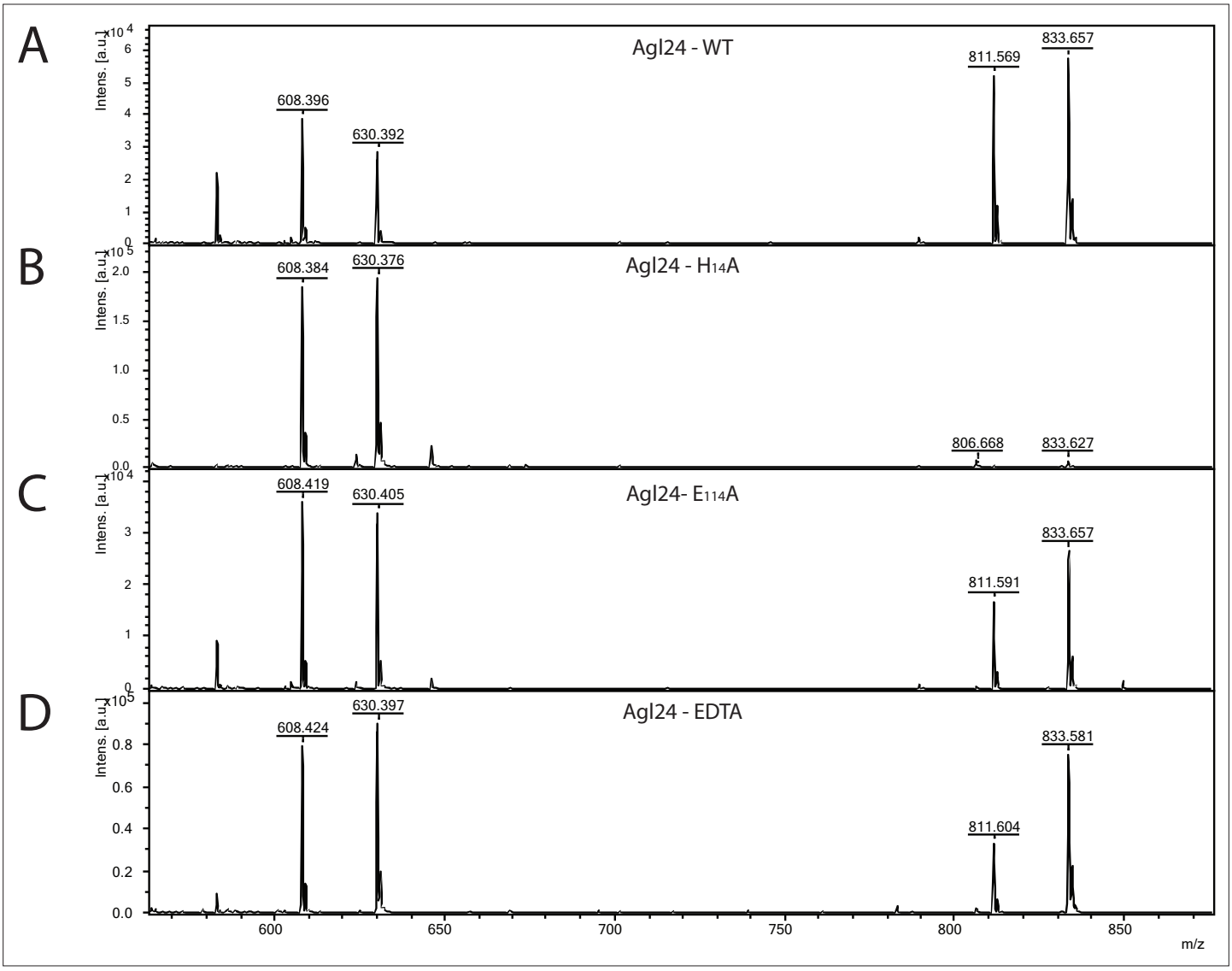

Figure 5. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) spectra of the in vitro archaeal glycosylation enzyme 24 (Agl24) reaction, the generated point mutations H14A and E114A and EDTA control. Spectra obtained from the purified enzymatic reaction mix with acceptor-1 (608 m/z [M-1H + 2Na] and 630 m/z [M-2H + 3Na]), UDP-GlcNAc and (**A**) Agl24-GFP wild-type (WT), (**B**) Agl24-H_{14}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_{14}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.

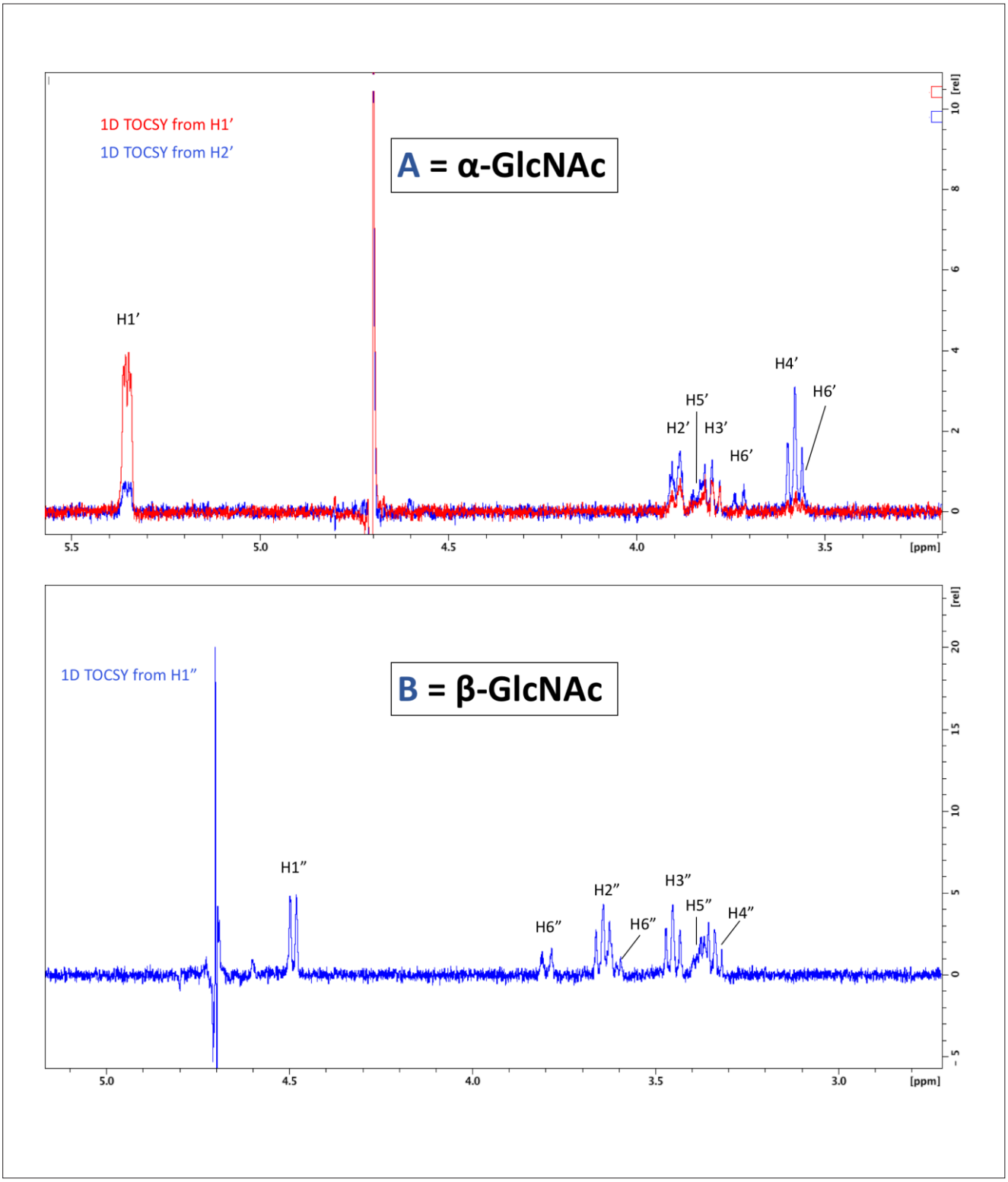

Figure 5—figure supplement 1. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) spectra of the in vitro archaeal glycosylation enzyme 24 (Agl24) reaction to assess the metal dependency for the specificity activity. Spectra obtained from the purified enzymatic reaction mix with acceptor-1, UDP-GlcNAc, and Agl24-GFP without the 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.

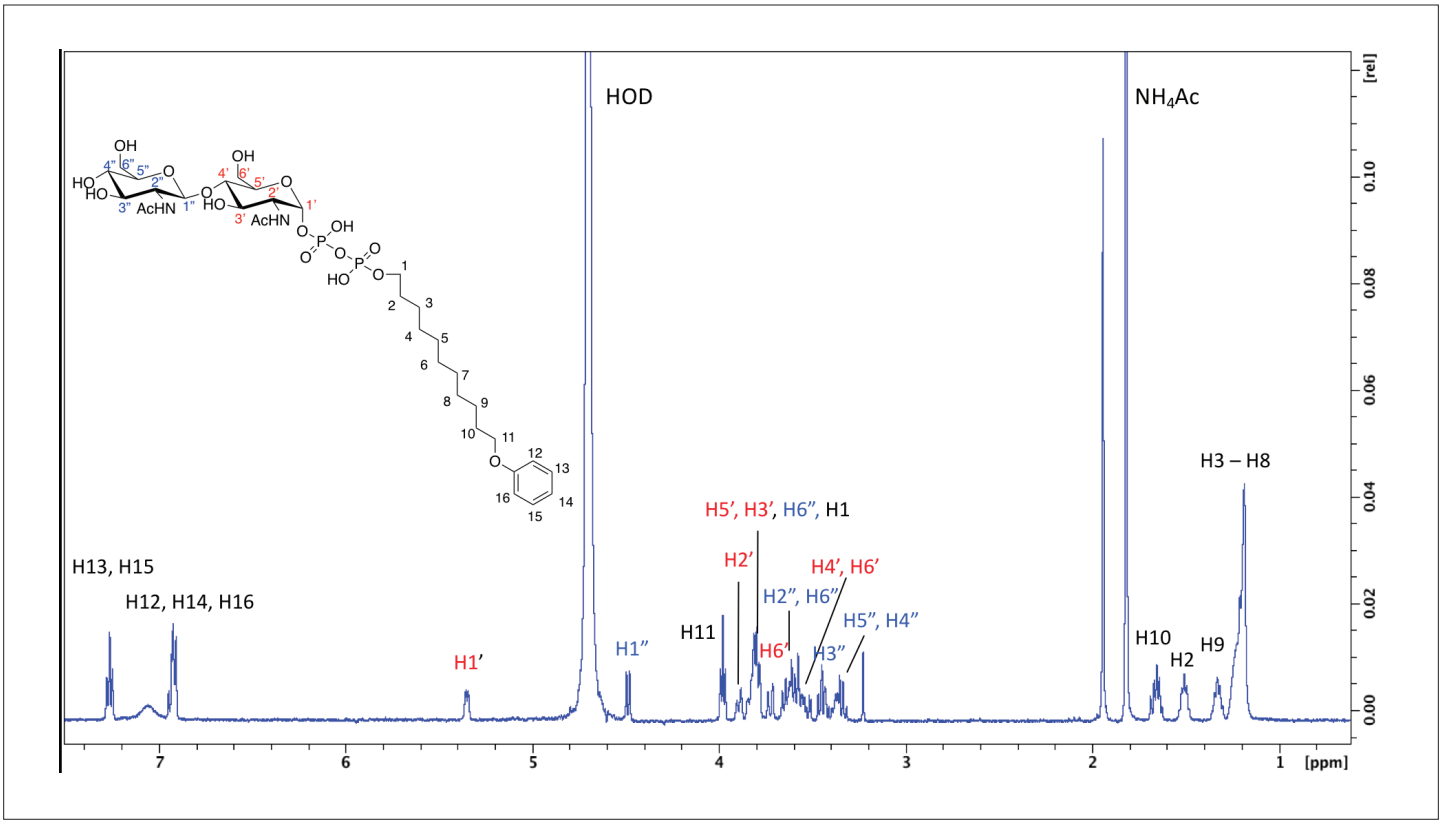

Figure 6. [1] H NMR spectra of purified archaeal glycosylation enzyme 24 (Agl24) enzymatic product. Spectra were acquired in D_2O at 293 K on a Bruker AVANCE III HD 500 MHz NMR spectrometer fitted with a 5 mm QCPI cryoprobe. Chemical shifts are reported with respect to the residual HDO signal at δ_H 4.70 ppm.

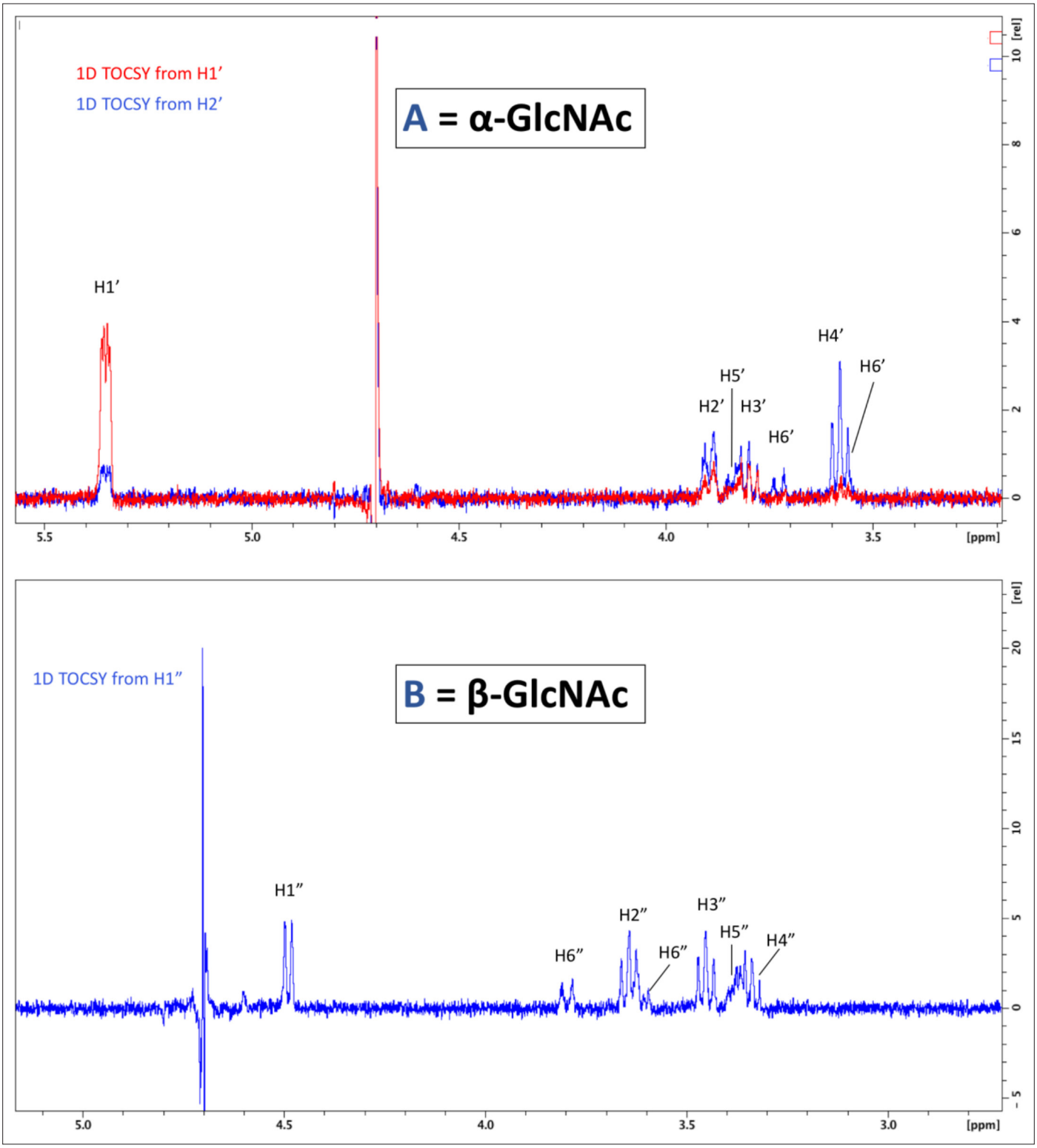

Figure 6—figure supplement 1. 1D total correlation spectroscopy (TOCSY) correlation spectra. To visualize all protons of the α-linked GlcNAc residue, irradiation was set for the shift of H1′ and H2′ (top). To visualize protons of the β-linked GlcNAc, irradiation of H1″ was sufficient to see all protons (bottom).

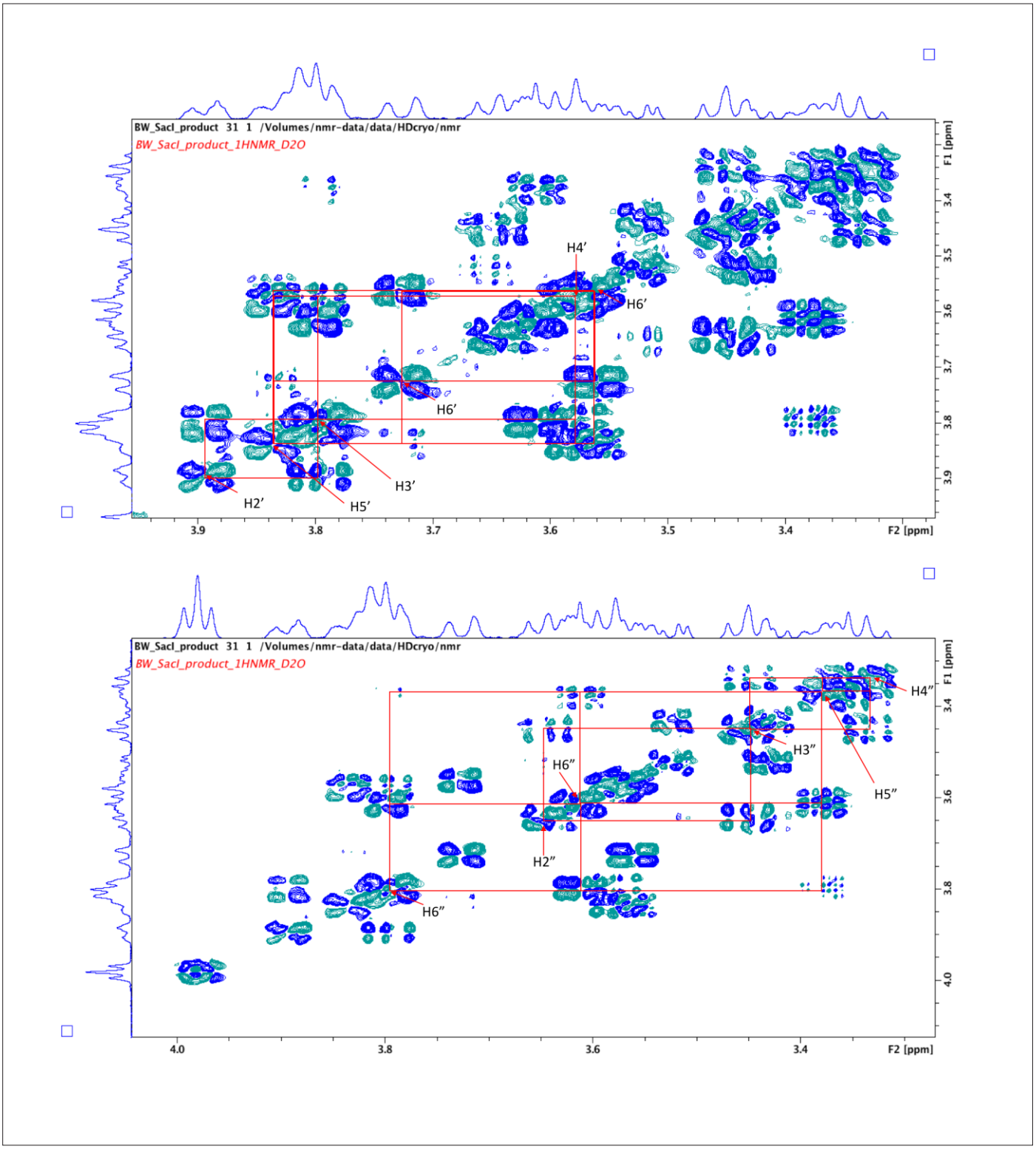

Figure 6—figure supplement 2. 2D COSY spectra used to assign proton signals arising from α-GlcNAc (top) and β-GlcNAc (bottom).

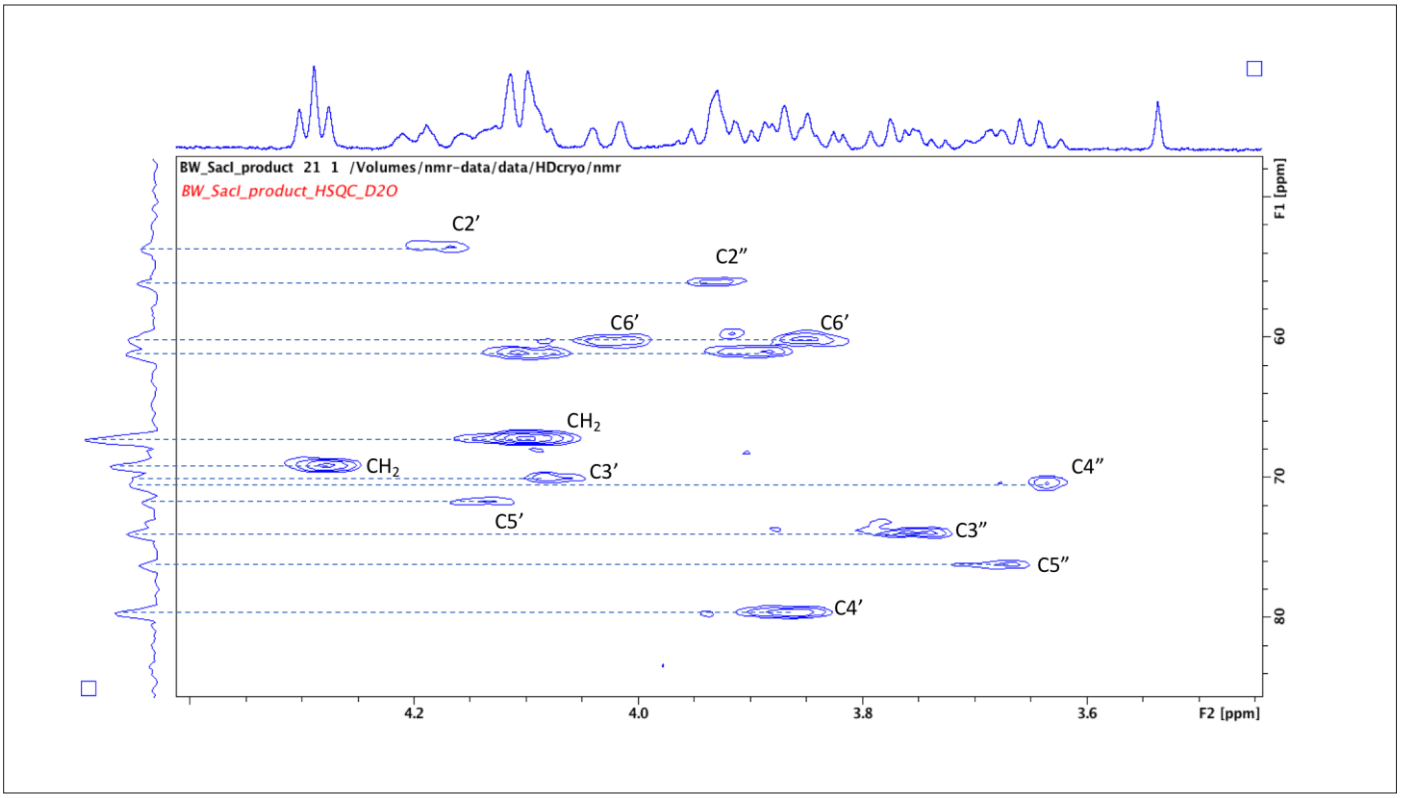

Figure 6—figure supplement 3. 2D HSQC spectra used to assign carbon signals arising from both α- and β-linked GlcNAc residues. Note the signal for C4′ is significantly shifted (79.6 ppm) due to the presence of the glycosidic linkage at this position.

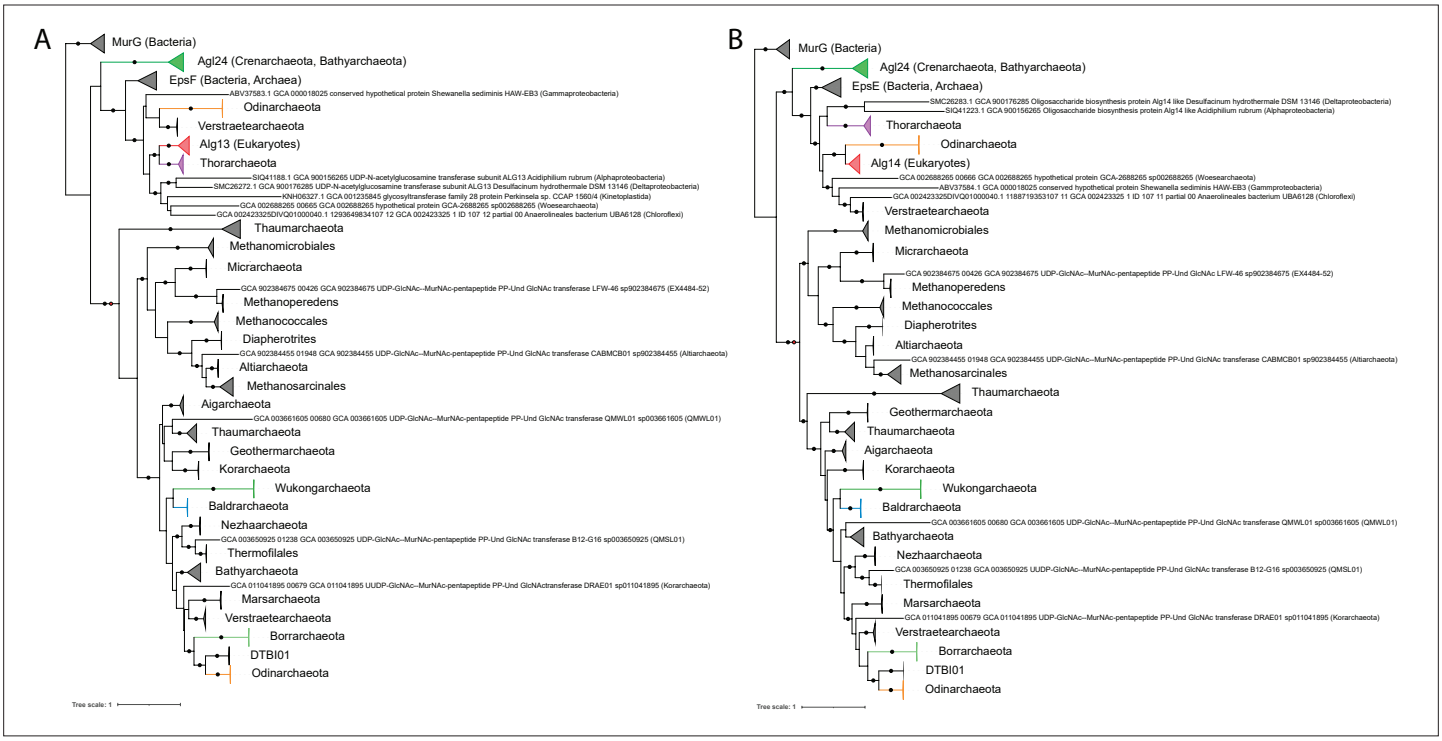

Figure 7. Single gene phylogenies of (**A**) Alg13/EpsF and (**B**) Alg14/EpsE homologs. The phylogenies were rooted using the MurG sequences from the corresponding (***Lombard, 2016***) datasets as the respective outgroup. Black dots indicate strongly supported branches (ultrafast bootstrap ≥ 95 and aLRT SH-like support ≥ 80). Red dots indicate the position of the minimal ancestor deviation (MAD) outgroup-free root in phylogenies without the MurG outgroup (doi.org/10.7910/DVN/9KSWQR). Clades of interest (archaeal glycosylation enzyme 24 [Agl24]-like containing Saci1262, Asgard lineages, and Eukaryotes) are colored. Taxa names are given according to common usage, with a few exceptions where the Genome Taxonomy Database (GTDB) classification was necessary for accuracy. The uncollapsed trees with numerical branch supports are given in ***Figure 7—figure supplement 1*** and ***Figure 7—figure supplement 2***, respectively. The concatenated dataset tree and its uncollapsed version are given in ***Figure 7—figure supplement 3*** and ***Figure 7—figure supplement 4***, respectively.

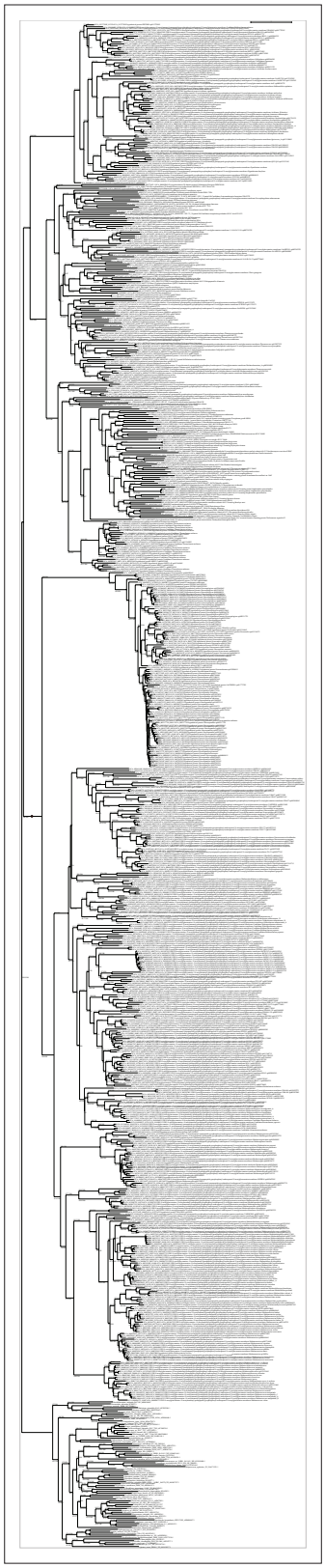

Figure 7—figure supplement 1. Uncollapsed phylogeny of Alg13/EpsF. The tree was rooted using the MurG sequences from the corresponding (***Lombard, 2016***) dataset as outgroup. Branch supports

Figure 7—figure supplement 1 continued on next page

Figure 7—figure supplement 1 continued

are given as SH-aLRT (%)/ultrafast bootstrap (%). The red dot indicates the position of the minimal ancestor deviation (MAD) outgroup-free root in phylogenies without the MurG outgroup.

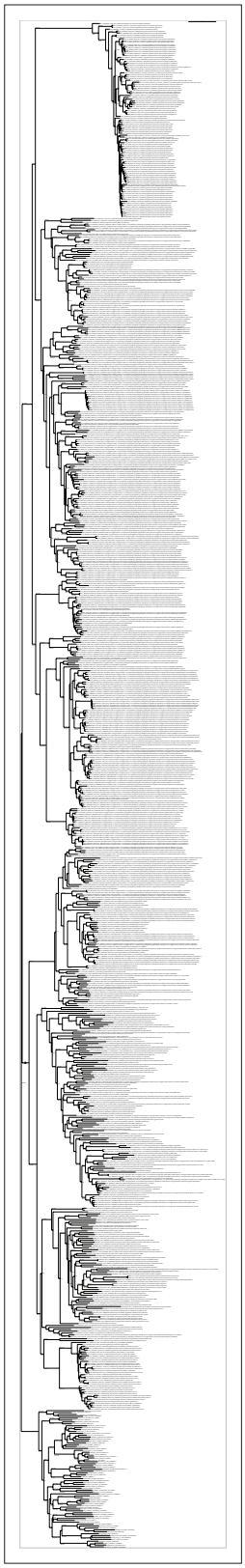

Figure 7—figure supplement 2. Uncollapsed phylogeny of Alg14/EpsE. The tree was rooted using the MurG sequences from the corresponding (***Lombard, 2016***) dataset as outgroup. Branch supports

Figure 7—figure supplement 2 continued on next page

Figure 7—figure supplement 2 continued

are given as SH-aLRT (%)/ultrafast bootstrap (%). The red dot indicates the position of the minimal ancestor deviation (MAD) outgroup-free root in phylogenies without the MurG outgroup .

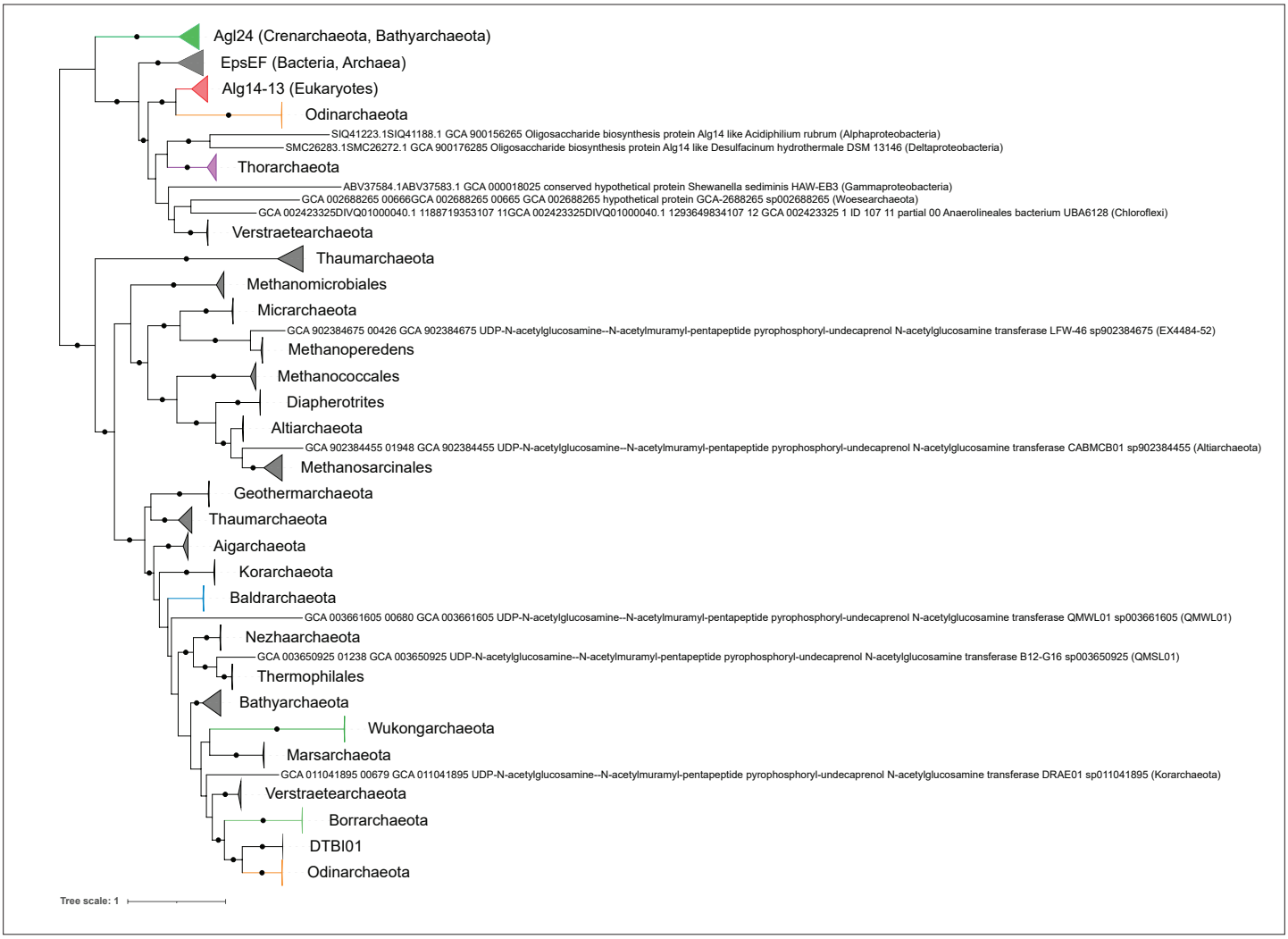

Figure 7—figure supplement 3. Phylogeny of concatenated Alg14-13/EpsEF. Black dots indicate strongly supported branches (ultrafast bootstrap ≥ 95 and aLRT SH-like support ≥ 80). Clades of interest (archaeal glycosylation enzyme 24 [Agl24]-like containing Saci1262, Asgard lineages, and Eukaryotes) are colored. Taxa names are given according to common usage, with a few exceptions where the Genome Taxonomy Database (GTDB) classification was necessary for accuracy. The tree was rooted by the minimal ancestor deviation (MAD) outgroup-free method.

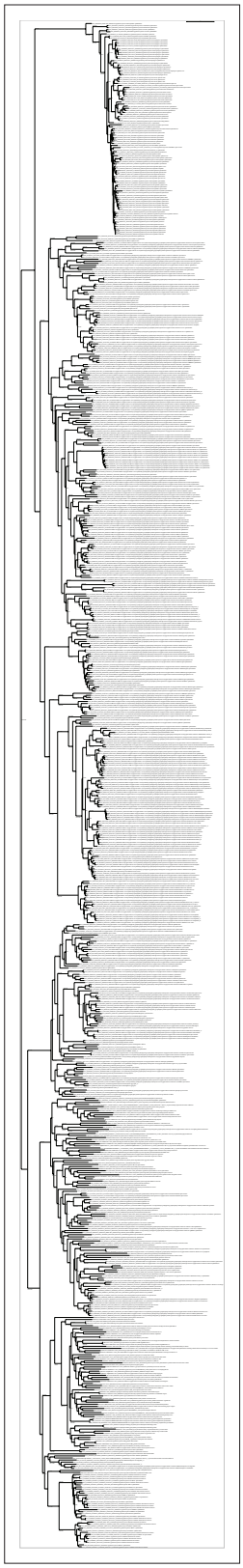

Figure 7—figure supplement 4. Uncollapsed phylogeny of concatenated Alg14-13/EpsEF. The tree was rooted by the minimal ancestor deviation (MAD) outgroup-free method. Branch supports are given as SH-aLRT (%)/ultrafast bootstrap (%).

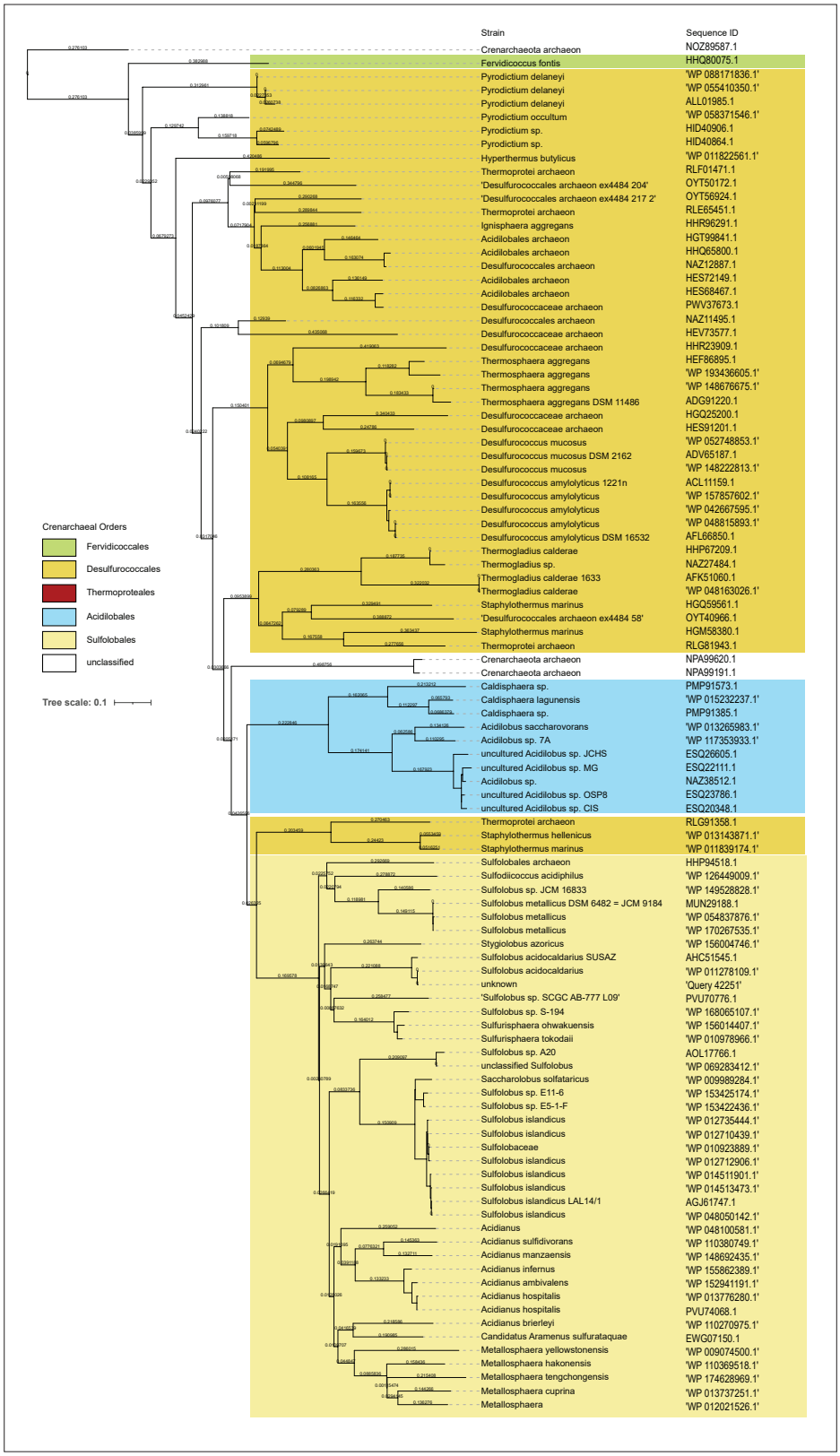

Figure 7—figure supplement 5. Universal distribution of archaeal glycosylation enzyme 24 (Agl24) homologs in Crenarchaeota, except for the Order Thermoproteaces. The BLAST analyses (https://blast.ncbi.nlm.nih.gov) of Agl24 (Q4J9C3) with the restriction to Crenarchaeota (Taxid:28889; 126 genomes) revealed 100 sequences (30–100% sequence identity), lacking any homology within the 27 genomes of the order Thermoproteaces. The

Figure 7—figure supplement 5 continued on next page

Figure 7—figure supplement 5 continued

orders of Crenarchaeaota are background colored: Fervidicoccales (green), Acidilobales (blue), Desulfurococcales (orange), Sulfolobales (yellow), Thermoproteales (red).