Structure of the two-component S-layer of the archaeon *Sulfolobus acidocaldarius*

Lavinia Gambelli[1,2†], Mathew McLaren[1,3], Rebecca Conners[1,3], Kelly Sanders[1,3], Matthew C. Gaines[1,3], Lewis Clark[1], Vicki A. M. Gold[1,3], Daniel Kattnig[1,2], Mateusz Sikora[4,5], Cyril Hanus[6,7], Michail N. Isupov[8], Bertram Daum[1,3*]

[1] Living Systems Institute, University of Exeter, Exeter, EX4 4QD, United Kingdom.

[2] Faculty of Environment, Science and Economy, University of Exeter, Exeter, EX4 4QF, United Kingdom.

[3] Faculty of Health and Life Sciences, University of Exeter, Exeter, EX4 4QD, United Kingdom.

[4] Department of Theoretical Biophysics, Max Planck Institute for Biophysics, 60438 Frankfurt, Germany

[5] Malopolska Centre of Biotechnology, Gronostajowa 7A, 30-387 Kraków, Poland.

[6] Institute of Psychiatry and Neurosciences of Paris, Inserm UMR1266 - Université Paris Cité, 75014 Paris, France

[7] GHU Psychiatrie et Neurosciences de Paris, 75014 Paris, France

[8] Henry Wellcome Building for Biocatalysis, Biosciences, Faculty of Health and Life Sciences, University of Exeter, Exeter, EX4 4QD, United Kingdom.

† Current address: Medical Research Council Laboratory of Molecular Biology, Francis Crick Ave, Cambridge, CB2 0QH.

*Corresponding author. Email: b.daum2@exeter.ac.uk

Keywords: archaea, *Sulfolobus acidocaldarius*, surface layer, electron cryo-microscopy, single particle analysis, electron cryo-tomography, subtomogram averaging.

Abstract

Surface layers (S-layers) are resilient two-dimensional protein lattices that encapsulate many bacteria and most archaea. In archaea, S-layers usually form the only structural component of the cell wall and thus act as the final frontier between the cell and its environment. Therefore, S-layers are crucial for supporting microbial life. Notwithstanding their importance, little is known about archaeal S-layers at the atomic level. Here, we combined single particle cryo electron microscopy (cryoEM), cryo electron tomography (cryoET) and Alphafold2 predictions to generate an atomic model of the two-component S-layer of *Sulfolobus acidocaldarius*. The outer component of this S-layer (SlaA) is a flexible, highly glycosylated, and stable protein. Together with the inner and membrane-bound component (SlaB), they assemble into a porous and interwoven lattice. We hypothesise that jackknife-like conformational changes, changes play important roles in S-layer assembly.

Introduction

The prokaryotic cell envelope includes a cytoplasmic membrane and a cell wall, which provide structural integrity to the cell and mediate the interaction between the extracellular and intracellular environment. The cell wall differs in composition and structure across prokaryotes[1]. In bacteria, a peptidoglycan (murein) layer encapsulates the cytoplasmic membrane, and this is in turn enclosed by a second membrane in Gram-negative bacteria[2]. Generally, the archaeal cell wall lacks an outer membrane, but a variety of cell wall elements, including pseudomurein, methanochondroitin and protein sheaths have been described[3]. Most prokaryotes exhibit a porous glycoprotein surface layer (S-layer) as the outermost component of their cell wall[1]. In archaea, S-layers are the simplest and most commonly found cell wall structure[1,3–5].

The prokaryotic cell envelope is exposed to a variety of environmental conditions, which, in the case of extremophiles, can be unforgiving (low/high pH, high temperature and salinity). Therefore, S-layers reflect the cellular need for both structural and functional plasticity, allowing archaea to thrive in diverse ecosystems. Archaeal S-layers maintain the cell shape under mechanical, osmotic and thermal stress, selectively allow molecules to enter or leave the cell, and create a quasiperiplasmic compartment (similar to the periplasmic space in Gram-negative bacteria)[3–5]. S-layer glycoproteins are also involved in cell-cell recognition[6] and mediate virus-host interactions[7,8].

Structurally, an S-layer is a pseudocrystalline array of (glyco)proteins (surface layer proteins, SLPs). The ordered nature of an S-layer is what sets it apart from

other protein sheaths[1–3,9]. S-layers usually consist of thousands of copies of one SLP species. These SLPs self-assemble on the cell surface predominantly at mid-cell[1,10], giving rise to an oblique (p1, p2), square (p4) or hexagonal (p3, p6) symmetry[9]. In archaea, the hexagonal symmetry is the most common[4]. The S-layer is highly porous. Depending on the species, the pores can occupy up to about 70 % of the S-layer surface and have different sizes and shapes[4,9]. Such an assembly provides a highly stable and flexible 2D lattice[11,12]. Archaeal SLPs range from 40 - 200 kDa in molecular mass and show little sequence conservation[1]. The most common post-translational modification of SLPs is glycosylation. Most archaeal SLPs are N- and/or O-glycosylated and the composition of the glycans is highly diverse[4,5]. Thermophilic and hyperthermophilic archaea show a higher number of glycosylation sites on SLPs compared to mesophilic archaea, suggesting that glycans support thermostability[13]. Another common aspect of archaeal S-layers is their binding of divalent metal ions[12,14,15], which have been shown to be essential for S-layer assembly and anchoring in bacteria[16,17]. Atomic models of assembled bacterial S-layers have been reported, including that of *Clostridium difficile*[18], *Caulobacter crescentus*[19,20], *Deinococcus radiodurans*[21], However, archaeal S-layers have been less well explored at this level of detail. So far, atomic models for domains of *Methanosarcina* SLPs[22,23], and more recently, a structure of the *Haloferax volcanii* S-layer have been described[15].

Sulfolobus acidocaldarius is a hyperthermophilic and acidophilic archaeon of the Crenarchaeota phylum and thrives in acidic thermal soils and hot springs worldwide. It grows at pH ~2-3 and temperatures ranging from 65 °C to 90 °C[24]. The *Sulfolobus* S-layer is composed of two repeating glycoproteins, SlaA and SlaB. In *S.*

acidocaldarius, SlaA contains 1,424 amino acids and has a molecular mass of 151 kDa, whereas SlaB comprises 475 amino acids and has a mass of 49.5 kDa[25]. Comparative sequence analysis and molecular modelling predicted that SlaA is a soluble protein rich in β-strands[26]. On the other hand, SlaB has been predicted to contain three consecutive β-sandwich domains at the N-terminus and a membrane-bound coiled-coil domain at the C-terminus[26]. Across the Sulfolobales, SlaA shows higher sequence and structural variability compared to SlaB[26]. Early 2D crystallography and electron microscopy experiments described the *S. acidocaldarius* S-layer as a "smooth", highly porous, hexagonal (p3) lattice[25,27]. Recently, we investigated the architecture of the *S. acidocaldarius* S-layer by electron cryo-tomography (cryoET)[28]. The S-layer has a bipartite organisation with SlaA and SlaB forming the extracellular- and intracellular-facing layers, respectively. Dimers of SlaA and trimers of SlaB assemble around hexagonal and triangular pores, creating a ~30 nm thick canopy-like framework. However, the resolution was limited, and secondary structure details were unresolved. *Sulfolobus* mutants lacking SlaA and/or SlaB show morphological aberrations, higher sensitivity to hyperosmotic stress and alterations of the chromosome copy number, suggesting that in these species the S-layer plays key roles in cell integrity, maintenance, and cell division[29].

Here, we investigated the *S. acidocaldarius* S-layer and its components using a combination of single particle cryoEM and cryoET. We solved the atomic structure of SlaA and investigated its stability across extreme pH ranges. Moreover, we combined cryoEM data and Alphafold2 to build a complete *in situ* atomic model of this S-layer and propose insights into its dynamics and assembly.

Results

Structure and N-glycosylation of $SlaA_{30\text{-}1,069}$ at acidic pH

To solve the structure of the *S. acidocaldarius* SLP SlaA, we disassembled the S-layer by changing the pH from acidic to basic and purified the native protein using size exclusion chromatography. We have previously shown that *S. acidocaldarius* SlaA purified in this way reforms S-layers upon shifting the pH back to acidic[28]. This demonstrates that after disassembly, SlaA remains in a “native”, reassembly-competent form.

CryoEM grids with suspensions of the protein were plunge frozen at pH 4, before the protein had time to reassemble into S-layers. The acidic pH was chosen to account for the natural conditions in which *S. acidocaldarius* thrives. The structure of SlaA was determined from cryoEM movies, using the single particle analysis (SPA) pipeline in Relion 3.1[30] (Fig. 1-Figure Supplement 1, Fig. 1-Figure Supplement 2a,d; Supplementary File 1a). The final cryoEM map had a global resolution of 3.1 Å (Fig. 1-Figure Supplement. 3a, b; 4a).

Because SlaA has virtually no homology with other structurally characterised proteins, the cryoEM map was used to build an atomic model *de novo* (Fig. 1a; Fig. 1-Figure Supplement 4b; Video 1). Residues 30 to 1,069 (~ 70 % of the sequence) were clearly defined in the cryoEM map. The N-terminal signal peptide (predicted to be residues 1-24) is cleaved prior to S-layer assembly[26]. A few N-terminal residues and residues 1,070-1,424 at the C-terminus were not resolved by SPA, likely due to their high flexibility (Fig. 1-Figure Supplement 5a; Video 2). $SlaA_{30\text{-}1,069}$ is a Y-shaped protein. It consists mostly of β-strands and contains only a few short α-helices (Fig. 1a,b, Fig. 1-Figure Supplement 4c,d). The polypeptide chain is arranged into four

domains ($D1_{30-234}$, $D2_{235-660,701-746}$, $D3_{661-700,747-914}$, $D4_{915-1,069}$), as defined by SWORD[31] (Fig. 1c).

Of those domains, only D4 shows significant similarity to known structures – the domain 3 of complement C5 (PDB ID: 4E0S) according to DALI[32]. A disulphide bond links D3 and D4 (Cys_{677}-$Cys_{1,017}$) (Fig. 1-Figure Supplement4d), however, the density of this bond is not visible in the cryoEM map, likely due to electron beam damage[33].

The structure of the missing C-terminus ($SlaA_{914-1,424}$) was predicted (including D4 to aid alignment) using Alphafold[34] and revealed two additional β-domains, D5 and D6, (Fig. 1c, Fig. 1-Figure Supplement6). Alphafold predicted five different conformations of $SlaA_{914-1,424}$, which differed with regards to the position of D5-D6 relative to D1-D4, suggesting an in-plane flexibility between these two parts of the protein around a hinge (amino acids $A_{1,067}$-$L_{1,071}$) between D4 and D5 (Fig. 1c, Fig. 1-Figure Supplement6). Similar conformations were also observed in 2D classes of our cryoEM dataset (Fig. 1-Figure Supplement 5a, Video 2), as well as a low-resolution 3D refinement of SlaA purified from the related species *Saccharolobus solfataricus* (Fig. 1-Figure Supplement 5b,c), substantiating the Alphafold predictions in Fig. 1-Figure Supplement 6. The predicted extremes of the conformational space of SlaA are shown in Fig. 1c and 1d. These describe stretched (“open”) and flapped (“closed”) conformations. The highly variable positions of D5-D6 seen in the 2D classes, suggest that these domains do not adopt discrete positions, but rather move about freely in the soluble form of the SlaA subunit. It is probable that this jackknife-like flexibility aids SlaA’s assembly into an interwoven S-layer. If some of this flexibility is retained in the assembled S-layer, it will enable it to adopt various

degrees of curvature, necessitated by its ability to encapsulate large cells, as well as small exosomes.

SlaA is expected to be highly glycosylated; its sequence contains 31 predicted N-glycosylation sites[35]. Our cryoEM map of $SlaA_{30\text{-}1,069}$ shows 19 glycan densities (Fig. 2), largely in agreement with the prediction of 20 sequons located in this portion of the protein[35]. The 19 glycosylated Asn residues in $SlaA_{30\text{-}1,069}$ are listed in Fig. 2e. The remaining predicted glycosylation sites reside in domains D5 and D6, in which eight sites were confirmed to be glycosylated by mass spectrometry analysis[35]. Therefore, the entire SlaA protein contains a total of 27 confirmed glycans.

The N-glycans were modelled into the cryoEM densities based on their known chemical structure[36]. The complete glycan is a tribranched hexasaccharide, containing a 6-sulfoquinovose (QuiS). Not all glycosylation sites had clear density to model the entire hexasaccharide. Instead, several forms of apparently truncated glycans were fitted into the cryoEM map (Fig. 2b-d). Most glycans (47 %) were built as pentasaccharides, lacking the glucose bound to QuiS in the mature glycan; 15 % of the glycan pool could be modelled with the whole hexasaccharide structure.

As shown for other glycoproteins, such as the spike proteins of coronavirus[37], glycans are usually much more dynamic than polypeptides and rapidly explore large conformational spaces, generating potentially bulky glycan shields over hundreds of nanoseconds. To evaluate the morphology and span of such shields, a reductionist molecular dynamics simulation approach (GlycoSHIELD)[38] was used to graft plausible arrays of glycan conformers onto open and closed conformations of SlaA monomers with D5 and D6 domains (Fig. 2 g,h). Glycan volume occupancy was comparable on the two conformations of the monomers (Fig. 2 g,h).

Both closed and open conformations showed a similar number of possible glycan conformers (with the closed slightly more than the open form; Figure 2-Figure Supplement 1). This signifies that neither SlaA conformation is entropically favoured over the other, which allows for the observed free jackknife movement between D1-4 and D5-6 (Video 2).

SlaA at different pH conditions

SlaA assembly and disassembly are pH-sensitive processes[28]. A pH shift from acidic (~pH 4) to alkaline (~pH 10) induces the disassembly of the lattice into its component subunits, while a reassembly occurs upon shifting the pH back to acidic[28]. Asking whether this pH shift-induced assembly and disassembly mechanism is based on a conformational change or partial unfolding of SlaA, we investigated the structure of SlaA at different pH conditions. Purified SlaA proteins were frozen at pH 7 and pH 10 and their structure was determined using the SPA pipeline in Relion3[39] (Figure 3-Figure Supplement 1a,b; Supplementary File 1a) and 3.1 (Figure 3-Figure Supplement 2, Supplementary File 1a; Figure 1-Figure Supplement 22b-f) . The resulting cryoEM maps had global resolutions of 3.9 Å for SlaA at pH 7 and 3.2 Å for SlaA at pH 10 (Fig. 3a; Figure 1-Figure Supplement 3). As for SlaA at pH 4, domains D5 and D6 were too flexible to be resolved in the cryoEM maps. Strikingly, the cryoEM maps of $SlaA_{30\text{-}1,069}$ at the three pH conditions were virtually identical, demonstrating a remarkable pH stability of this protein. The mean r.m.s.d. value of Ca atoms between the pH 4 and pH 10 structures was 0.79 Å (min. = 0.02 Å; max. = 2.6 Å) (Fig. 3b; Video 3), confirming that $SlaA_{30\text{-}1,069}$ maintains its structure unchanged across a surprisingly broad pH range. This suggests that a pH-induced conformational change or unfolding in $SlaA_{30\text{-}1,069}$ is not the cause for S-layer

disassembly. However, because D5 and D6 were not resolved in our map, a structural rearrangement affecting these domains remains a possibility.

A variation in pH can dramatically affect protein-protein interactions by changing the overall electrostatic surface potential of the protein complex[40,41]. An analysis of the surface charges of SlaA, including the glycans, at pH 4, 7 and 10 revealed that the overall protein charge changes from positive at pH 4 to negative at pH 10 (Fig. 3c-e). A comparison of the surface charge between glycosylated and non-glycosylated SlaA (Figure 3-Figure Supplement 4) showed that the glycans contribute considerably to the negative charge of the protein at higher pH values. This change in electrostatic surface potential may be a key factor in disrupting protein-protein interactions within the S-layer, causing its disassembly at alkaline pH.

Atomic model of the *S. acidocaldarius* S-layer

In a previous study, we determined the location of SlaA and SlaB within the S-layer lattice by cryoET of whole cells and isolated S-layers[28]. However, due to the limited resolution of the cryoEM maps and the lack of SlaA and SlaB atomic models, the details of the S-layer structure could not be explored. To improve on the available cryoEM map of the *S. acidocaldarius* S-layer and investigate the atomic structure of the lattice, we performed cryoET and subtomogram averaging (STA) on *S. acidocaldarius* exosomes with improved imaging conditions and processing techniques. Exosomes are naturally secreted S-layer-encapsulated vesicles, with a diameter of about 90-230 nm[42]. To analyse the *in situ* structure of the S-layer, we performed STA using Warp[43], Relion 3.1[30] and M[44] and obtained a cryoEM map at 11.2 Å resolution (Figure 4-Figure Supplement 1 and 2). We fitted our structure of

237 SlaA into the S-layer map, which provides an atomic model of the assembled lattice
238 (Fig. 4a,b; Figure 4-Figure Supplement 1 d-i).
239 When observed in the direction parallel to the S-layer plane, the exosome-
240 encapsulating S-layer displays a positive curvature, with a curvature radius of the
241 cryoEM average map of 84 nm (Fig. 4c,d). SlaA assembles into a sheet with a
242 thickness of 95 Å. The SlaA subunits adopt an angle of about 28° with respect to the
243 membrane plane (Fig. 4d). As a result of this inclination, effectively two zones in the
244 SlaA assembly can be distinguished: an outer zone constituted by D1, D2, D3 and
245 D4, and one inner zone formed by D5 and D6 (Fig. 4c,d).
246 Six SlaA monomers assemble around a hexagonal pore of 48 Å in diameter (glycans
247 not included) (Fig. 4a). The D1 domains of these six monomers project into and
248 define the shape of the hexagonal pore, together with D3 and D4 domains. The
249 triangular pores that surround the hexagonal pores have a diameter of ~85 Å and are
250 defined by the D2, D4, D5 and D6 domains of three SlaA molecules (Fig. 4e). The
251 D3 domain of each monomer overlaps with the D4 domain of the following monomer
252 along the hexagonal ring in a clockwise fashion. The D5 and D6 domains of each
253 SlaA subunit project towards the cytoplasmic membrane. Two SlaA monomers
254 dimerise through the D6 domains, with each SlaA dimer spanning two adjacent
255 hexagonal pores (Fig. 4b,e, Figure 4-Figure Supplement 3 and 4). Thus, protein-
256 protein interactions between two adjacent hexagonal pores occur through the
257 dimerising D6 domains of each SlaA dimer and the D2 domains of overlapping SlaA
258 monomers. The SlaA dimer includes an angle of 160° between the two monomers,
259 and a total length of 420 Å (Figure 4-Figure Supplement 3). While SlaA was not
260 resolved as a dimer in our single particle analysis, we could confirm these dimers in
261 tomograms of negatively stained S-layers (Figure 4-Figure Supplement 4), which

show similar dimensions and structure as in our assembly model. Their co-existence with assembled S-layer may indicate that SlaA dimers are an intermediate of S-layer assembly or disassembly.

Modelling of glycan shields in the assembled structure showed that glycans fill large gaps seen between SlaA's globular domains and significantly protrude into the lumen of the triangular and hexagonal pores (Fig. 4 f-h). In the assembled S-layer, the interaction sites between SlaA largely occur via glycosylated surfaces, leaving most glycans unaffected (Figure 2-Figure Supplement 1). Reduction of glycan conformational freedom is overall small between isolated and assembled SlaA monomers. Instead, the glycoshields appear to delineate protein-protein interfaces, which may "guide" the self-assembly of the S-layer, substantiated by the fact that any restriction of glycan flexibility would be entropically unfavourable. Similarly, a glycan-guided assembly mechanism has been suggested for the assembly of cadherins in the desmosome[45].

To get a handle on the structure of the entire S-layer, we used Alphafold v2.2.0[34] and SymmDock[46] and predicted the monomeric and trimeric SlaB structure. The predicted SlaB structure consists of three N-terminal β-sandwich domains and a 132 amino acid long C-terminal α-helix (Fig. 5a,b; Figure 5-Figure Supplement 1). The domain architecture and trimeric arrangement of SlaB agree with the sequence-based molecular modelling described previously[26]. The TMHMM – 2.0 server predicted the N-terminal amino acids 448-470 as transmembrane helix. The hydrophobicity plot (Figure 5-Figure Supplement 2e) confirms a hydrophobic region corresponding to the predicted transmembrane helix (Figure 5-Figure Supplement 2a,e). The protein is predicted to have 14 N-glycosylation sites, of which six are located along the C-terminal α-helix (Figure 5-Figure Supplement 2b-d). The

287 electrostatic surface potential calculated at pH 4 shows that the C-terminal α-helix is
288 mostly neutral (Figure 5-Figure Supplement 2f). In contrast, the three β-sandwich
289 domains have greater electrostatic potential. While D2 is mostly positive, D3 carries
290 distinct negatively charged patches (Figure 5-Figure Supplement 2f). These patches
291 may play a role in electrostatic interactions between SlaB's D3 domain and the
292 mainly positively charged SlaA.

293 By combining SPA and STA with structural predictions, we were able to build
294 a complete *S. acidocaldarius* S-layer model (Fig. 5 c-e; Figure 5-Figure Supplement
295 3). The Alphafold predictions of the SlaB trimer superimposed remarkably well into
296 the corresponding densities visible in our STA map at low threshold values, and
297 flexible fitting using Namdinator[47] further improved the fit (Figure 5-Figure
298 Supplement 3).

299 In the assembled lattice, SlaB trimers occupy alternating triangular pores
300 around each hexagonal pore[28]. The SlaB trimer has a tripod-like shape, with its long
301 axis perpendicular to the membrane plane and that formed by SlaA. Three Ig-like
302 domains branch away from the trimer's symmetry axis and face the SlaA canopy,
303 whereas three α-helices form a coiled coil, which at the predicted transmembrane
304 region insert into the resolved exosome membrane (Figure 5-Figure Supplement 3c).

305 The lattice is a ~35 nm thick macromolecular assembly, in which each SlaB
306 trimer interacts with three SlaA dimers. This interaction is mediated by the positively
307 charged D6 dimerising domains of SlaA and the negatively charged N-terminal Ig-
308 like D3 domains of SlaB.

309

310 **Discussion**

311 The *Sulfolobales* S-layer lattice stands out from others because it is a two-
312 component lattice, consisting of the S-layer-forming SlaA and the membrane anchor
313 SlaB. In 2019, we reported on the structure of the *S. acidocaldarius* S-layer obtained
314 from STA on whole cells and isolated S-layer sheets[28]. With the new information
315 provided in the current study, we were able to improve on the model we proposed
316 previously. The new data confirm the overall p3 S-layer lattice symmetry, in which
317 the unit cell is constituted by one SlaB trimer and three SlaA dimers ($SlaB_3$/$3SlaA_2$).
318 Each SlaB trimer occupies alternating triangular pores and each SlaA dimer spans
319 two adjacent hexagonal pores. In agreement with our previously published model,
320 we observe that SlaB binds to alternating triangular pores. Because each SlaB
321 monomer interacts with the dimerisation domains of SlaA dimers, the SlaB trimer
322 occupancy of all triangular pores would likely be unfavourable due to steric
323 hindrance. Additionally, alternating SlaB throughout the array would reduce the
324 protein synthesis costs for this protein by 50%. SlaB trimers occupying every second
325 triangular pore also effectively create an S-layer with a variety of pore sizes,
326 modulating the exchange of molecules with the environment.
327 Using exosomes and a new image processing approach, we were able to improve
328 the resolution and eliminate the missing wedge in our subtomogram average of the
329 *S. acidocaldarius* S-layer. The new map enabled us to build a revised model of the
330 *S. acidocaldarius* S-layer assembly (Fig. 4 and 5). Here, the SlaA dimer (Figure 4 –
331 Figure Supplement 3a) spans an angle of 160° and extends over 42 nm, instead of
332 23 nm, as previously proposed[28]. The increased length is largely a result of the
333 unexpected positioning of domains D5 and D6, which were previously not accounted
334 for (Figure 4-Figure Supplement 3).

SLPs of extremophilic archaea generally show a high degree of glycosylation, potentially aiding their survival in extreme environments[48]. SlaA is predicted to contain 31 N-glycosylation sites[35] and the $SlaA_{30\text{-}1,069}$ cryoEM map showed 19 clear densities corresponding to N-glycosylation sequons. The cryoEM map contained densities for the complete hexasaccharide[35,36] on the SlaA surface, as well as various glycan intermediates. We cannot rule out the possibility that our cryoEM map could not resolve the complete hexasaccharide on all sequons due to the flexibility of the glycans. In any case, the presence of a heterogeneous family of glycans has previously been reported[35], with nano-LC-ES-MS/MS used to analyse the structure of the glycans linked to the C-terminal portion of SlaA (residues 961-1,395), and a heterogenous degree of glycosylation was observed including all intermediates from monosaccharide to complete hexasaccharide. The presence of a heterogeneous family of glycans has also been shown, for example, in the SLP of *H. volcanii*[49] and the archaellum of *Methanothermococcus thermolithotrophicus*[50]. In archaea, the final step in protein glycosylation is catalysed by the oligosaccharyl transferase AglB[51]. The enzyme is promiscuous, meaning that AglB can load glycans of variable length on the lipid carrier[52]. While AglB is essential for the viability of *S. acidocaldarius*[51], it remains to be determined whether the heterogenous composition of its glycans is to be attributed to AglB loading glycan precursors onto SlaA and/or glycan hydrolysis due to the harsh environmental conditions. A future study involving the genetic or enzymatic ablation of glycosylation sites would shed more light on the roles that surface glycans play in S-layer structure, stability, and function.

Metal ions are often bound to SLPs and have recently been demonstrated to play a crucial role in S-layer assembly and cell-surface binding[14,16,17,19,20,53,54]. In the bacterium *C. crescentus*, whose S-layer has been investigated in detail, Ca^{2+} ions

are essential for intra- and inter-molecular stability of the S-layer lattice[16,19]. Moreover, analogous results have been obtained for the S-layer of *Geobacillus stearothermophilus*[17]. The SLP of the archaeon *H. volcanii* has also been recently confirmed to bind cations[15]. The *S. acidocaldarius* S-layer is no exception and its assembly is a Ca^{2+}-dependent process[28]. Interestingly, the $SlaA_{30\text{-}1,069}$ cryoEM map did not reveal any anomalous densities that could be attributed to ions. It is therefore possible that cations are harboured in the D5 and D6 domains that were not resolved, and / or at the protein-protein interfaces within the assembled lattice, which at this point cannot be defined at the side-chain level due to the limited resolution of our sub-tomogram average.

In a recent work, von Kügelgen *et al*. presented the structure of the *H. volcanii* S-layer[15]. Therefore, the *H. volcanii* and *S. acidocaldarius* S-layers are currently the only two archaeal S-layers for which complete atomic models are available. *H. volcanii* is a halophilic archaeon of the Euryarchaeota phylum. As the *S. acidocaldarius* S-layer, the *H. volcanii* lattice also exhibits a hexagonal symmetry, but different architecture. The *H. volcanii* S-layer is constituted by a single glycosylated SLP named csg. SlaA (1,424 residues) and csg (827 residues) both consist of six domains (Figure 5-Figure Supplement 4b). However, while all csg domains adopt Ig-like folds, SlaA is built up from domains of more complex topology. In csg, the domains are arranged linearly, whereas SlaA adopts an extended Y-shape (Figure 5-Figure Supplement 4 a,b). Ig-like domains are widespread among SLPs in different archaeal phyla, including the order Sulfolobales[15]. In fact, the SlaA protein of *Metallosphaera sedula* is predicted to consist of seven Ig-like domains (Figure 5-Figure Supplement 4 d)[15]. The different domain architecture that we

observe for *S. acidocaldarius* SlaA highlights the great divergence of S-layers among microorganisms.

Assembled csg forms hexagonal (13 Å), pentameric (6 Å), and trimeric (10 Å) pores much smaller than the hexagonal (48 Å) and trimeric (85 Å) pores of the *S. acidocaldarius* lattice. In both cases, the pore size is further reduced by glycans projecting into the pores. The glycans could regulate the permeability of the S-layer in a fashion similar to the hydrogel regulating the permeability of the nuclear pore complexes[55]. It is currently unknown which evolutionary parameters resulted in species-specific S-layer pore sizes. It may be speculated that, for example, these pores have co-evolved with and adapted their size according to certain secreted protein filaments, such as pili. *S. acidocaldarius* produces four such filaments - archaella[56], A-pili[57], UV-inducible pili and threads[58]. Of these four filaments, only threads, with a diameter of ~40 Å, would be able to pass through the hexagonal pores of the S-layer without the need for a widening of the pores or a partial S-layer disassembly. It is thus tantalising to speculate that the hexagonal S-layer pores have evolved to accommodate threads, perhaps as a scaffold for their assembly.

S-layers are intrinsically flexible structures as to encapsulate the cell entirely. In the case of *H. volcanii*, csg assembles around hexameric as well as pentameric pores on the surface of both exosomes and whole cells[15]. Such pentameric “defects” confer enough flexibility to the array to encase the cell in areas of low and high membrane curvature. Interestingly, we did not observe an analogous phenomenon for the *S. acidocaldarius* S-layer on whole cells or exosomes. However, symmetry breaks have been observed on S-layers isolated from whole cells at the edges where the lattice changes orientation[59]. Furthermore, additional flexibility may be provided by the SlaA dimeric interface, as well as by loop regions linking the SlaA

domains. In fact, only single loops link D1-D2, D3-D4, D4-D5 and D5-D6. While the reciprocal position of D3-D4 is stabilised by the disulphide bond (Cys_{677}-$Cys_{1,017}$), the loops connecting D1-D2, D4-D5 and D5-D6 may allow the flexibility necessary for SlaA to be incorporated in this highly interwoven, yet deformable protein network.

Electrostatic interactions are critical for proper protein folding and function. Moreover, changes in surface charge have been shown to affect protein-protein interactions. Particularly, the pH plays a key role in determining the surface charge of proteins due to polar amino acid residues on the protein surface[40,41]. Remarkably, $SlaA_{30\text{-}1,069}$ proved stable over a vast pH range and its tertiary structure remains virtually unchanged. Thus, we propose that is likely not pH-induced unfolding or conformational changes in SlaA that cause S-layer disassembly at alkaline pH.

The surface net charge of SlaA shifts from positive to negative from pH 4 to pH 10. The observed reversal in electrostatic potential at rising pH values is a manifestation of deprotonation of amino acid residues, as the concentration of hydrogen ions (H^+) in the solution decreases. The loss of protons can reduce or abolish the ability of side chains to form hydrogen bonds, and as a result, hydrogen bonds involving these groups can be weakened or broken. The weakening or abolishment of these bonds (in particular those involving acidic amino acids) could therefore be a key factor in pH-induced disassembly. Conversely, the lowering of the pH will re-protonate these residues, facilitate the formation of hydrogen bonds, and thus the assembly of the S-layer. However, it is important to note that the effects of pH on hydrogen bonding in proteins can be complex. Thus, further experimentation would be required to test this hypothesis.

Considerations regarding the pH stability of $SlaA_{30\text{-}1,069}$ can be extended to the entirety of the protein using pH stability predictions, which suggest virtually no

difference in pH-dependent protein stability across ionic strength and pH values for both $SlaA_{30\text{-}1,069}$ and SlaA (Figure 5-Figure Supplement 5a-d). This suggests that domains D5 and D6 equally do not unfold at alkaline pH. Analogous predictions of protein stability were obtained for SlaB (Figure 5-Figure Supplement 5 e,f), where the net charge is slightly positive across pH 2-8. For comparison, we ran the same predictions on the *C. crescentus* and *H. volcanii* S-layer proteins RsaA and csg, respectively (Figure 5-Figure Supplement 6). Among SlaA, SlaB, RsaA and csg, we observe that SlaA and SlaB are expected to be the most stable at different pH values. Notably, csg is most stable at acidic pH and progressively less so at neutral and alkaline pH. This prediction is confirmed by experimental data[60], which additionally showed pH-dependent protein folding rearrangements and protein unfolding. It is to be considered that this prediction does not include glycosylation[61], which enhances S-layer stability, especially in the case of Sulfolobales[48,51,62,63]. The resilience of SlaA at temperature and pH shifts can likely be attributed to two main factors: the high glycosylation level, and the fact that ~ 56% of $SlaA_{30\text{-}1,069}$ has a defined secondary structure, which allows the formation of intramolecular bonds[64].

S-layers are often necessary for the survival of microorganisms in nature but can also be of great interest for synthetic biology. Therefore, a greater understanding of their structural details will strongly aid their nanotechnological uses, which have already shown remarkable potential in biomedical[18,65,66] and environmental applications[67–70].

Methods

S. acidocaldarius strains and growth conditions

Cells of *S. acidocaldarius* strain MW001 were grown in basal Brock medium* at pH 3[24] as previously described[28]. Briefly, cells were grown at 75 °C, 150 rpm, until an OD600 of >0.6 was reached. Cells were then centrifuged at 5,000 g (Sorvall ST 8R) for 30 min at 4 °C. The cell fraction was stored at -20 °C for S-layer isolation, whereas the supernatant was stored at 4 °C for exosomes isolation.

*Brock media contain (per l): 1.3 g $(NH_4)2SO_4$, 0.28 g KH_2PO_4, 0.25 g $MgSO_4 \cdot 7H_2O$, 0.07 g $CaCl_2 \cdot 2H_2O$, 0.02 g $FeCl_2 \cdot 4H_2O$, 1.8 mg $MnCl_2 \cdot 4H_2O$, 4.5 mg $Na_2B_4O_7 \cdot 10H_2O$, 0.22 mg $ZnSO_4 \cdot 7H_2O$, 0.05 mg $CuCl_2 \times 2H_2O$, 0.03 mg $NaMoO_4 \cdot 2H_2O$, 0.03 mg $VOSO_4 \times 2H_2O$, 0.01 mg $CoSO_4 \cdot 7H_2O$, and 0.01 mg uracil.

S-layer isolation and disassembly

The S-layer isolation and disassembly were performed as previously described[28]. Briefly, frozen cell pellets from a 50 ml culture were incubated at 40 rpm (Stuart SB3) for 45 min at 37 °C in 40 ml of buffer A (10 mM NaCl, 1 mM phenylmethylsulfonyl fluoride, 0.5% sodium lauroylsarcosine), with 10 µg/ml DNase I. The samples were pelleted by centrifugation at 18,000 × g (Sorvall Legend XTR) for 30 min and resuspended in 1.5 ml of buffer A, before further incubation at 37 °C, for 30 min. After centrifugation at 14,000 rpm for 30 min (Sorvall ST 8R), the pellet was purified by resuspension and incubation in 1.5 ml of buffer B (10 mM NaCl, 0.5 mM $MgSO_4$, 0.5% SDS) and incubated for 15 min at 37 °C. To remove SlaB from the assembled S-layers, washing with buffer B was repeated three more times. Purified Sla-only S-layers were washed once with distilled water and stored at 4 °C. The removal of SlaB was confirmed by SDS/PAGE analysis. S-layers were disassembled by increasing the pH to 10 with the addition of 20 mM $NaCO_3$ and 10 mM $CaCl_2$ and incubated for 2 hours at 60 °C, 600 rpm (Thermomixer F1.5, Eppendorf).

SlaA purification

After disassembly the sample containing SlaA was further purified using gel filtration chromatography. A total of 100 µl containing 10 mg/ml of disassembled protein were loaded onto a Superdex 75 Increase 10/300 GL (GE Healthcare) using 300 mM NaCl for elution. At the end of the run, the fractions containing SlaA were dialysed against 30 mM acetate buffer (0.1 M CH_3COOH, 0.1 M CH_3COONa) at pH 4, 150 mM Tris-HCl at pH 7, or 20 mM $NaCO_3$ at pH 10, with the aim to compare the SlaA protein structure at different pH values. The purity of the fractions was assessed by SDS/PAGE analysis and negative staining with 1% uranyl acetate on 300 mesh Quantifoil copper grids with continuous carbon film (EM Resolutions).

CryoEM workflow for single particle analysis (SPA)

Grid preparation

The purified SlaA samples at pH 4 and 10 (3 µl of ~0.1 mg/ml) were applied to 300 mesh copper grids with graphene oxide-coated lacey carbon (EM Resolutions) without glow discharge. Grids were frozen in liquid ethane using a Mark IV Vitrobot (Thermo Fisher Scientific, 4 °C, 100 % relative humidity, blot force 6, blot time 1 sec) with Whatman 597 filter paper. The purified SlaA at pH 7 was applied to glow discharged R 1.2/1.3 300 mesh copper grids with holey carbon. The freezing procedure was kept the same as for the samples at pH 4 and 10 besides the blot time of 2 sec.

Data collection

Micrographs were collected on a 200 kV FEI Talos Arctica TEM, equipped with a Gatan K2 Summit direct detector using EPU software (Thermo Fisher Scientific) (Supplementary File 1a). Data were collected in super-resolution at a nominal magnification of 130,000x with a virtual pixel size of 0.525 Å at a total dose of ~60 e^- /Å^2. A total of 3,687 movies (44 fractions each), 3,163 movies (44 fractions each), and 5,046 movies (60 fractions each), with a defocus range comprised between -0.8 and -2.4 µm, were collected for samples at pH 4, pH 7 and pH 10, respectively.

Image processing

Initial steps of motion correction (MotionCor 2[71]) and CTF estimation (CTF-find 4[72]) were performed in Relion 3.0[39] and Relion 3.1[30] for datasets at pH 4 and 7, whereas Warp[43] was used for the pH 10 dataset. Further steps of 2D and 3D classification, refinement, CTF refinement and polishing were performed using Relion 3.1. For a detailed workflow of the three datasets see Supplementary Fig. 1, 8a,b and 9. The refined maps were post-processed in Relion 3.1 as well as using DeepEMhancer[73]. The produced maps had a resolution of 3.1 Å, 3.9 Å and 3.2 Å at pH 4, 7 and 10, respectively, by gold-standard FSC 0.143.

Model building and validation

The SlaA atomic model was built *de novo* using the cryoEM map at pH 4 in Buccaneer[74], refined using REFMAC5[75] and rebuilt in COOT[76]. The glycans were modelled in COOT with the refinement dictionary for the unusual sugar 6-sulfoquinovose prepared using JLigand[77]. This atomic model was then positioned into the cryoEM maps at pH 10 and pH 7 using ChimeraX[78] and refined using

REFMAC5 and COOT. All models were further refined using Isolde[79] and validated using Molprobity[80] in CCP4[81].

Exosome isolation

S. acidocaldarius exosomes were isolated from the supernatant obtained after cell growth. The procedure was adapted from Ellen *et al*., 2009[42]. The supernatant was split into 8 fractions and exosomes were pelleted in two runs of ultracentrifugation (Optima LE-80K, Beckman Coulter) at 125,000 g for 45 min at 4 °C. The pellet was resuspended in 2 ml (per fraction) of the supernatant and ultracentrifuged (Optima MAX-TL, Beckman Coulter) at 12,000 rpm (TLA55 rotor, Beckman Coulter) for 10 min at 4 °C. The pellet (containing intact cells and cell debris) was discarded, and the supernatant was ultracentrifuged (Optima MAX-TL, Beckman Coulter) at 42,000 rpm (TLA55 rotor, Beckman Coulter) for 90 min at 4 °C. The pellet containing the isolated exosomes was resuspended in MilliQ water at a concentration of 15 mg/ml. The purity of the sample was assessed by negative staining with 1% uranyl acetate on 300 mesh Quantifoil copper grids with continuous carbon film (EM Resolutions).

CryoEM workflow for subtomogram averaging

Grid preparation

The isolated exosomes were mixed 1:1 with 10 nm colloidal gold conjugated protein A (BosterBio) and 3 µl droplets were applied four times on glow discharged 300 mesh Quantifoil copper R2/2 grids (EM Resolutions). The grids were blotted with 597 Whatman filter paper for 4 sec, using blot force 1, in 95 % relative humidity, at 21 °C, and plunge-frozen in liquid ethane using a Mark IV Vitrobot (FEI).

Data collection

Micrographs were collected on two microscopes: a 200 kV FEI Talos Arctica TEM, equipped with a Gatan K2 Summit direct detector and a 300 kV Thermo Fisher Titan Krios G3 with a Thermo Fisher Falcon 4i direct detector and SelectrisX energy filter, both using the Tomo 4 package. Tilt series on the Talos/K2 were collected in super-resolution at a nominal magnification of 63,000 x with a virtual pixel size of 1.105 Å at a total dose of ~83 e-/Å^2. The tilts were collected from –20 deg to 60 deg, in 3 degree steps (2 fractions per tilt). Tilt series on the Krios/Falcon 4 were collected as conventional MRC files at 4k x 4k, nominal magnification of 64,000 x and a pixel size of 1.9 A at a total dose of ~ 83 e-/A2. Tilts were collected from -60 deg to 60 deg in 3 degree steps in a dose-symmetric scheme with groupings of 2 (6 fractions per tilt). A nominal defocus range between -4 and -6 µm was used for both collections. A total of 86 positions were collected, 28 on the Talos and 58 on the Krios.

Electron cryotomography and subtomogram averaging

Initial subtomogram averaging was performed using only data collected on the Talos. Motion correction was performed using the IMOD[82] program alignframes. IMOD was also used for the tomogram reconstruction. Initial particle picking on all 28 tomograms was performed using seedSpikes and spikeInit as part of the PEET software package[83] with a total of 12,010 particles picked. For initial subtomogram averaging, the picked particles were CTF corrected and extracted using the Relion STA pipeline[84]. 2D classification, initial model generation, 3D classification and initial refinements were all performed using Relion 3.1[30]. A resolution of 16.1 Å was reached using 1,313 particles and C3 symmetry.

For higher resolution averaging, the tilt series from both datasets were processed using the Warp-Relion-M pipeline[44]. Motion correction and CTF estimation of the movies were performed in Warp[43]. The poor quality tilts were excluded and Aretomo[85] was used to provide alignments on the resulting tilt series stacks for tomogram reconstruction in Warp. Deconvolved tomograms were used to visualise the exosomes and, as above, seedSpikes and spikeInit were used to generate initial particle coordinates for the S-layer. A total of 22,950 particles were picked and subsequently extracted in Warp at a pixel size of 10 Å/px. The two datasets were processed separately with several rounds of refinement and classification until they reached a resolution of 20 Å with C3 symmetry. For both datasets, the 16.1 Å map from the initial averaging was used, low-pass filtered to 60 Å. The two maps were visually compared and found to be different sizes, so the pixel size of the Talos data was adjusted. The tomograms were reprocessed and particles re-extracted at 10 Å/px then refined until a resolution of 20 Å was again achieved. The particles were combined together then refined in M to a resolution of 16 Å (C3 symmetry). The particles were extracted at a pixel size of 5 Å/px. Further refinement and 3D classification resulted in a 14 Å resolution. A final iteration in M resulted in a resolution of 11.2 Å with 2,771 particles used in the refinement.

The model of the assembled S-layer was built by initial rigid body fitting the SlaA structure determined by single particle analysis into the sub-tomogram average using ChimeraX[78]. The C-terminal domains of SlaA that were predicted in Alphafold2[34] were then added to each SlaA. Hereby, only SlaA in the extended conformation could be reconciled with the map. Next, the SlaB trimers were predicted in Alphafold2 and fitted into the trimeric stalks that connected the S-layer

canopy with the membrane. Finally, the model was refined using Namdinator[47], a molecular dynamics–based flexible fitting software.

Structure analysis and presentation

The electrostatic potential of the protein was derived using APBS (Adaptive Poisson-Boltzmann Solver)[86] based on the PARSE force field for the protein as available through PDB2PQR[87]. Where available, the charges of the glycans were assigned based on the GLYCAM force field[88]; charges of the hydrogens were combined with their central heavy atom. The charge assignment depends on the bonding topology, i.e. occupied linkage positions. Supplementary File 1b summarizes the mapping of residue from the structure file to GLYCAM residue names. For residue SMA, charge assignments are not available from the GLYCAM force field; these were derived based on RESP calculations conducted for the methoxy-derivatives on the HF/6-31G*//HF/6-31G* level of theory and employing a hyperbolic restraint equal to 0.010 in the charge fitting step[89,90]. The total charge of the newly derived residue was constrained to -0.8060 e and -1 e for the 1-substituted and 1,4-substituted SMA (referred to as SG0 and SG4 in Supplementary File 1c & 1d), respectively, in agreement with the conventions of the GLYCAM force field. In assembling the final charge assignment, the charge of the linking ND2 atom of the glycosylated Asn residues of the protein were altered to compensate for the polarization charge of the attached saccharide unit. The electrostatic charge was visualised using VMD[91] (http://www.ks.uiuc.edu/Research/vmd/).

The structure of *S. acidocaldarius* SlaA was visualised using UCSF Chimera[92], Chimera X v.1.3 and v1.4[78], and Pymol[93]. The structural domains of SlaA were assigned using SWORD[31].

Heatmaps for net charge, and pH and ionic strength-dependent protein stability were obtained using Protein-Sol (https://protein-sol.manchester.ac.uk/)[61]. For SlaB (Supplementary Fig. 17) the signal-peptide was predicted using InterPro[94], the trans-membrane region was predicted using TMHMM – 2.0[95], the N-glycosylation sites (sequons N-X-S/T) were predicted using GlycoPP v1.0[96].

Molecular dynamics simulations (MDS)

Conformation arrays of glycans were grafted on protein structure using GlycoSHIELD[38]. In brief, glycan systems (GlcNAc[2],Man[2],QuiS[1],Glc[1] N-linked to neutralised glyc-Asp-gly tripeptides) were modelled in CHARMM-GUI[97] and solvated using TIP3P water models in the presence of 150mM NaCl and configured for simulations with CHARMM36m force fields[98,99]. MDS were performed with GROMACS 2020.2 and 2020.4-cuda[100] in mixed GPU/CPU environments. Potential energy was first minimized (steepest descent algorithm, 5,000 steps) and were equilibrated in NVT ensemble (with 1 fs time-steps using Nose-Hoover thermostat). Atom positions and dihedral angles were restrained during the equilibration, with initial force constants of 400, 40 and 4 kJ/mol/nm^2 for restraints on backbone positions, side chain positions and dihedral angles, respectively. The force constants were gradually reduced to 0. Systems were additionally equilibrated in NPT ensemble (Parrinello-Rahman pressure coupling with the time constant of 5 ps and compressibility of 4.5 10^{-5} bar^{-1}) over the course of 10 ns with a time step of 2 fs. Hydrogen bonds were restrained using LINCS algorithm. During the production runs, a velocity-rescale thermostat was used and the temperature was kept at 351K. Production runs were performed for a total duration of 3μs and snapshots of atom positions stored at 100 ps intervals.

Glycan conformers were grafted using GlycoSHIELD with a distance of 3.25 Å between protein α-carbons and glycan ring-oxygens. Glycan conformers were shuffled and subsampled for representation of plausible conformations on displayed renders. Graphics were generated with ChimeraX[78].

Data availability

The SlaA atomic coordinates were deposited in the Protein Data Bank (https://www.rcsb.org/) with accession numbers 7ZCX, 8AN3, and 8AN2 for pH 4, 7 and 10, respectively. The cryoEM maps were deposited in the EM DataResource (https://www.emdataresource.org/) with accession numbers EMD-14635, EMD-15531 and EMD-15530 for pH 4, 7 and 10, respectively, and EMD-18127 for the *in situ* S-layer map obtained by cryoET.

Other structural data used in this study are: *H. volcanii* csg (PDB ID: 7PTR, http://dx.doi.org/10.2210/pdb7ptr/pdb), and *C. crescentus* RsaA ((N-terminus PDB ID: 6T72, http://dx.doi.org/10.2210/pdb6t72/pdb, C-terminus PDB ID: 5N8P, http://dx.doi.org/10.2210/pdb5n8p/pdb).

Acknowledgements

We acknowledge Ufuk Borucu for help with data collection, and the GW4 Facility for High-Resolution Electron Cryo-Microscopy, funded by the Wellcome Trust (202904/Z/16/Z and 206181/Z/17/Z) and BBSRC (BB/R000484/1). We also acknowledge Alexander Neuhaus for assistance with data analysis. We thank IDRIS

for the allocation of high-performance computing resources (allocations #2020-AP010711998 and #2021-A0100712343 to CH).

For this project, L.G., B.D., M.M, MG, RC and KS were funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 803894). M.M. was also funded by a BBSRC New Investigator Research Grant (BB/R008639/1) to VG and RC by the University of Exeter and a Wellcome Trust Seed Award in Science (210363/Z/18/Z) awarded to VG, as well as a Wellcome Trust Seed Award in Science (212439/Z/18/Z) awarded to BD. MS was supported by a Fonds zur Förderung der wissenschaftlichen Forschung Schrödinger fellowship (J4332-B28). Work in the laboratory of CH is supported by the Agence Nationale de la Recherche (grants #ANR-16-CE16-0009-01

Figures

Figure 1

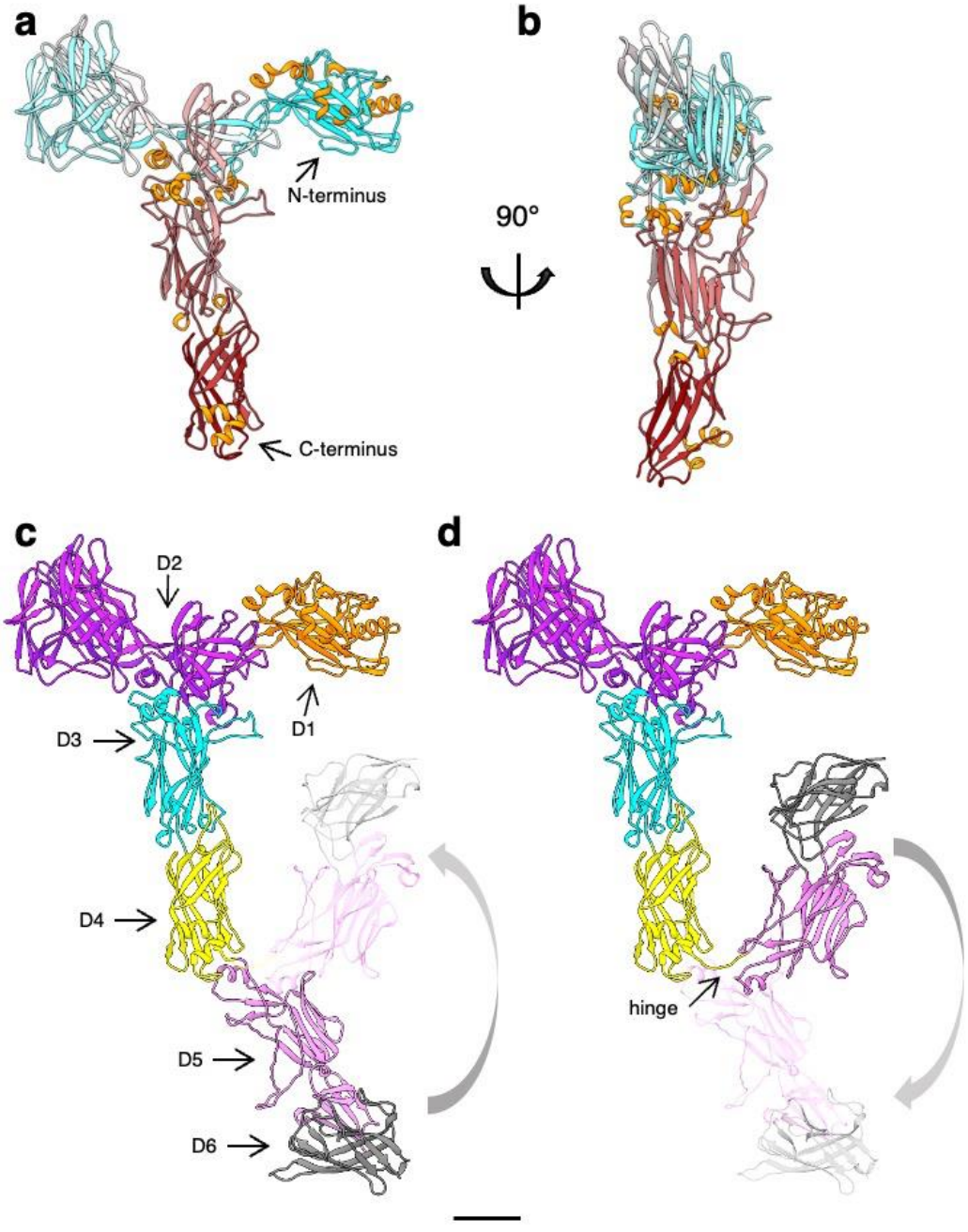


Figure 1. Atomic model of *S. acidocaldarius* S-layer protein SlaA at pH 4. a and b, $SlaA_{30\text{-}1,069}$ atomic model obtained by single particle cryoEM in ribbon representation and cyan-grey-maroon colours (N-terminus, cyan; C-terminus, maroon) with α-helices highlighted in orange. **c** and **d**, SlaA atomic models highlighting six domains: $D1_{30\text{-}234}$ (orange), $D2_{235\text{-}660,701\text{-}746}$ (purple), $D3_{661\text{-}700,747\text{-}914}$ (cyan), $D4_{915\text{-}1,074}$ (yellow), $D5_{1,075\text{-}1,273}$ (pink), and $D6_{1,274\text{-}1,424}$ (grey). D5 and D6 were predicted using Alphafold. A flexible hinge exists between D4 and D5. D5 and D6 are thus free to move relative to D1-

D4 in the isolated SlaA particle (highlighted by an arrow between a stretched (c) and a flapped (d) conformation). Scale bar, 20 Å.

Figure 2

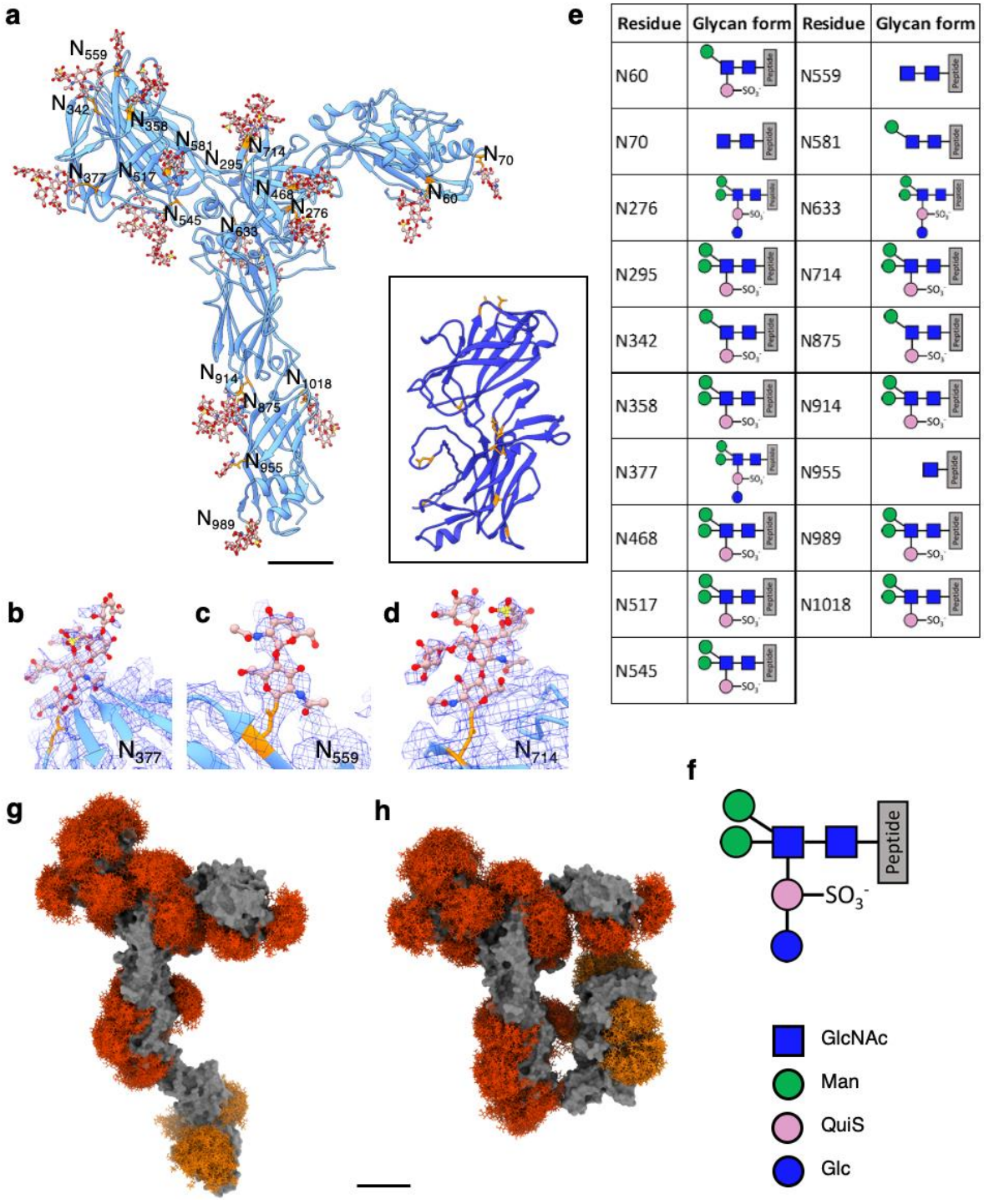


Figure 2. N-glycosylation of *S. acidocaldarius* SlaA. **a**, SlaA atomic model in ribbon representation. $SlaA_{30\text{-}1,069}$ as solved by cryoEM is in cornflower blue; $SlaA_{1,070\text{-}1,424}$ as predicted by

Alphafold is in blue (boxed). All 19 Asn-bound glycan molecules (stick representation; glycans in rusty brown, Asn in orange) in $SlaA_{30\text{-}1,069}$ are modelled fitting the cryoEM map. In the glycans, O atoms are shown in red, N in blue and S in yellow. Eight glycosylated Asn are highlighted (stick representation, orange) in $SlaA_{1,070\text{-}1,424}$ based on Peyfoon *et al*., 2010[35]. Scale bar, 20 Å. **b**-**d** are example close-ups of glycosylation sites with superimposed cryoEM map (blue mesh). (**b**) shows the full hexasaccharide on Asn_{377}, (**c**) shows $GlcNAc_2$ on Asn_{559}, and (**d**) shows a pentasaccharide lacking Glc_1 on Asn_{714}. **e**, list of glycosylation sites and associated glycans of $SlaA_{30\text{-}1,069}$. The schematic glycan representation (**f**) is equivalent to Peyfoon *et al*., 2010[35]. Blue square, N-acetylglucosamine; green circle, mannose; pink circle, 6-sulfoquinovose; blue circle, glucose. **g** and **h**, GlycoSHIELD models (red, orange) showing the glycan coverage of the protein (solid grey). Glycan shields corresponding to glycosylation sites visualised by cryoEM are highlighted in red, glycan shields with the Alphafold model of the SlaA C-terminus are shown in orange.

Figure 3

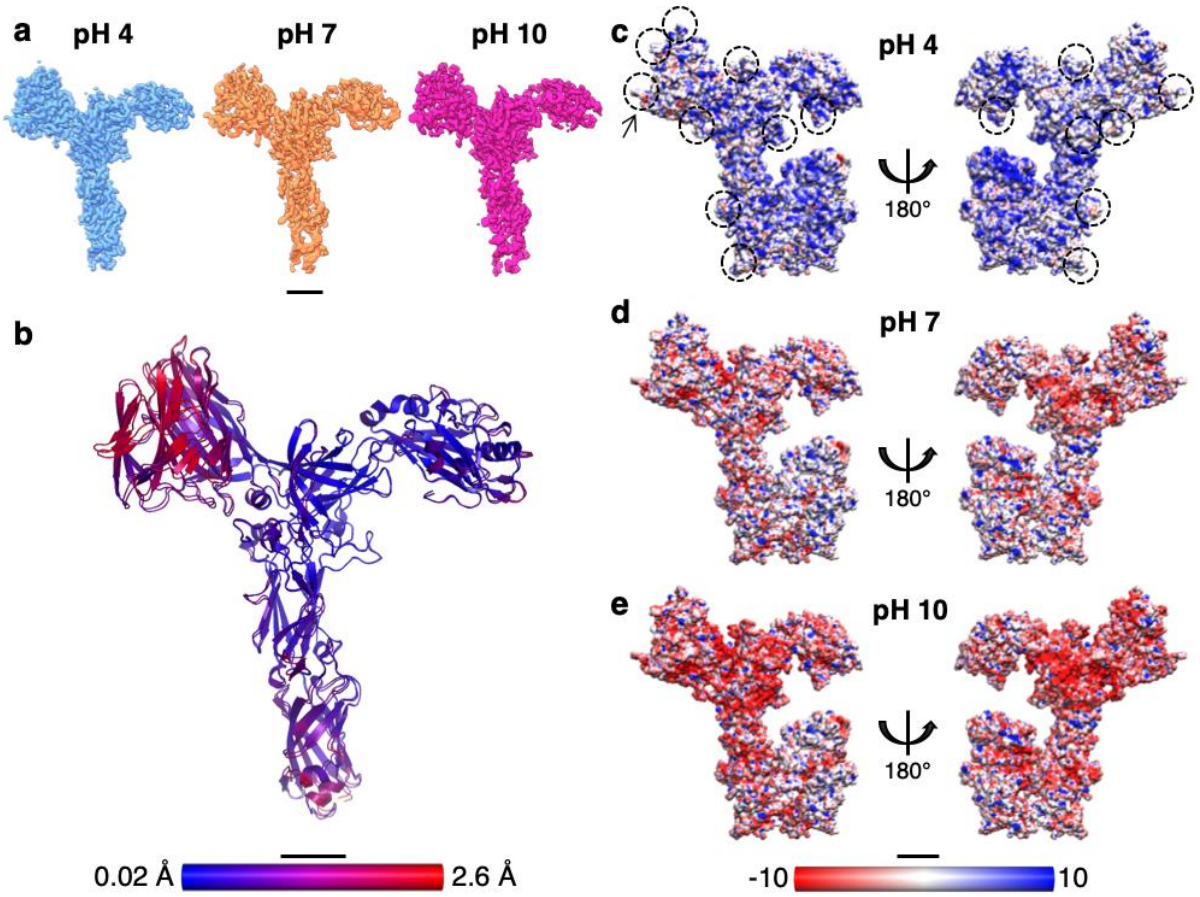


Figure 3. Structural comparison and electrostatic surface potentials of *S. acidocaldarius* SlaA at different pH conditions. a, $SlaA_{30\text{-}1,069}$ cryoEM maps at pH 4 (light blue, res. 3.1 Å), pH 7 (orange, res. 3.9 Å) and pH 10 (magenta, res. 3.2 Å). **b**, r.m.s.d. (root-mean-square deviation) alignment between $SlaA_{30\text{-}1069}$ atomic models at pH 4 and pH 10. Smaller deviations are shown in blue and larger deviations in red, with mean r.m.s.d. = 0.79 Å. **c-e**, electrostatic surface potentials of SlaA at pH 4 (**c**), pH 7 (**d**) and pH 10 (**e**). Models include Alphafold-predicted C-terminal domains (in closed conformation). Surfaces are coloured in red and blue for negatively and positively charged residues respectively. White areas represent neutral residues. In (**c**) some areas occupied by glycans are circled; the arrow points at one of the 6-sulfoquinovose residues displaying a negative charge at pH 4. Scale bar, 20 Å.

Figure 4

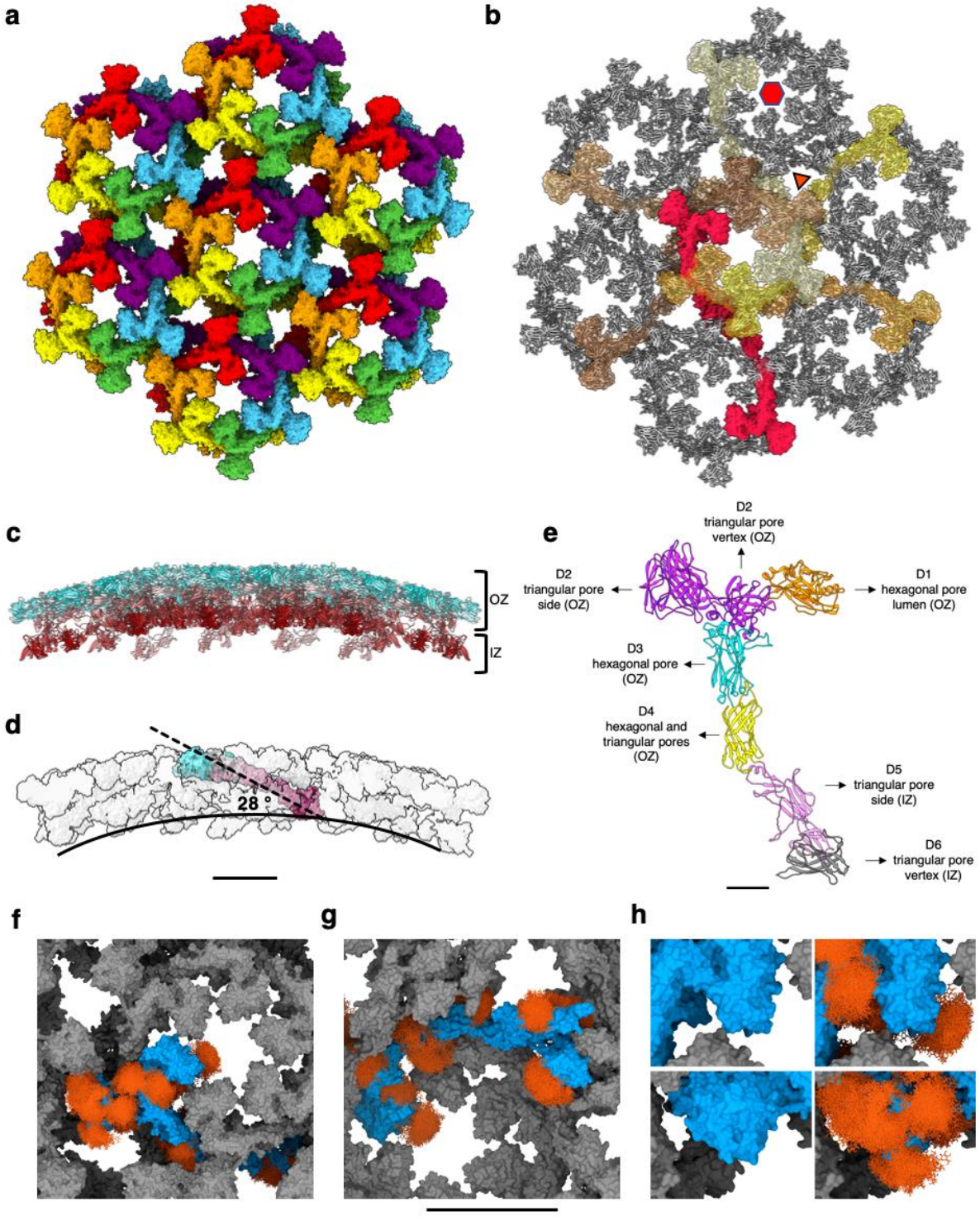


Figure 4. *S. acidocaldarius* SlaA assembly on the surface of exosomes. a, extracellular view of assembled SlaA monomers in surface representation and in randomly assigned colours. **b**, extracellular view of assembled SlaA in ribbon representation with SlaA dimers around a hexagonal pore highlighted in shades of red and surface representation. Each dimer spans two adjacent hexagonal pores. **c**, side view of the SlaA lattice (blue, N-terminus; red, C-terminus). It is possible to distinguish an outer zone (OZ) formed by domain D1, D2, D3 and D4, and an inner zone (IZ)

constituted by domains D5 and D6. **d**, one SlaA monomer (surface representation, N-terminus cyan, grey, C-terminus maroon) is highlighted within assembled SlaA in surface representation. SlaA assembles with its long axis (dashed line) forming a 28° angle to the membrane plane (solid line). **e**, location of SlaA domains in assembled S-layer. The SlaA domains are highlighted in different colours: $D1_{30\text{-}234}$ in orange, $D2_{235\text{-}660,701\text{-}746}$ in purple, $D3_{661\text{-}700,747\text{-}914}$ in cyan, $D4_{915\text{-}1,074}$ in yellow, $D5_{1,075\text{-}1,273}$ in pink, and $D6_{1,274\text{-}1,424}$ in grey. **f-h**, SlaA glycans modelled with GlycoSHIELD in the assembled S-layer. (**f**) shows the extracellular view; (**g**) shows the intracellular view; (**h**) shows insets of (**f**) at higher magnification without (left) and with (right) glycans. Glycans fill gaps unoccupied by the protein and significantly protrude in the lumen of the triangular and hexagonal pores. Scale bars in (**a-d** and **f-h**), 10 nm; in (**e**), 20 Å.

Figure 5

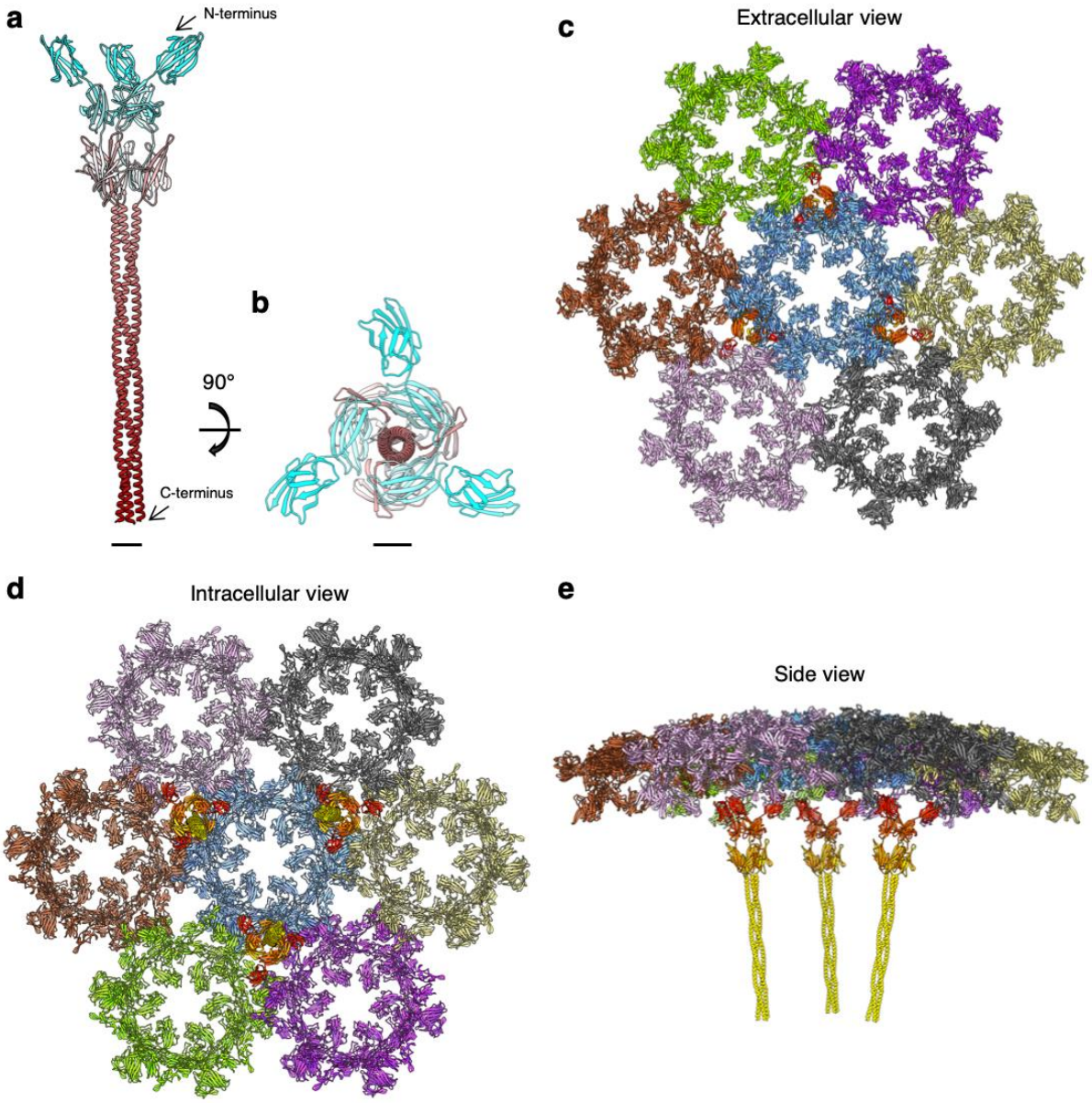


Figure 5. *S. acidocaldarius* S-layer assembly. a and **b**, SlaB trimer (ribbon representation, N terminus, cyan; C-terminus, maroon) as predicted by Alphafold v2.2.0[34]. **c-e**, ribbon representation of the assembled SlaA and SlaB components of the S-layer. (**c**), (**d**), and (**e**) are extracellular, intracellular and side views, respectively. SlaA proteins around each hexagonal pore are shown in different colours. SlaB trimers are shown in orange tones (N-termini are in darker shades and C-termini in lighter shades). Scale bar, (**a**) and (**b**), 20 Å; (**c-e**) 10 nm.

Supplementary Figures

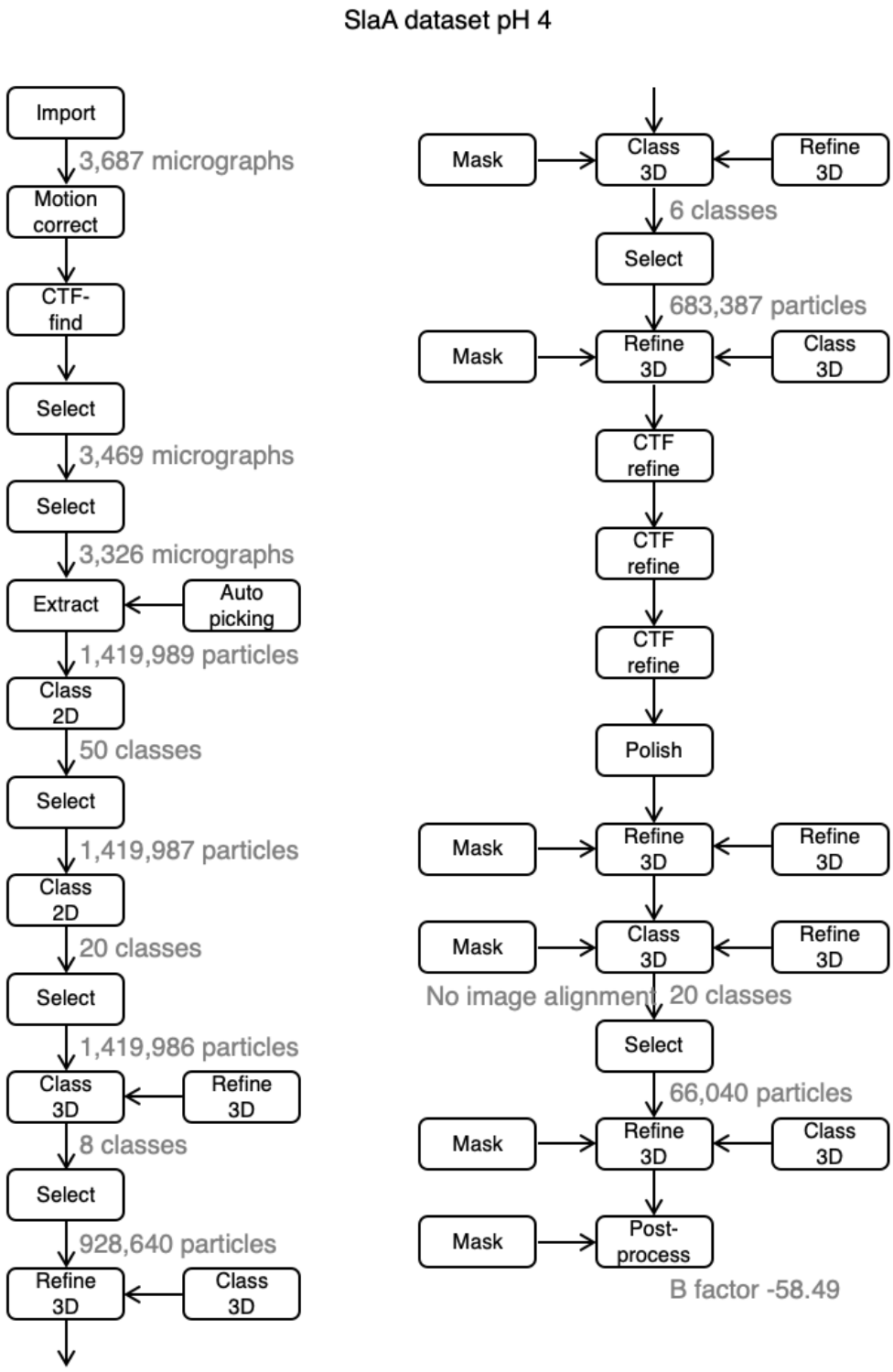


Fig. 1-Figure Supplement 1. Relion processing workflow for pH 4 dataset.

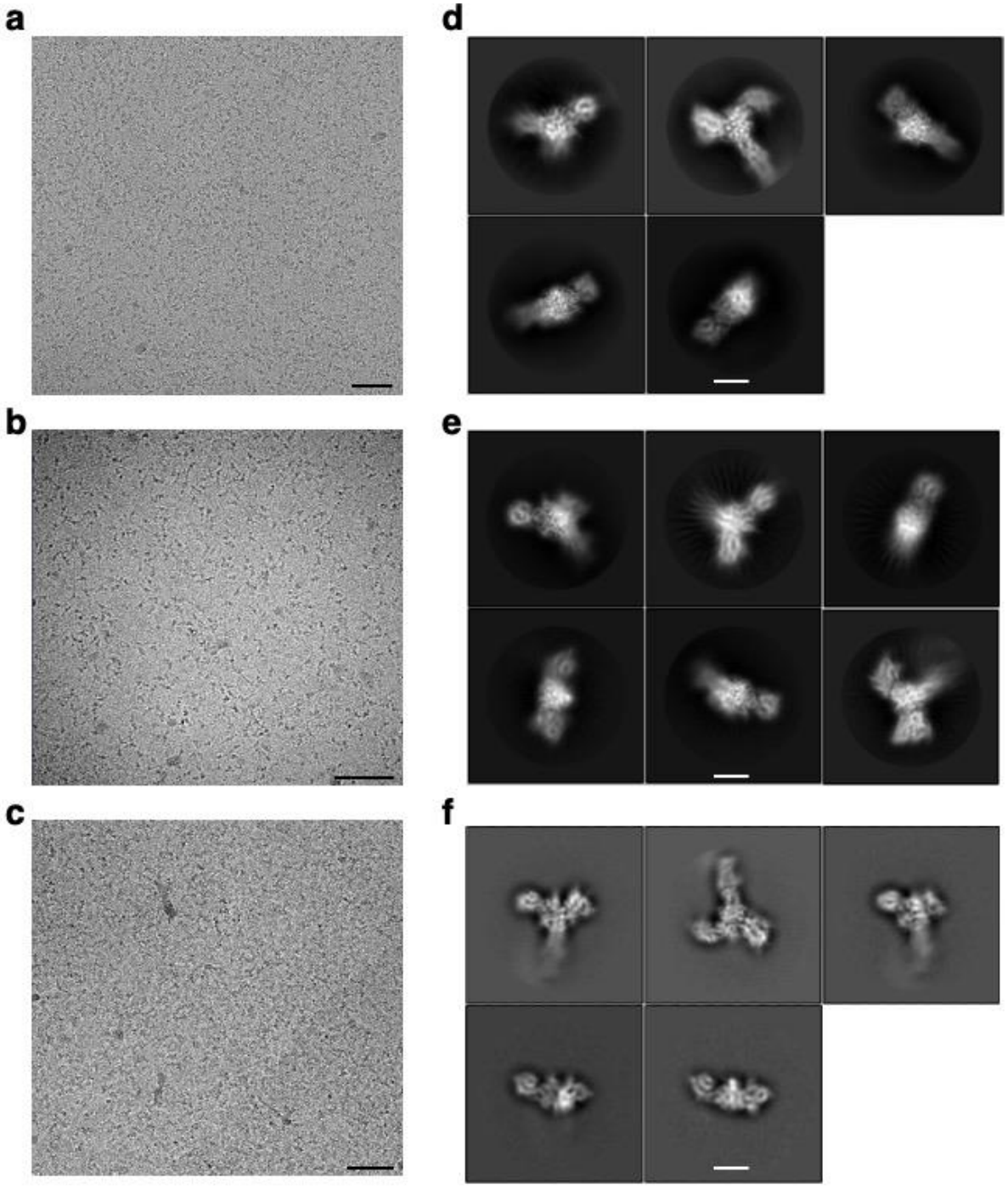


Fig. 1-Figure Supplement 2. SlaA representative data. a-c, representative cryoEM micrographs (from a total of 3,687 for (**a**), 3,163 for (**b**) and 5,046 for (**c**)). **d**-**f**, 2D classification examples of *S. acidocaldarius* SlaA polished particles in Relion at pH 4 (**a, d**), pH 7 (**b, e**) and pH 10 (**c, f**). Scale bar (**a-c**),100 nm; (**d-f**), 50 Å.

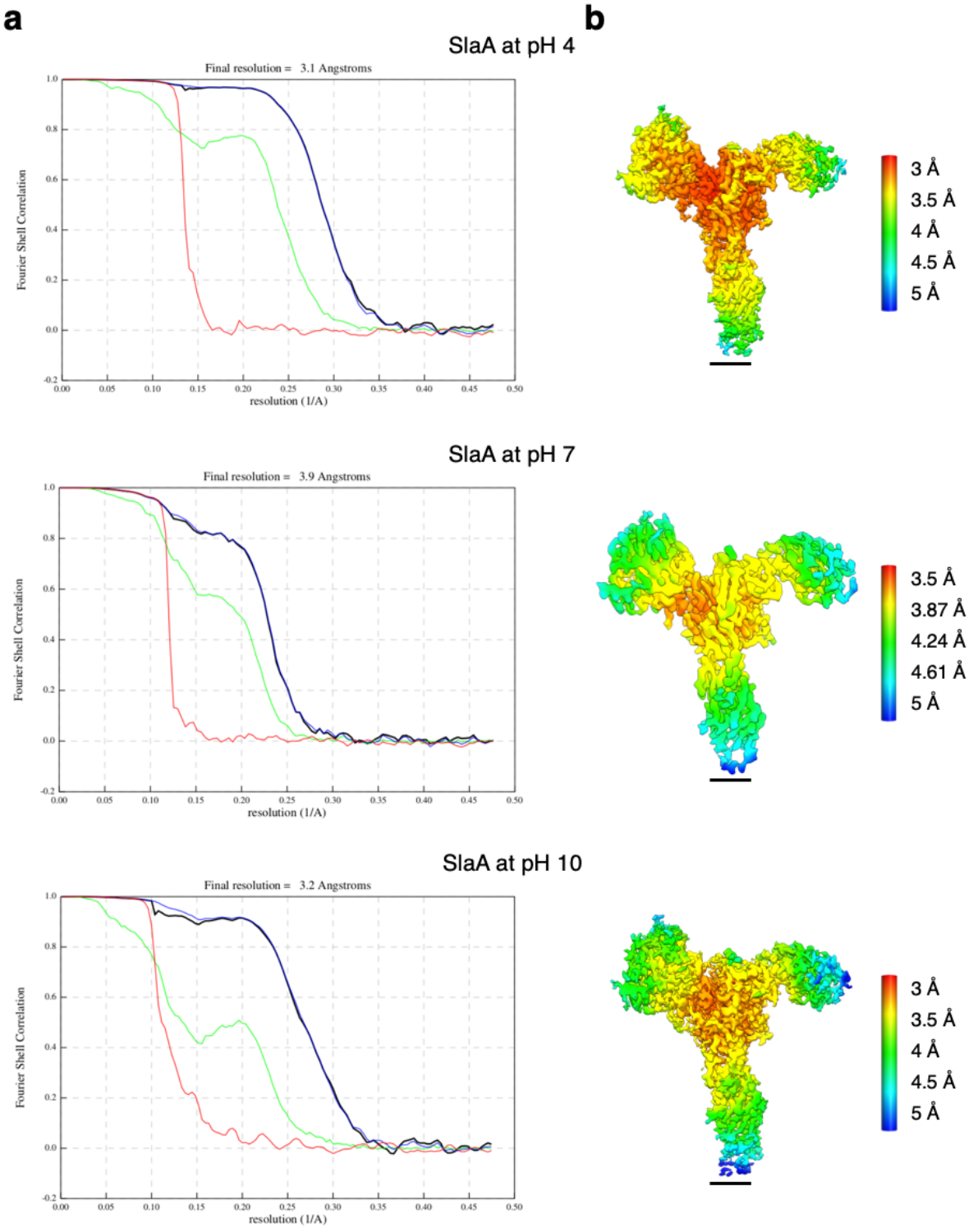


Fig. 1-Figure Supplement 3. SlaA data quality in Relion. **a**, gold standard FSC and **b**, local resolution estimations for the SlaA map obtained at pH 4, 7 and 10. Red, phase randomised masked; green, unmasked; blue, masked; black, corrected. Scale bar, 20 Å.

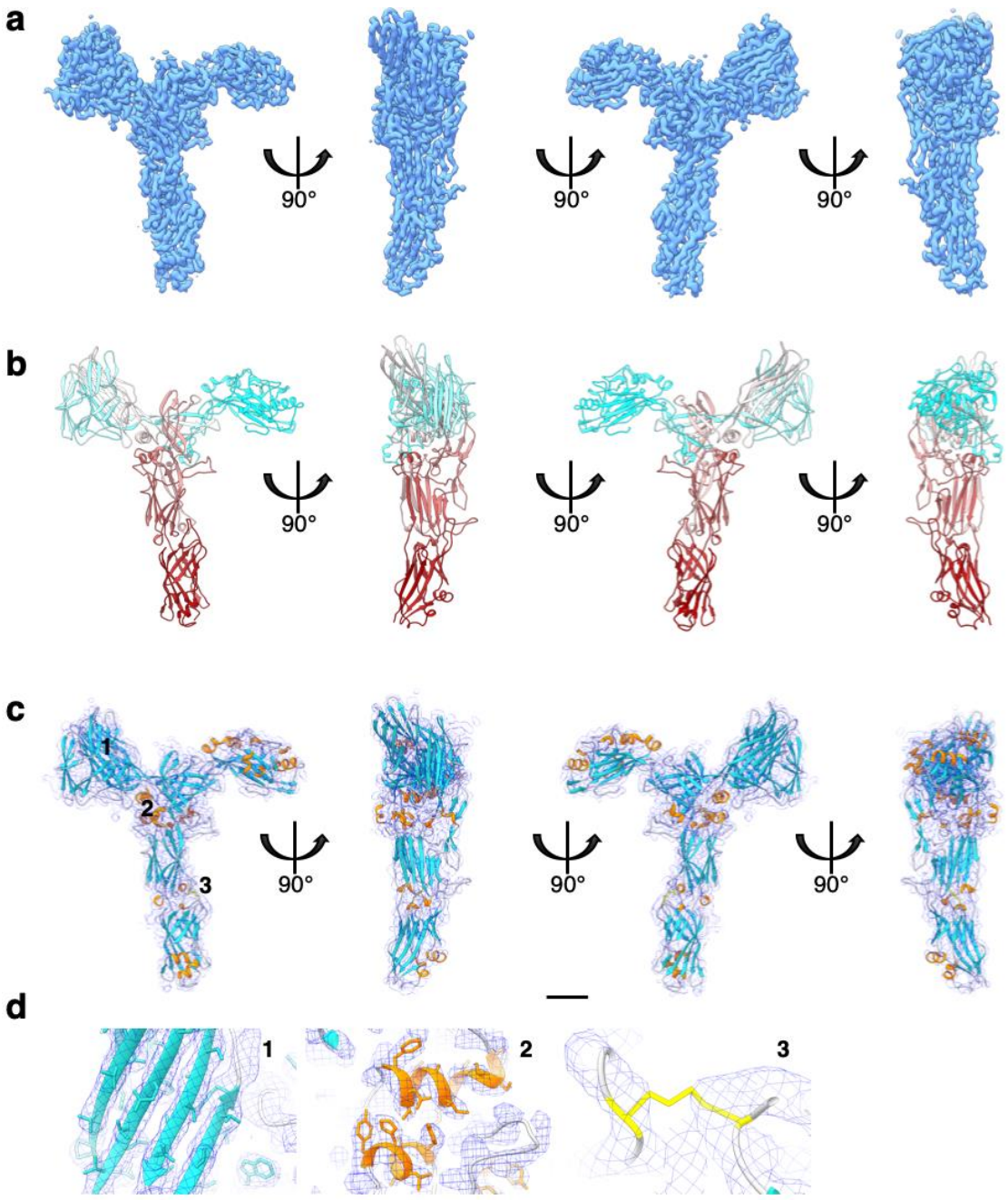


Fig. 1-Figure Supplement 4. $SlaA_{30-1069}$ cryoEM map and atomic model. a, $SlaA_{30-1,069}$ cryoEM map at 3.1 Å global resolution. **b**, atomic model of $SlaA_{30-1,069}$ (ribbon representation, cyan-grey-maroon colours. N-terminus, cyan; C-terminus, maroon). **c**, fitting of the atomic model (ribbon representation) into the cryoEM map (blue mesh). The loop regions are in grey, α-helices in orange, β-sheets in turquoise, and the disulphide bridge in yellow. **d**, close-ups of three example regions of β-sheets, α-helices and the disulphide bridge. Locations of the close-ups are labelled in (**d**) as 1, 2 and 3. Scale bar in (**a-c**), 20 Å.

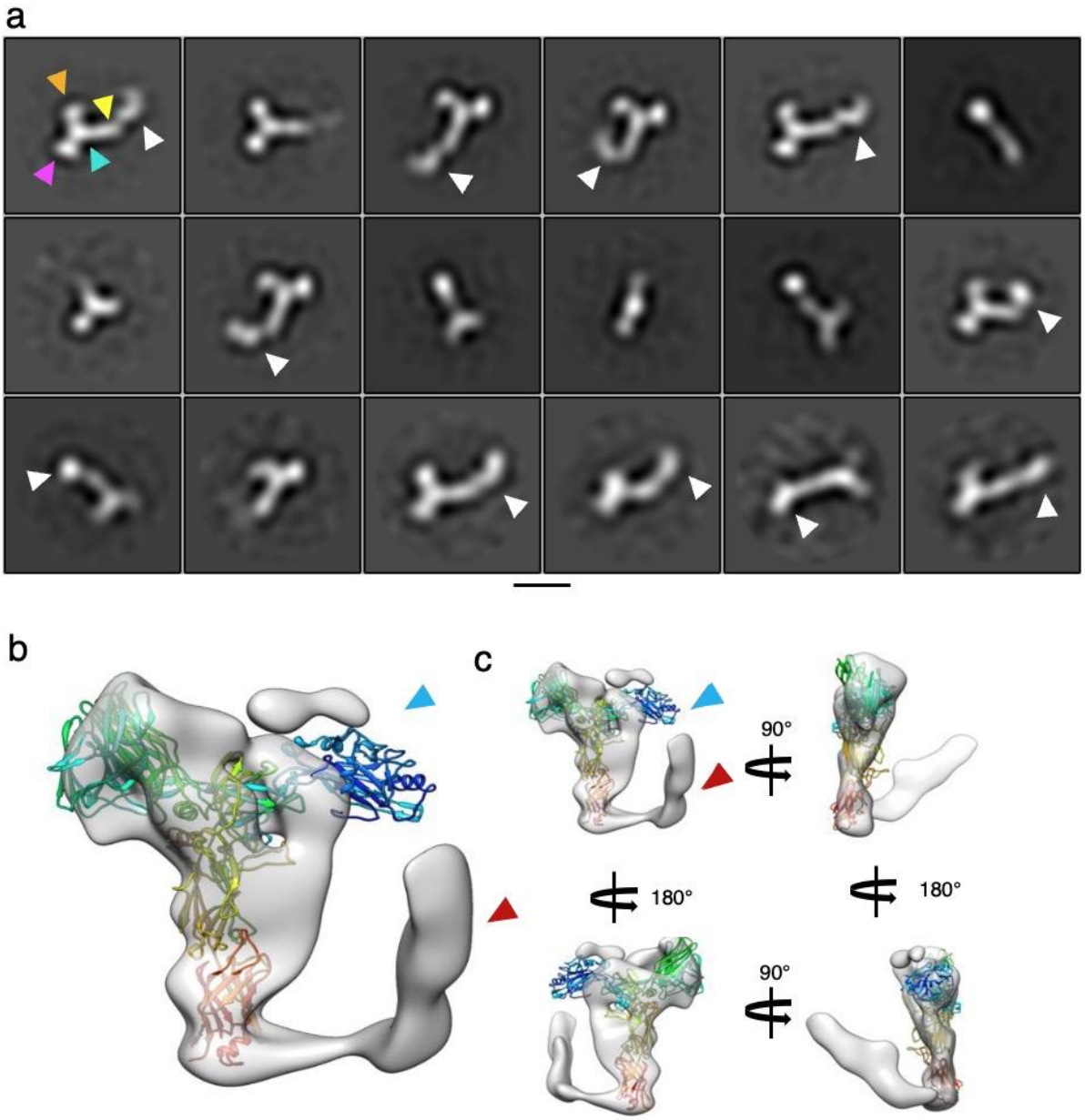


Fig. 1-Figure Supplement 5. SlaA flexibility. a, 2D classification of negative stained micrographs of SlaA purified from *S. acidocldarius*. The white arrowheads point at domains D5 and D6 in different orientations, highlighting the mobility of these domains. The arrowheads in the first class highlight D1 (orange), D2 (purple), D3 (cyan), and D4 (yellow). Scale bar, 100 Å. **b**, low-resolution 3D refinement of *Saccharolobus solfataricus* SlaA (transparent grey; 13.5 Å resolution) superimposed with the atomic model of *S. acidocaldarius* SlaA (rainbow ribbon). While in the *S. solfataricus* map, the N-terminus (blue arrowhead) is not resolved (due to flexibility), the C-terminus is visible (red arrowhead) and in a “closed” conformation, reminiscent of the Alphafold predications shown in Figures 1, 2 and 3, as well as Supplementary Figure 6. **c**, map and model from (**b**) shown in different orientations.

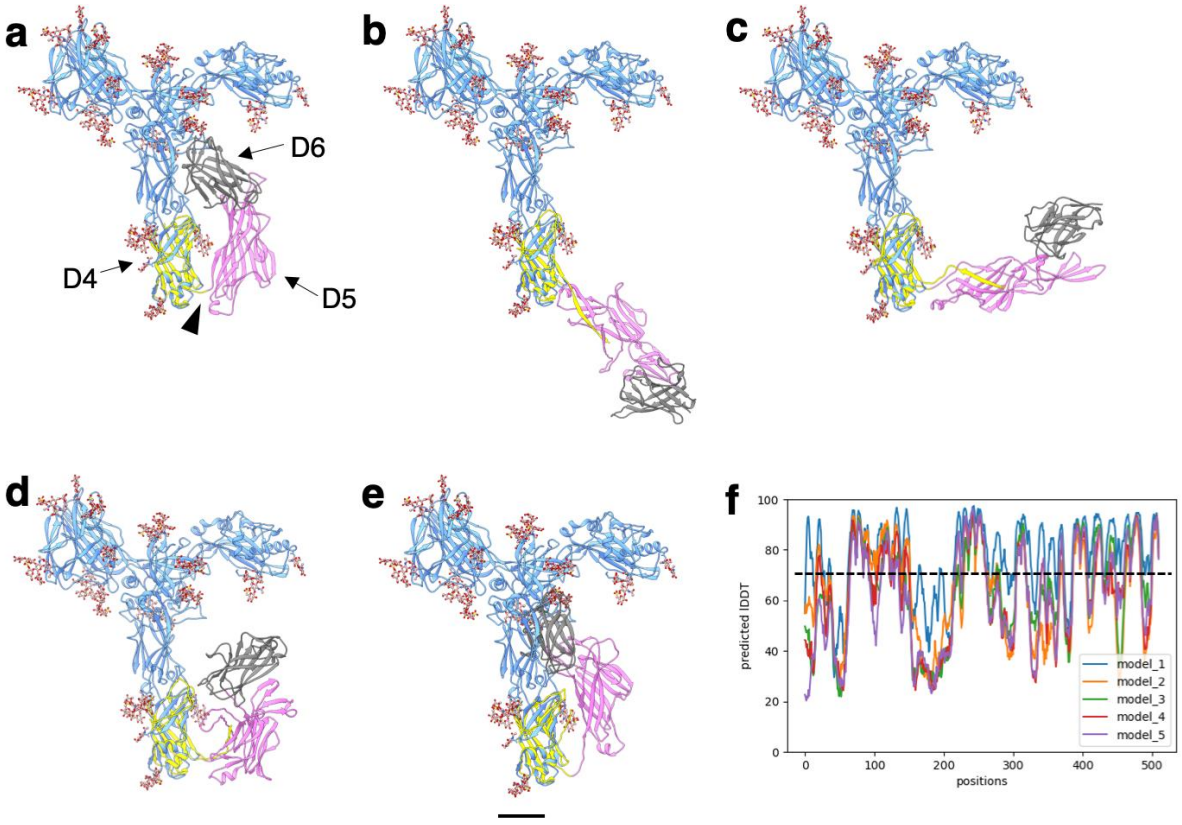


Fig. 1-Figure Supplement 6. Five Alphafold predictions of SlaA$_{914\text{-}1,424}$.

a-e, SlaA$_{30\text{-}1,069}$ is shown in ribbon representation and cornflower blue; glycans are in ball-stick representation and rusty brown. The Alphafold predictions are coloured according to domains, highlighting domains D5 (pink) and D6 (grey). Residues 914-1,069 (D4 in yellow) at the C-terminus of SlaA$_{30\text{-}1,069}$ were included in the prediction to aid alignment between SlaA$_{30\text{-}1,069}$ and D5-D6. The predicted D4 largely overlaps with the cryoEM structure of the SlaA$_{30\text{-}1,069}$ N-terminus. The black arrowhead (**a**) indicates the intramolecular hinge loop. **f**, pLDDT (per-residue confidence score) plot showing the per-residue confidence metric of the predicted models. The dashed line marks the threshold of predicted LDDT=70, above which the structures are expected to be modelled with high confidence. Scale bar, 20 Å.

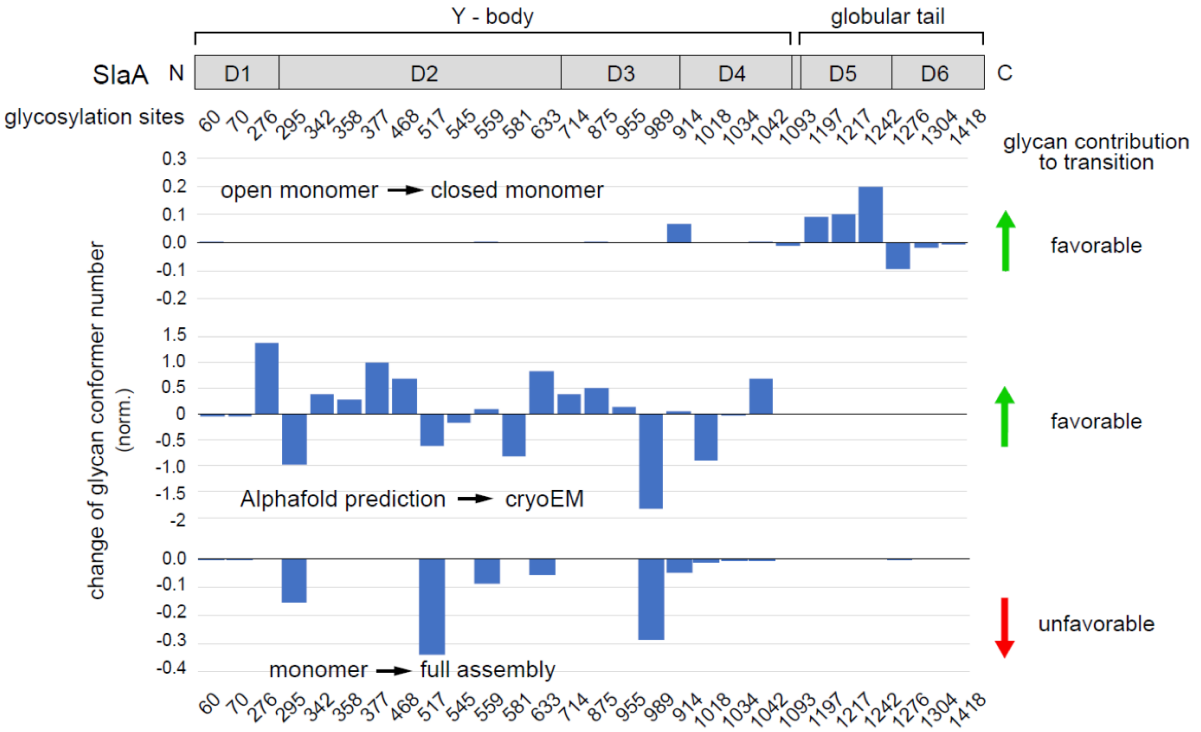


Figure 2-Figure Supplement 1. Entropic contribution of glycans to protein conformation. Position of N-glycosylated sites and globular domains on SlaA primary structure (upper panel) and changes of number of possible N-glycan conformers at each glycosylated sites (bar graphs) for SlaA with stretched (open) or flapped D5-D6 domains (closed), SlaA Y-body predicted by Alphafold v2.2.0 versus experimental cryoEM, and for monomer in isolation or in the assembled structure. Shown in the bar graphs are changes of glycan conformer numbers at individual glycosylation sites normalized by global changes for all glycans. Positive and negative values indicate an increase or a decrease of possible glycan conformations, respectively, indicative of favourable and unfavourable entropic contributions. The 7 last N-glycans of the protein were not taken into account for the Alphafold-cryoEM comparison plot, changing the scale of the Y axis compared to the two other plots. Green and red arrows on the right side of the figure indicate a total increase or decrease of glycan conformers during the transition between the two conformations of the protein that were compared in each plot, indicative of favourable and unfavourable entropic contributions to the conformation transition.

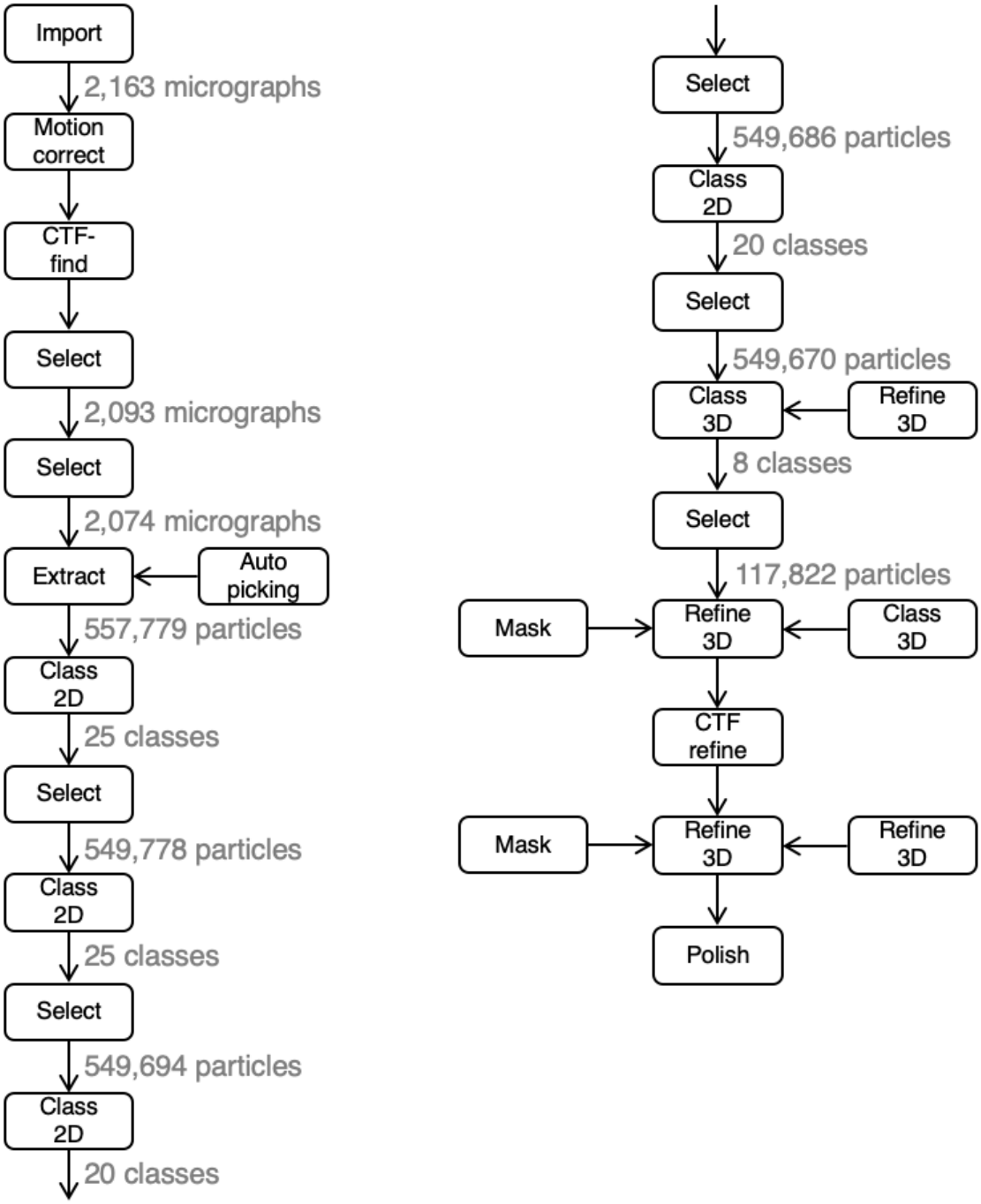


Figure 3-Figure Supplement 1a. Relion processing workflow for pH 7 dataset (part 1).

Relion processing workflow for pH 7 dataset (part 1).

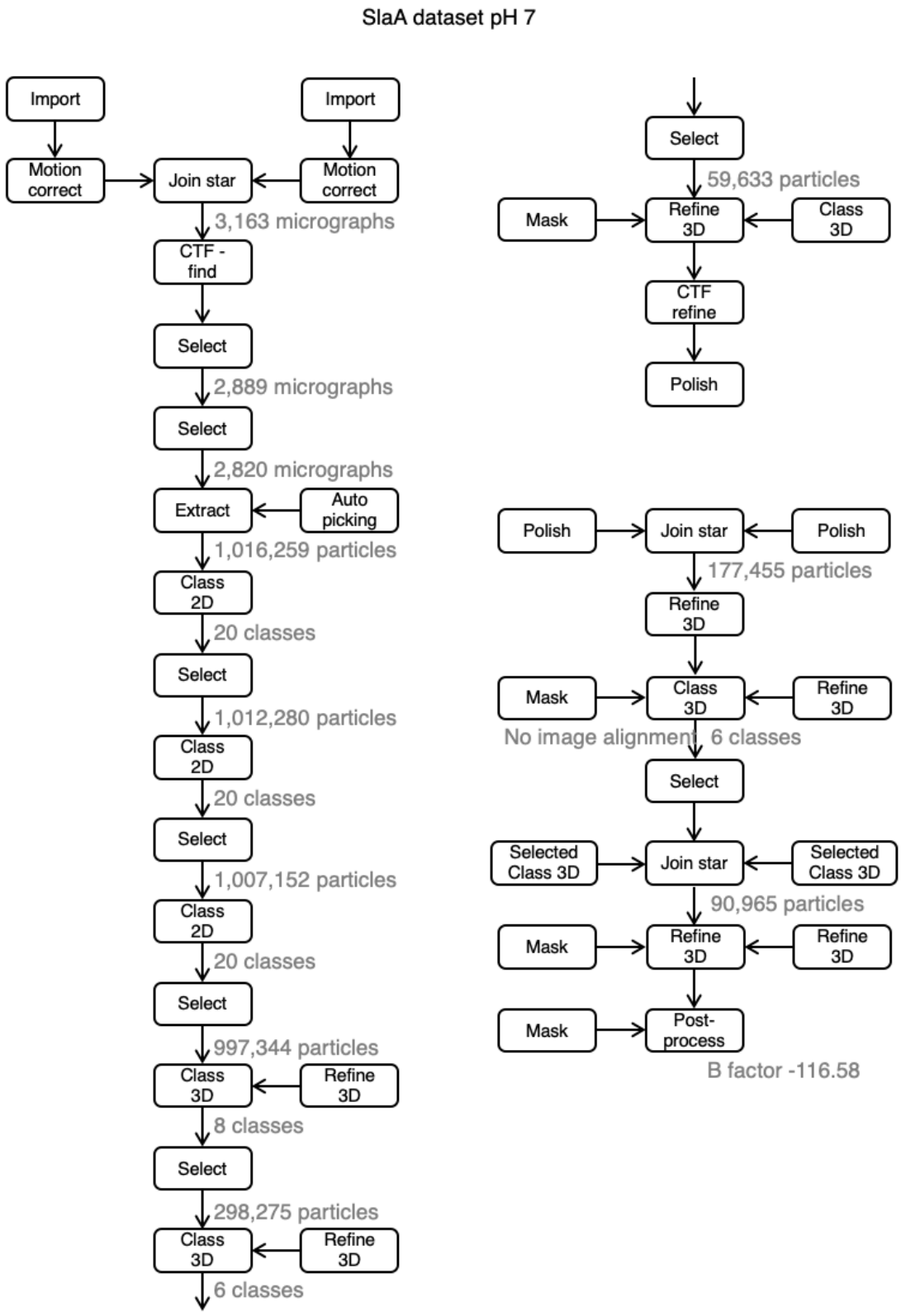


Figure 3-Figure Supplement 1b. Relion processing workflow for pH 7 dataset (part 2).

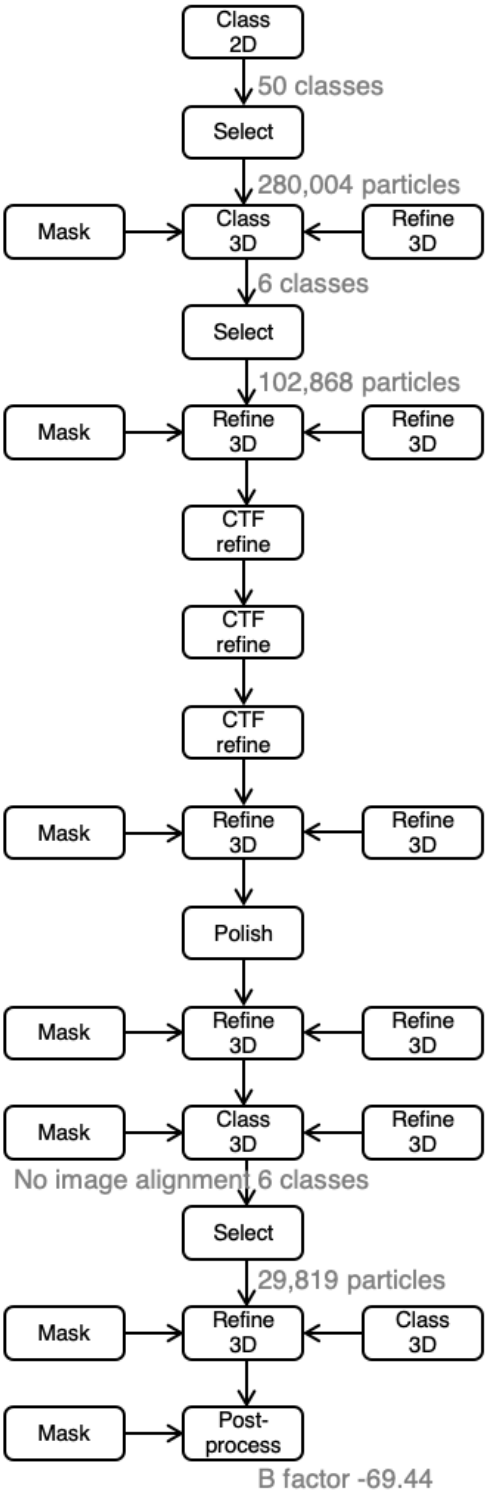


Figure 3-Figure Supplement 2. Relion processing workflow for pH 10 dataset.

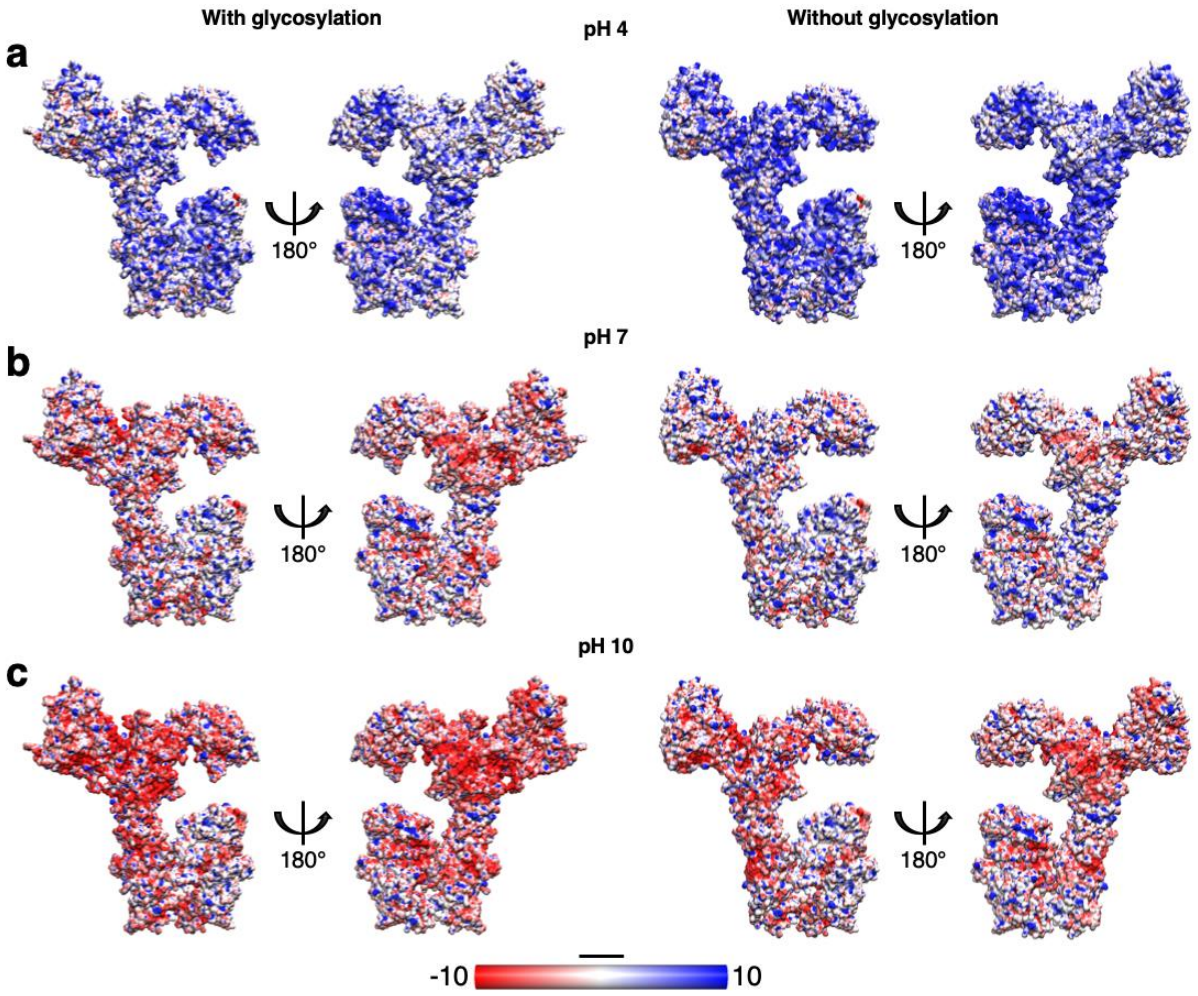


Figure 3-Figure Supplement 3. Impact of glycosylation on the electrostatic surface charge of SlaA at different pH values. Comparison of the SlaA electrostatic surface charge with and without glycans at pH 4 (**a**), 7 (**b**) and 10 (**c**). The glycosylation increases the overall surface negative charge of SlaA, particularly noticeable at pH 7 and 10. Scale bar, 20 Å**.**

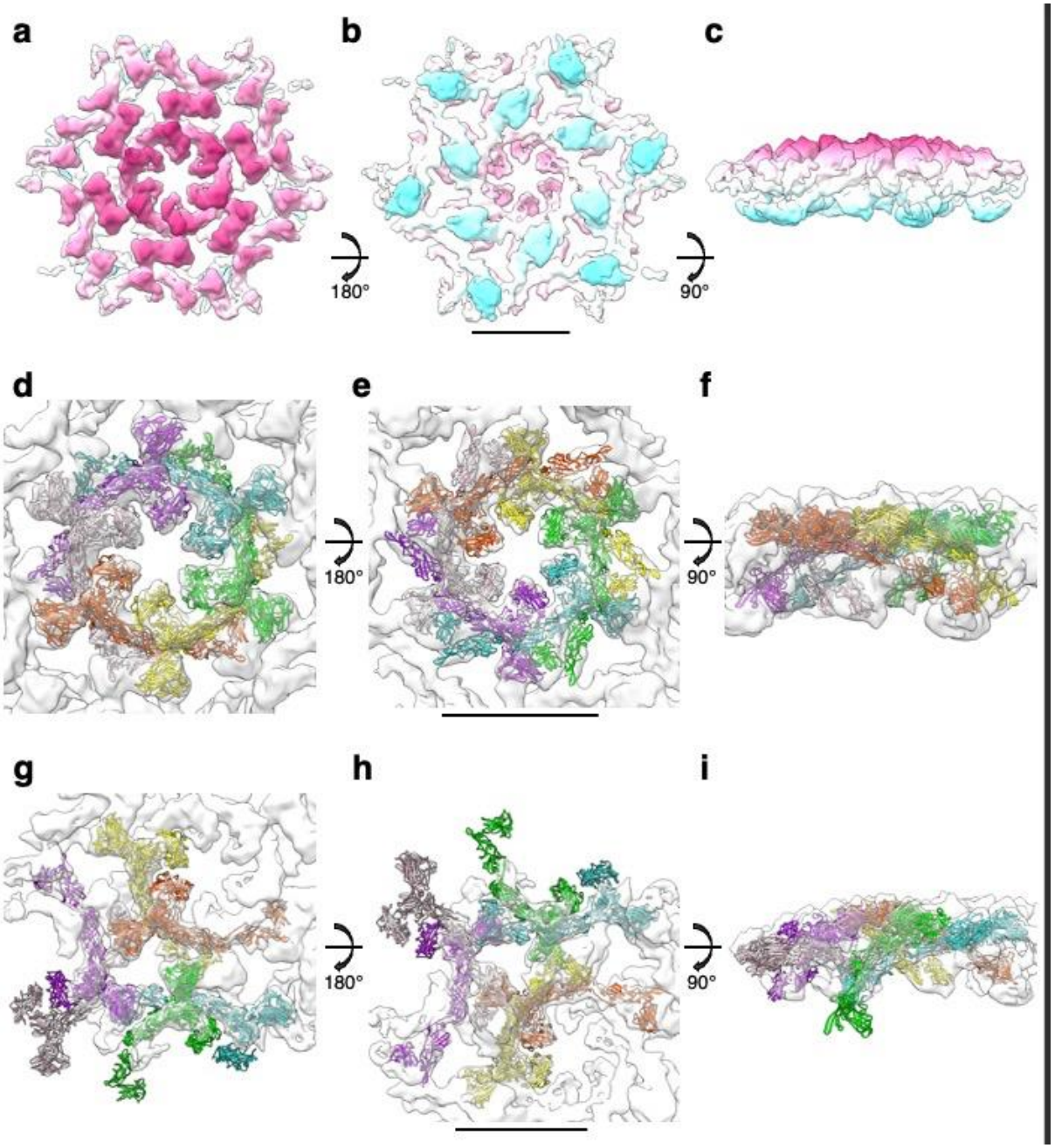


Figure 4-Figure Supplement 1. Subtomogram averaging of the S-layer on exosomes – SlaA fitting. a-**c**, cryoEM map of the S-layer assembled on exosomes in extracellular (**a**), intracellular (**b**), and side (**c**) views at 11.2 Å resolution. The map shows in magenta the membrane-distal and in cyan the membrane-proximal sides of the lattice. **d-f**, fitting of the SlaA hexamer model into the S-layer map in (**a-c**). SlaA is shown in ribbon representation in different colours. **g-i**, fitting of six SlaA monomers around a triangular pore. SlaA monomers are in different colours. Scale bar, 10 nm.

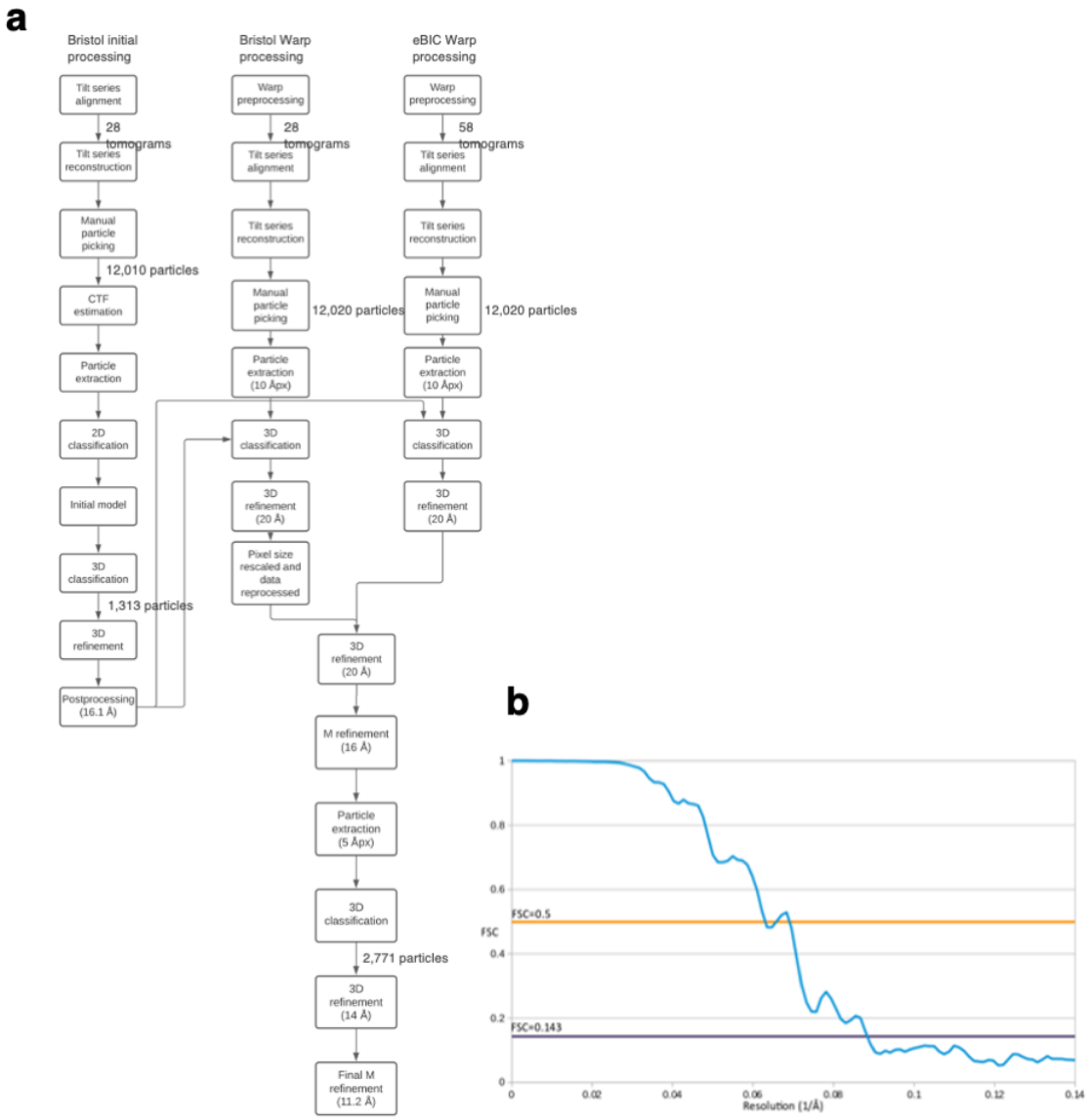


Figure 4-Figure Supplement 2. CryoET processing workflow and data quality for cryoEM map of S-layer on exosomes. a, subtomogram averaging processing workflow using Warp-Relion-M. **b**, gold standard FSC of the subtomogram averaging map.

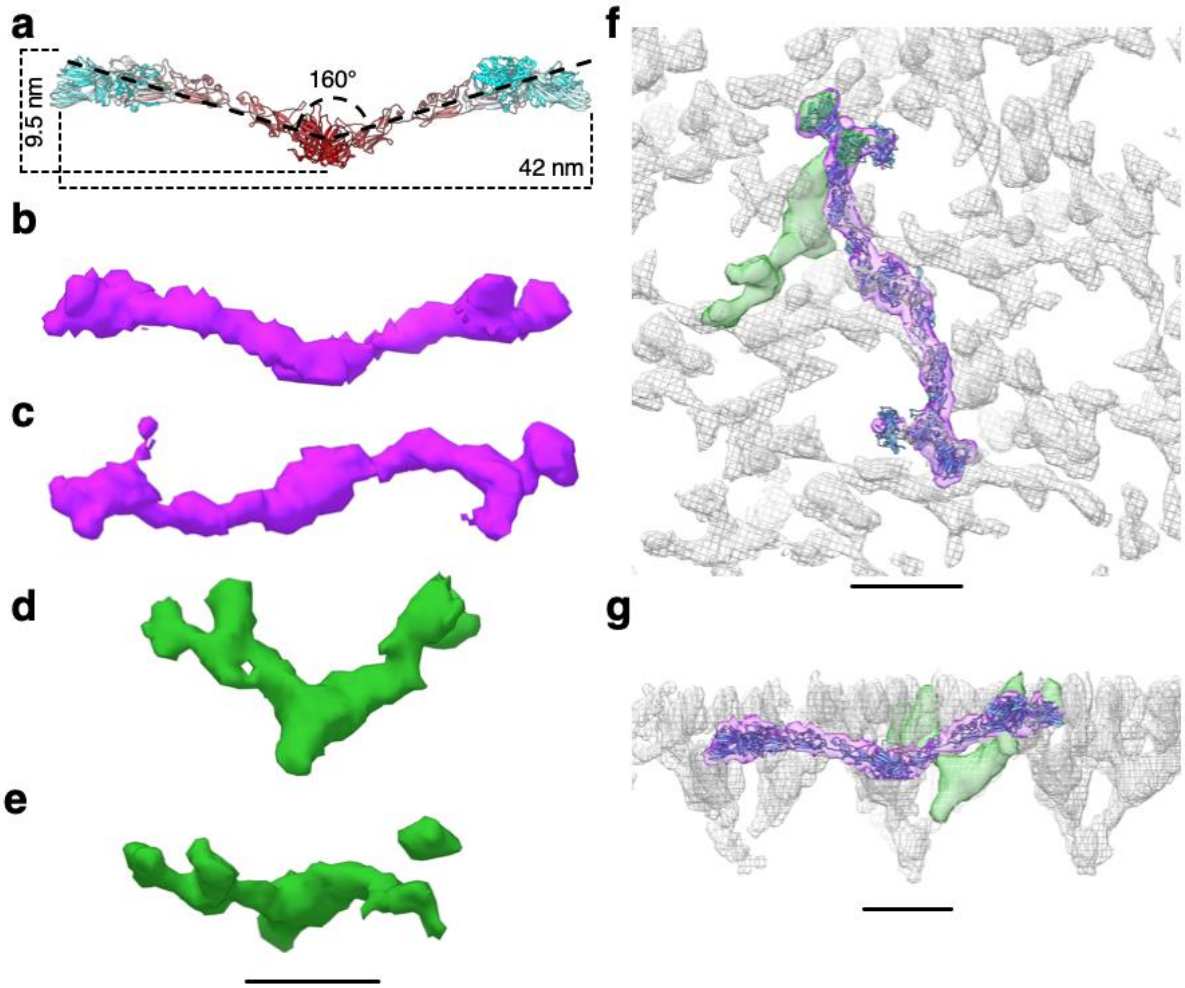


Figure 4-Figure Supplement 3. Comparison between current and previously reported[28] *S. acidocaldarius* SlaA assembly models.

a, side view of the SlaA dimer (ribbon representation in cyan-grey-maroon from N-terminus to C-terminus). Two SlaA monomers form an angle of 160°. The dimer has a height of 9.5 nm and a length of 42 nm. **b-e**, cryoEM densities extrapolated from the S-layer map published in 2019. The purple density in (**b**) and (**c**) (side and extracellular views, respectively) contain the SlaA dimer as presented in this work (**f** and **g**). The green density in (**d**) and (**e**) (side and extracellular views, respectively), represent the SlaA dimer as reported in 2019. **f** and **g**, show the atomic model of the SlaA dimer in (**a**) and the cryoEM densities in (**b-e**) fitting the S-layer cryoEM map (grey mesh) presented in 2019. Scale bar, 10 nm.

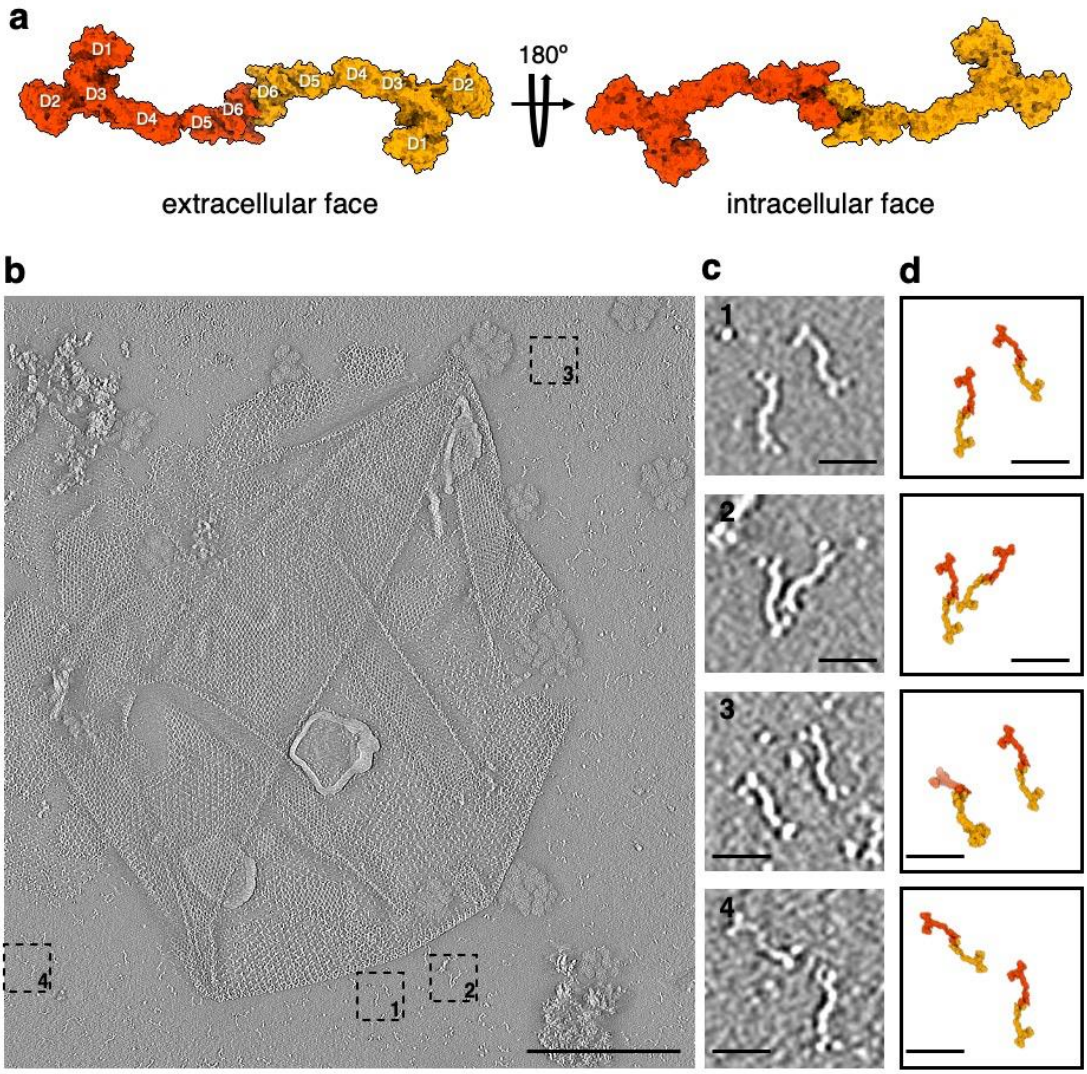


Figure 4-Figure Supplement 4. Isolated SlaA-only S-layer from S. acidocaldarius. a, surface representation of the atomic model of the SlaA dimer, as it occurs in the S-layer. **b**, negative stain electron tomography slice of isolated SlaA-only S-layer. **c**, 1-4 are cut-outs from (**b**) showing dimeric SlaA with their respective positions marked in (**b**). **d**, atomic models from (**a**) scaled and superimposed with the dimers seen in negative stain tomography (**c**). Scale bar (**b**), 200 nm; scale bars (**c-d**), 25 nm.

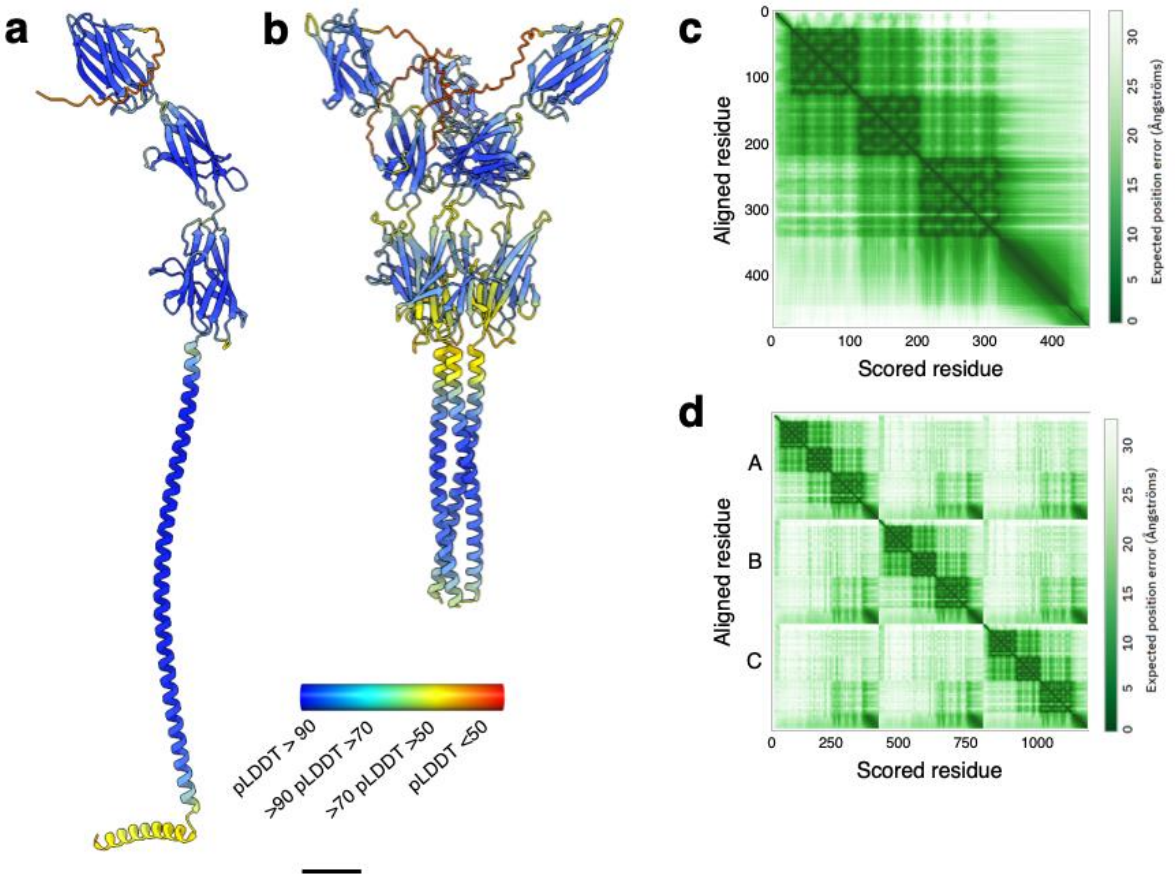
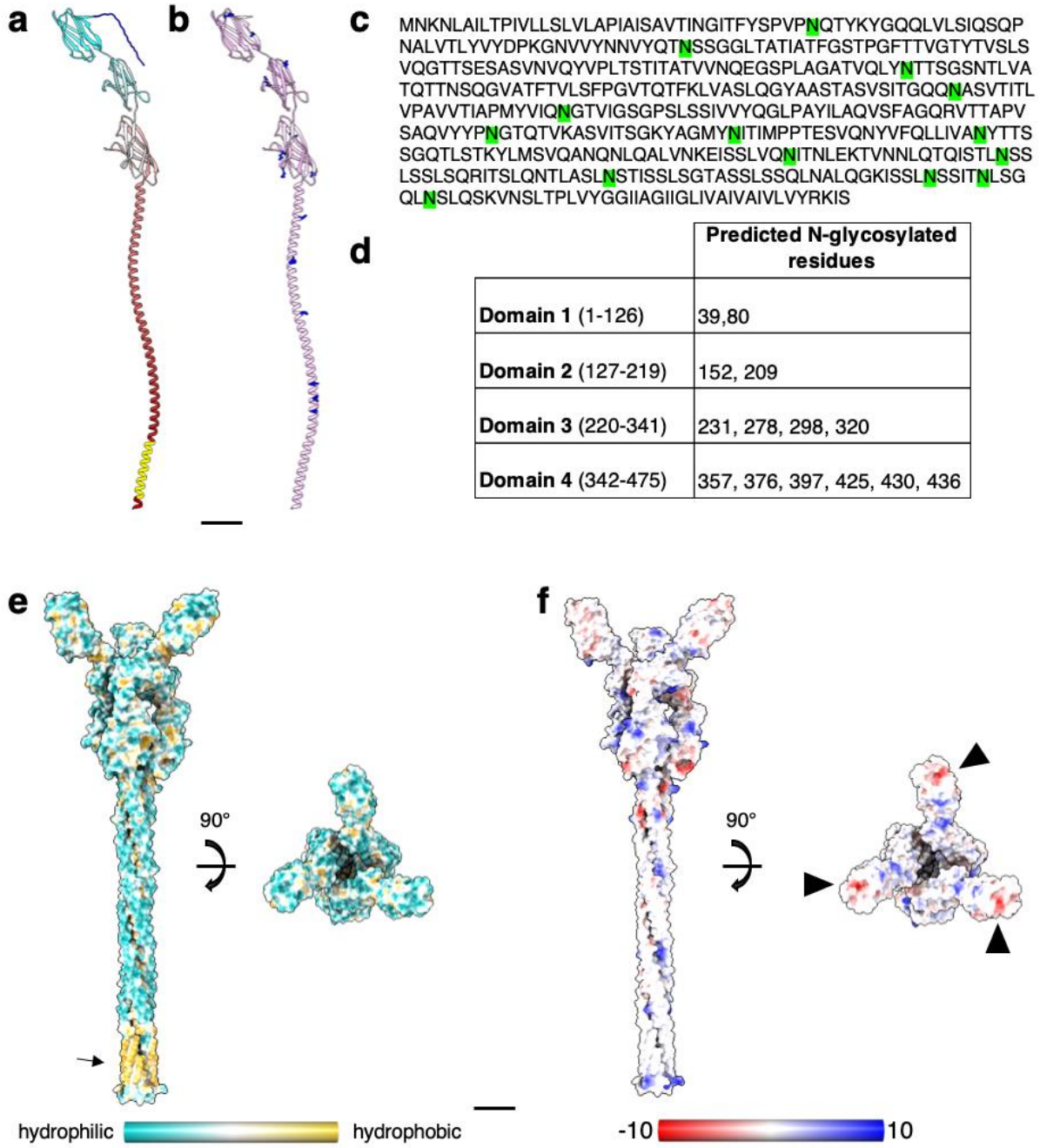


Figure 5-Figure Supplement 1. Alphafold v2.2.0 predictions of SlaB monomer and trimer. **a** and **b**, Alphafold v2.2.0 predictions of SlaB monomer and trimer, respectively. The ribbon is coloured by pLDDT (per-residue confidence score) where red indicates very low confidence and blue very high. The trimeric coiled coil of the SlaB trimer (**b**) is truncated at residue 400, and the complete trimeric coiled coil (Fig. 5d) was predicted using SymmDock. **c** and **d**, PAE (predicted aligned error) plots for (**a**) and (**b**), respectively. Scale bar, 20 Å.



	Predicted N-glycosylated residues
Domain 1 (1-126)	39,80
Domain 2 (127-219)	152, 209
Domain 3 (220-341)	231, 278, 298, 320
Domain 4 (342-475)	357, 376, 397, 425, 430, 436

Figure 5-Figure Supplement 2. Structural prediction of *S. acidocaldarius* SlaB.

a, atomic structure of SlaB as predicted by Alphafold v2.2.0 (ribbon representation, cyan-grey-maroon from N-terminus to C-terminus). Amino acids from 1-24 (blue) are predicted as signal peptide by InterPro. Amino acids from 448-470 (yellow) are predicted as trans-membrane helix by TMHMM – 2.0. **b**, putative N-glycosylation sites are labelled as predicted by GlycoPP v1.0. **c**, SlaB sequence with predicted N-glycosylated residues in green. **d**, table showing predicted N-glycosylation distribution across four SlaB domains. **e** and **f**, SlaB trimer (as predicted by Alphafold v2.2.0) surface representation showing hydrophobicity (from most hydrophilic in dark cyan, to white, to most

hydrophobic in dark goldenrod) in (**e**), and electrostatic surface potential (from mostly negative in red, to white, to mostly positive in blue) in (**f**). The arrow in (**e**) highlights the predicted hydrophobic trans-membrane region. Arrowheads in **f** indicate negatively-charged patches that may electrostatically interact with the mostly positively charged SlaA. Scale bar, 20 Å.

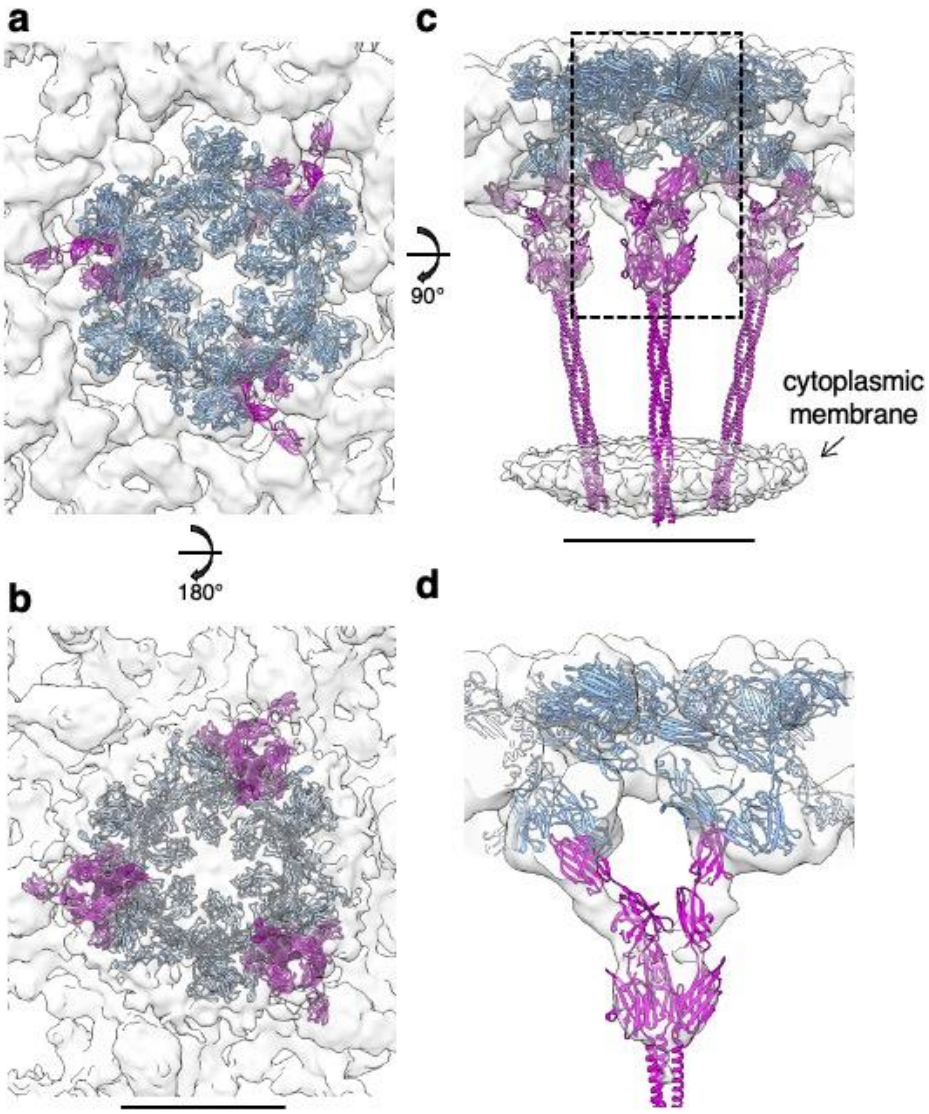


Figure 5-Figure Supplement 3. Subtomogram averaging of the S-layer on exosomes – SlaB fitting. a-d, SlaA hexamer (cornflower blue) and SlaB trimer (magenta) fitting into the cryoEM map at higher threshold than that presented in Supplementary Fig. 11. (**d**) is a slice through the highlighted region in (**c**) showing the interaction between SlaA and SlaB. Scale bar, 10 nm.

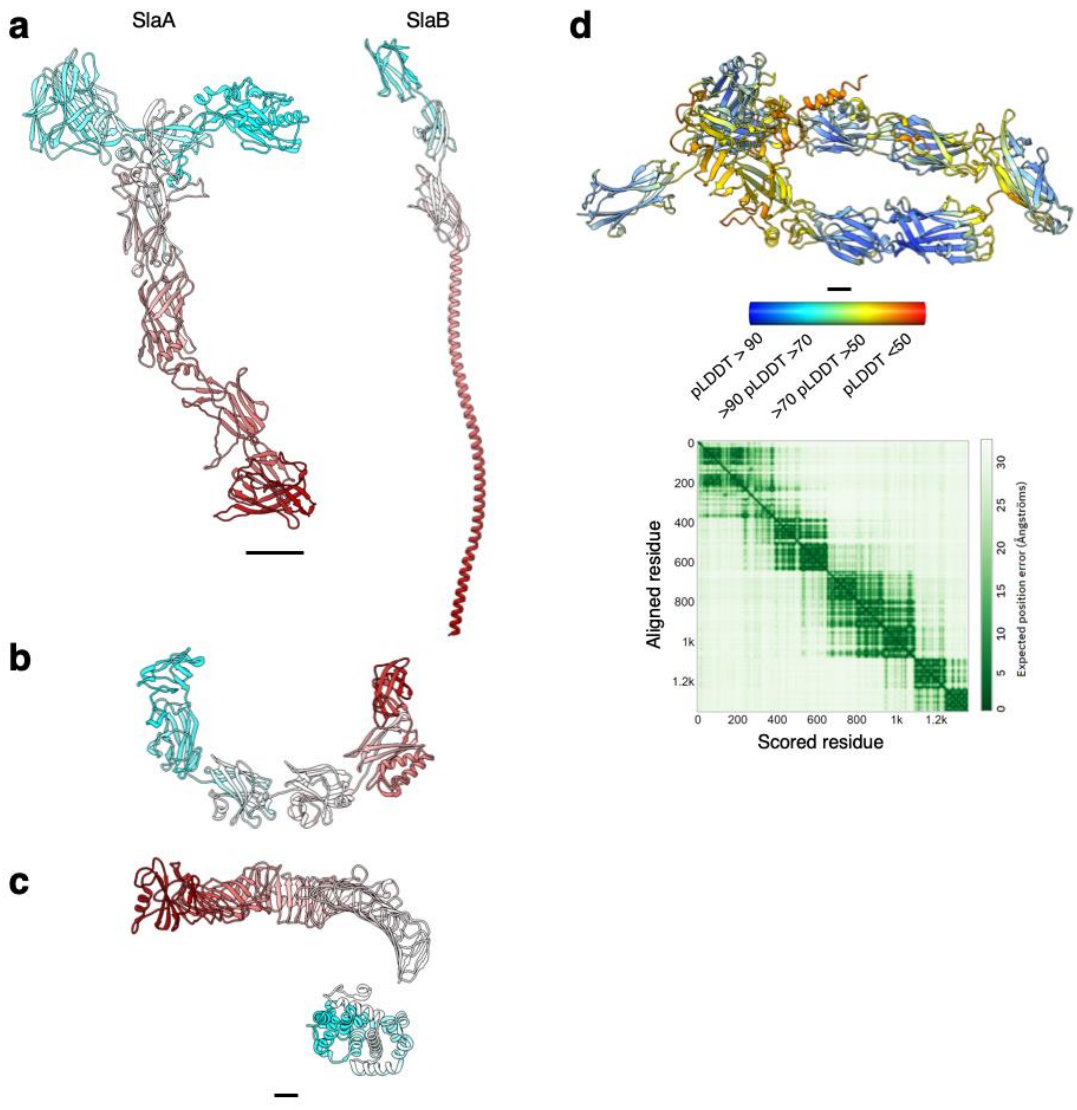


Figure 5-Figure Supplement 4. Structure of archaeal and bacterial S-layer proteins. a-c, atomic models are shown in ribbon representation in cyan-grey-maroon from the N-terminus to the C-terminus. *S. acidocaldarius* SlaA (domains D5 and D6 as predicted by Alphafold v 2.2.0) and SlaB (as predicted by Alphafold v.2.2.0) are in (**a**), *H. volcanii* csg (PDB ID: 7PTR, http://dx.doi.org/10.2210/pdb7ptr/pdb) is in (**b**), and *C. crescentus* RsaA (N-terminus PDB ID: 6T72, http://dx.doi.org/10.2210/pdb6t72/pdb, C-terminus PDB ID: 5N8P, http://dx.doi.org/10.2210/pdb5n8p/pdb) is in (**c**). **d**, SlaA atomic model of *M. sedula* (UNIPROT A4YHQ8, as predicted by Alphafold v2.2.0) and PAE (predicted aligned error) plot. The model is shown in ribbon representation coloured by pLDDT (per-residue confidence score) where red indicates very low confidence and blue very high. Scale bar, 20 Å.

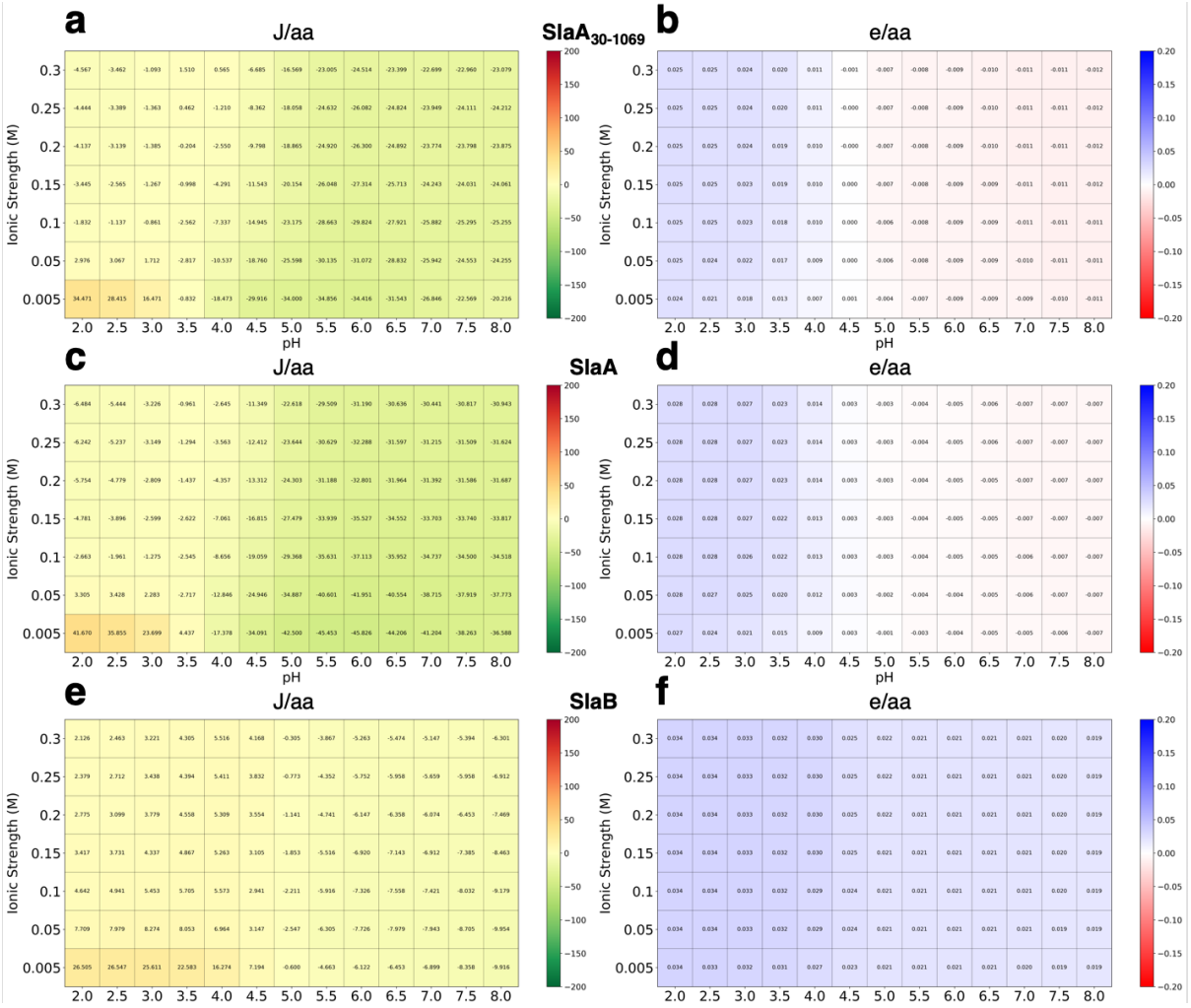


Figure 5-Figure Supplement 5. Stability and charge heatmaps for *S. acidocaldarius* SlaA$_{30\text{-}1,069}$, SlaA and SlaB. **a**, **c** and **e**, calculated folded state stability heatmaps for SlaA$_{30\text{-}1,069}$ (**a**), SlaA (**c**) and SlaB (**e**), respectively. SlaA$_{30\text{-}1,069}$, SlaA and SlaB are stable across pH 2-8. **b**, **d** and **f**, calculated charged heatmaps for SlaA$_{30\text{-}1,069}$ (**b**), SlaA (**d**) and SlaB (**f**), respectively. Surface charge shifts from positive to negative for SlaA$_{30\text{-}1,069}$ and SlaA from pH 2 to 8, whereas SlaB shows a largely consistent positive charge.

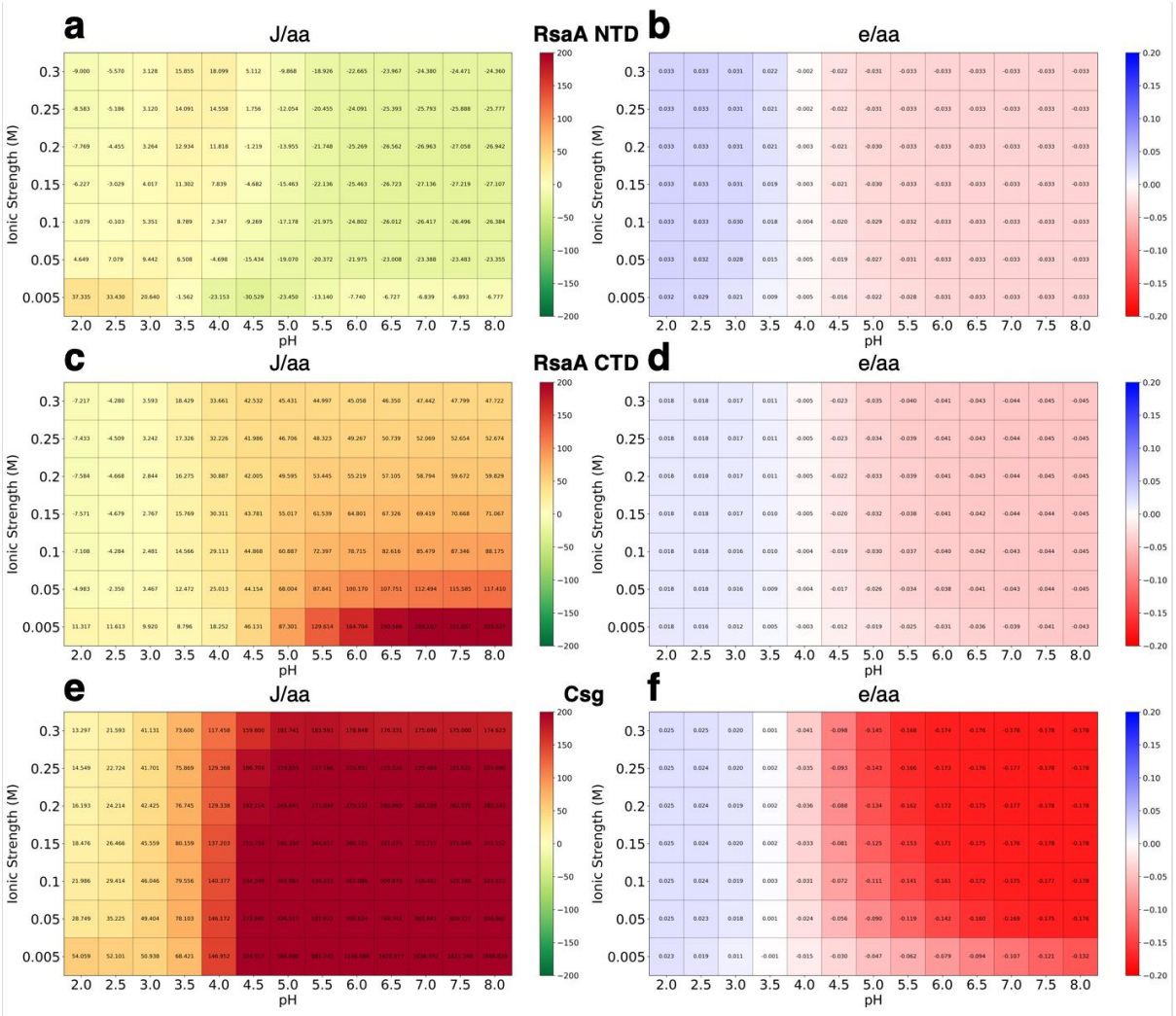


Figure 5-Figure Supplement 6. Stability and charge heatmaps for *C. crescentus and H. volcanii* S-layer proteins. **a**, **c** and **e**, calculated folded state stability heatmaps for *C. crescentus* S-layer protein RsaA N-terminus (**a**) and C-terminus (**c**) domains, and *H. volcanii* S-layer protein csg (**e**). The RsaA N-terminus domain is largely stable across pH 2-8; RsaA C-terminus domain becomes unstable at elevated pH and low ionic strength; csg's stability is greatly affected at neutral and high pH. **b**, **d** and **f**, calculated charged heatmaps for RsaA N-terminus (**b**) and C-terminus (**d**) domains, and csg (**f**). RsaA's and csg's surface charge shifts from positive to negative from pH 2 to 8. Csg shows a dramatic difference in surface charge from pH 3 to 5, becoming negatively charged.

Supplementary File 1 legends

Supplementary File 1a: Statistics of data collection, 3D reconstruction and validation.

Supplementary File 2b: Mapping of glycan residues from the structure file to residues of the GLYCAM force field or the newly charge-derived SG0 and SG4 residues, representing the 1-substituted and 1,4-substituted SMA.

Supplementary File 1c: RESP charges derived for residue SG0 on the HF/6-31G*//HF/6-31G* level of theory (see Methods for details).

Supplementary File 1d: RESP charges derived for residue SG4 on the HF/6-31G*//HF/6-31G* level of theory (see Methods for details).

Supplementary Videos

Supplementary Video 1. Atomic structure and glycosylation of $SlaA_{30\text{-}1069}$.

The $SlaA_{30\text{-}1069}$ cryoEM map is shown in cornflower blue. The atomic structure is shown in ribbon representation in cyan-grey-maroon from the N-terminus to the C-terminus. The glycosylated Asn residues are in orange and the glycans are represented as ball-stick in steel blue, with N atoms in blue, O in red and S in yellow.

Supplementary Video 2. Flexibility of SlaA.

Sequence of 2D classifications obtained in Relion 3 of negatively stained SlaA. D1-4 were aligned, showing the flexibility of D1, D5 and D6.

Supplementary Video 3. Comparison of $SlaA_{30\text{-}1069}$ structure at pH 4 and 10. r.m.s.d. alignment between $SlaA_{30\text{-}1069}$ atomic models at pH 4 and pH 10. Smaller deviations are shown in blue and larger deviations in red, with mean r.m.s.d. = 0.79 Å, as in Figure 3b.

Supplementary Video 4. Model of the assembled *S. acidocaldarius* S-layer.

References

1. Bharat, T. A. M., von Kügelgen, A. & Alva, V. Molecular Logic of Prokaryotic Surface Layer Structures. *Trends Microbiol* **29,** 405–415 (2021).
2. Fagan, R. P. & Fairweather, N. F. Biogenesis and functions of bacterial S-layers. *Nature Reviews Microbiology* vol. 12 Preprint at https://doi.org/10.1038/nrmicro3213 (2014).
3. Klingl, A., Pickl, C. & Flechsler, J. Archaeal Cell Walls. *Subcell Biochem* **92,** 471–493 (2019).
4. Albers, S. V. & Meyer, B. H. The archaeal cell envelope. *Nature Reviews Microbiology 2011 9:6* **9,** 414–426 (2011).
5. Rodrigues-Oliveira, T., Belmok, A., Vasconcellos, D., Schuster, B. & Kyaw, C. M. Archaeal S-layers: Overview and current state of the art. *Front Microbiol* **8,** 2597 (2017).
6. Shalev, Y., Turgeman-Grott, I., Tamir, A., Eichler, J. & Gophna, U. Cell surface glycosylation is required for efficient mating of Haloferax volcanii. *Front Microbiol* **8,** 1253 (2017).
7. Tittes, C., Schwarzer, S. & Quax, T. E. F. Viral Hijack of Filamentous Surface Structures in Archaea and Bacteria. *Viruses,* **13,** 164 (2021).
8. Schwarzer, S., Hackl, T., Oksanen, H. M. & Quax, T. E. F. Archaeal Host Cell Recognition and Viral Binding of HFTV1 to Its Haloferax Host. *mBio* **14,** (2023).
9. Sleytr, U. B., Schuster, B., Egelseer, E. M. & Pum, D. S-layers: principles and applications. *FEMS Microbiol Rev* **38,** 823 (2014).
10. Abdul-Halim, M. F. *et al.* Lipid anchoring of archaeosortase substrates and midcell growth in haloarchaea. *mBio* **11,** (2020).
11. Engelhardt, H. & Peters, J. Structural research on surface layers: A focus on stability, surface layer homology domains, and surface layer-cell wall interactions. *J Struct Biol* **124,** 276–302 (1998).
12. Engelhardt, H. Are S-layers exoskeletons? The basic function of protein surface layers revisited. *J Struct Biol* **160,** 115–124 (2007).
13. Meyer, B. H. & Albers, S. V. Hot and sweet: Protein glycosylation in crenarchaeota. *Biochem Soc Trans* **41,** 384–392 (2013).
14. Cohen, S., Shilo, M. & Kessel, M. Nature of the salt dependence of the envelope of a Dead Sea archaebacterium, Haloferax volcanii. *Archives of Microbiology,* **156,** 198–203 (1991).
15. von Kügelgen, A., Alva, V. & Bharat, T. A. M. Complete atomic structure of a native archaeal cell surface. *Cell Rep* **37,** 110052 (2021).

16. Herdman, M. *et al.* High-resolution mapping of metal ions reveals principles of surface layer assembly in Caulobacter crescentus cells. *Structure* **30,** 215-228.e5 (2022).

17. Baranova, E. *et al.* SbsB structure and lattice reconstruction unveil Ca2+ triggered S-layer assembly. *Nature* **487,** 119–122 (2012).

18. Lanzoni-Mangutchi, P. *et al.* Structure and assembly of the S-layer in C. difficile. *Nat Commun* **13,** (2022).

19. Bharat, T. A. M. *et al.* Structure of the hexagonal surface layer on Caulobacter crescentus cells. *Nat Microbiol* **2,** (2017).

20. von Kügelgen, A. *et al.* In Situ Structure of an Intact Lipopolysaccharide-Bound Bacterial Surface Layer. *Cell* **180,** 348-358.e15 (2020).

21. von Kügelgen, A. *et al.* Interdigitated immunoglobulin arrays form the hyperstable surface layer of the extremophilic bacterium Deinococcus radiodurans. *Proc Natl Acad Sci U S A* **120,** (2023).

22. Jing, H. *et al.* Archaeal surface layer proteins contain β propeller, PKD, and β helix domains and are related to metazoan cell surface proteins. *Structure* **10,** 1453–1464 (2002).

23. Arbing, M. A. *et al.* Structure of the surface layer of the methanogenic archaean Methanosarcina acetivorans. **109,** 11812–11817 (2012).

24. Brock, T. D., Brock, K. M., Belly, R. T. & Weiss, R. L. Sulfolobus: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Archiv für Mikrobiologie* **84,** 54–68 (1972).

25. Grogan, D. W. Organization and interactions of cell envelope proteins of the extreme thermoacidophile Sulfolobus acidocaldarius. *Can J Microbiol* **42,** 1163–1171 (1996).

26. Veith, A. *et al.* Acidianus, Sulfolobus and Metallosphaera surface layers: structure, composition and gene expression. *Mol Microbiol* **73,** 58–72 (2009).

27. Lyon, I. *et al.* Structure of the S-layer of Sulfolobus acidocaldarius. *Nature* **299,** 840–842 (1982).

28. Gambelli, L. *et al.* Architecture and modular assembly of Sulfolobus S-layers revealed by electron cryotomography. *Proc Natl Acad Sci U S A* **116,** 25278–25286 (2019).

29. Zhang, C. *et al.* Cell structure changes in the hyperthermophilic crenarchaeon Sulfolobus islandicus lacking the S-layer. *mBio* **10,** (2019).

30. Scheres, S. H. W. Amyloid structure determination in RELION-3.1. *Acta Crystallogr D Struct Biol* **76,** 94–101 (2020).

31. Postic, G., Ghouzam, Y., Chebrek, R. & Gelly, J.-C. An ambiguity principle for assigning protein structural domains. *Sci Adv*. **3,** e1600552 (2017)

32. Holm, L. Using Dali for Protein Structure Comparison. *Methods Mol Biol* **2112,** 29–42 (2020).

33. Kato, K. *et al.* High-resolution cryo-EM structure of photosystem II reveals damage from high-dose electron beams. *Communications Biology* **4,** 1–11 (2021).

34. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (2021).

35. Peyfoon, E. *et al.* The S-layer glycoprotein of the crenarchaeote Sulfolobus acidocaldarius is glycosylated at multiple sites with chitobiose-linked N-glycans. *Archaea* 754101, (2010).

36. Zähringer, U., Moll, H., Hettmann, T., Knirel, Y. A. & Schäfer, G. Cytochrome b558/566 from the archaeon Sulfolobus acidocaldarius has a unique Asn-linked highly branched hexasaccharide chain containing 6-sulfoquinovose. *Eur J Biochem* **267,** 4144–4149 (2000).

37. Sikora, M. *et al.* Computational epitope map of SARS-CoV-2 spike protein. *PLoS Comput Biol* **17,** e1008790 (2021).

38. Gecht, M. *et al.* GlycoSHIELD: a versatile pipeline to assess glycan impact on protein structures. *bioRxiv* 2021.08.04.455134 (2022) doi:10.1101/2021.08.04.455134.

39. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7,** (2018).

40. Jensen, J. Calculating pH and salt dependence of protein-protein binding. *Curr Pharm Biotechnol* **9,** 96–102 (2008).

41. Zhang, Z., Witham, S. & Alexov, E. On the role of electrostatics on protein-protein interactions. *Phys Biol* **8,** 035001 (2011).

42. Ellen, A. F. *et al.* Proteomic analysis of secreted membrane vesicles of archaeal Sulfolobus species reveals the presence of endosome sorting complex components. *Extremophiles* **13,** 67–79 (2009).

43. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nature Methods* **16,** 1146–1152 (2019).

44. Tegunov, D., Xue, L., Dienemann, C., Cramer, P. & Mahamid, J. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nat Methods* **18,** (2021).

45. Sikora, M. *et al.* Desmosome architecture derived from molecular dynamics simulations and cryo-electron tomography. *Proc Natl Acad Sci U S A* **117,** 27132–27140 (2020).

46. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* **33,** W363–W367 (2005).

47. Kidmose, R. T. *et al.* Namdinator - Automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ* **6**, (2019).

48. Jarrell, K. F. *et al.* N-Linked Glycosylation in Archaea: a Structural, Functional, and Genetic Analysis. *Microbiol Mol Biol Rev* **78**, 304 (2014).

49. Abu-Qarn, M. *et al.* Haloferax volcanii AglB and AglD are involved in N-glycosylation of the S-layer glycoprotein and proper assembly of the surface layer. *J Mol Biol* **374**, 1224–1236 (2007).

50. Kelly, J. F. *et al.* Identification of a novel N-linked glycan on the archaellins and S-layer protein of the thermophilic methanogen, methanothermococcus thermolithotrophicus. *Journal of Biological Chemistry* **295**, 14618–14629 (2020).

51. Meyer, B. H. & Albers, S. V. AglB, catalyzing the oligosaccharyl transferase step of the archaeal N-glycosylation process, is essential in the thermoacidophilic crenarchaeon Sulfolobus acidocaldarius. *Microbiologyopen* **3**, 531 (2014).

52. Cohen-Rosenzweig, C., Guan, Z., Shaanan, B. & Eichler, J. Substrate Promiscuity: AglB, the Archaeal Oligosaccharyltransferase, Can Process a Variety of Lipid-Linked Glycans. *Appl Environ Microbiol* **80**, 486–496 (2014).

53. Lupas, A. *et al.* Domain structure of the Acetogenium kivui surface layer revealed by electron crystallography and sequence analysis. *J Bacteriol* **176**, 1224–1233 (1994).

54. Herrmann, J. *et al.* A bacterial surface layer protein exploits multistep crystallization for rapid self-assembly. *Proc Natl Acad Sci U S A* **117**, 388–394 (2020).

55. D'Angelo, M. A. & Hetzer, M. W. Structure, dynamics and function of nuclear pore complexes. *Trends Cell Biol* **18**, 456–466 (2008).

56. Szabó, Z. *et al.* Flagellar motility and structure in the hyperthermoacidophilic archaeon Sulfolobus solfataricus. *J Bacteriol* **189**, 4305–4309 (2007).

57. Henche, A. L. *et al.* Structure and function of the adhesive type IV pilus of Sulfolobus acidocaldarius. *Environ Microbiol* **14**, 3188–3202 (2012).

58. Fröls, S. *et al.* UV-inducible cellular aggregation of the hyperthermophilic archaeon Sulfolobus solfataricus is mediated by pili formation. *Mol Microbiol* **70**, 938–952 (2008).

59. Pum, D., Messner, P. & Sleytr, U. B. Role of the S layer in morphogenesis and cell division of the archaebacterium Methanocorpusculum sinense. *J Bacteriol* **173**, 6865–6873 (1991).

60. Rodrigues-Oliveira, T. *et al.* Environmental factors influence the Haloferax volcanii S-layer protein structure. *PLoS One* **14**, e0216863 (2019).

61. Hebditch, M. & Warwicker, J. Web-based display of protein surface and pH-dependent properties for assessing the developability of biotherapeutics. *Scientific Reports* **9,** 1–9 (2019).

62. Meyer, B. H. *et al.* Sulfoquinovose synthase – an important enzyme in the N-glycosylation pathway of Sulfolobus acidocaldarius. *Mol Microbiol* **82,** 1150–1163 (2011).

63. Yurist-Doutsch, S., Chaban, B., VanDyke, D. J., Jarrell, K. F. & Eichler, J. Sweet to the extreme: protein glycosylation in Archaea. *Mol Microbiol* **68,** 1079–1084 (2008).

64. Vogt, G., Woell, S. & Argos, P. Protein thermal stability, hydrogen bonds, and ion pairs. *J Mol Biol* **269,** 631–643 (1997).

65. Luo, G. *et al.* Slp-coated liposomes for drug delivery and biomedical applications: Potential and challenges. *International Journal of Nanomedicine* vol. 14 Preprint at https://doi.org/10.2147/IJN.S189935 (2019).

66. Fioravanti, A., Mathelie-Guinlet, M., Dufrêne, Y. F. & Remaut, H. The Bacillus anthracis S-layer is an exoskeleton-like structure that imparts mechanical and osmotic stabilization to the cell wall . *PNAS Nexus* **1,** (2022).

67. Charrier, M. *et al.* Engineering the S-Layer of Caulobacter crescentus as a Foundation for Stable, High-Density, 2D Living Materials. *ACS Synth Biol* **8,** 181–190 (2019).

68. Pallares, R. M. *et al.* Precision Engineering of 2D Protein Layers as Chelating Biogenic Scaffolds for Selective Recovery of Rare-Earth Elements. *J Am Chem Soc* **144,** 854–861 (2022).

69. Zhang, T. *et al.* Optimization of Encapsulation Using Milk Polar Lipid Liposomes with S-Layer Protein and Transport Study of the ACE-Inhibitory Peptide RLSFNP. *J Agric Food Chem* **69,** 7049–7056 (2021).

70. Schuster, B. & Sleytr, U. B. S-Layer Ultrafiltration Membranes. *Membranes (Basel)* **11,** (2021).

71. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10,** 584–590 (2013).

72. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192,** 216–221 (2015).

73. Sanchez-Garcia, R. *et al.* DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Communications Biology 2021 4:1* **4,** 1–8 (2021).

74. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62,** 1002–1011 (2006).

75. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallogr D Biol Crystallogr.* **67,** 355–367 (2011).

76. Emsley, P. *et al*. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66,** 486–501 (2010).

77. Lebedev, A. A. *et al.* JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr D Biol Crystallogr* **68,** 431–440 (2012).

78. Pettersen, E.F *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30,** 70–82 (2021).

79. Croll, T.I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74,** 519–530 (2018).

80. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallog* **66,** 12–21 (2009).

81. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67,** 235–242 (2011).

82. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer Visualization of Three-Dimensional Image Data Using IMOD. *J Struct Biol* **116,** 71–76 (1996).

83. Nicastro, D. *et al.* The molecular architecture of axonemes revealed by cryoelectron tomography. *Science* **313,** 944–948 (2006).

84. Bharat, T. A. M. & Scheres, S. H. W. Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nat Protoc* **11,** 2054–2065 (2016).

85. Zheng, S. *et al.* AreTomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction. *J Struct Biol X* **6,** (2022).

86. Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Science* **27,** 112–128 (2018).

87. Dolinsky, T. J. *et al.* PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* **35,** (2007).

88. Kirschner, K. N. *et al.* GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J Comput Chem* **29,** 622 (2008).

89. Breneman, C. M. & Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J Comput Chem* **11,** 361–373 (1990).

90. Dupradeau, F. Y. *et al.* The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Physical Chemistry Chemical Physics* **12,** 7821–7839 (2010).

91. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J Mol Graph* **14,** 33–38 (1996).

92. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25,** 1605–1612 (2004).

93. Delano, W. L. The PyMOL Molecular Graphics System. *CCP4 Newsletter on protein crystallography* **40,** (2002).

94. Blum, M. *et al.* The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res* **49,** D344–D354 (2021).

95. Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol* **305,** 567–580 (2001).

96. Chauhan, J. S., Bhat, A. H., Raghava, G. P. S. & Rao, A. GlycoPP: A Webserver for Prediction of N- and O-Glycosites in Prokaryotic Protein Sequences. *PLoS One* **7,** e40155 (2012).

97. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29,** 1859–1865 (2008).

98. Park, S. J. *et al.* CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology* **29,** 320–331 (2019).

99. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods 2016 14:1* **14,** 71–73 (2016).

100. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2,** 19–25 (2015).