O-glycosylation of intrinsically disordered regions regulates homeostasis of membrane proteins in streptococci

Mohammad M. Rahman[1], Svetlana Zamakhaeva[1], Jeffrey S. Rush[2], Catherine T. Chaton[2], Cameron W. Kenner[1], Yin Mon Hla[3], Ho-Ching Tiffany Tsui[3], Vladimir N. Uversky[4], Malcolm E. Winkler[3], Konstantin V. Korotkov[2], and Natalia Korotkova[1, 2*]

[1]Department of Microbiology, Immunology and Molecular Genetics, University of Kentucky, Lexington, Kentucky, USA

[2]Department of Molecular and Cellular Biochemistry, University of Kentucky, Lexington, Kentucky, USA

[3]Department of Biology, Indiana University Bloomington, Bloomington, Indiana, USA

[4]Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, Florida, USA

*Correspondence and request for materials should be addressed to N.K. (email: nkorotkova@uky.edu)

Short title: O-glycosylation of IDRs of membrane proteins in streptococci

Keywords: *O*-glycosylation, intrinsically disordered region, secreted peptidyl-prolyl isomerase, foldase, biofilm, *Streptococcus mutans*, *Streptococcus pyogenes*, *Streptococcus pneumoniae*

Abstract

Proteins harboring intrinsically disordered regions (IDRs) lacking stable secondary or tertiary structures are abundant across the three domains of life. These regions have not been systematically studied in prokaryotes. Our genome-wide analysis identifies extracytoplasmic serine/threonine-rich IDRs in several biologically important membrane proteins in streptococci. We demonstrate that these IDRs are *O*-glycosylated with glucose by glycosyltransferases GtrB and PgtC2 in *Streptococcus pyogenes* and *Streptococcus pneumoniae*, and with N-acetylgalactosamine by a Pgf-dependent mechanism in *Streptococcus mutans*. Absence of glycosylation leads to a defect in biofilm formation under ethanol-stressed conditions in *S. mutans*. We link this phenotype to the C-terminal IDR of a post-translocation secretion chaperone PrsA. *O*-glycosylation of the IDR protects this region from proteolytic degradation. The IDR length attenuates the efficiency of glycosylation and, consequently, the expression level of PrsA. Taken together, our data reveal that *O*-glycosylation of IDRs functions as a dynamic switch of protein homeostasis in streptococci.

Introduction

Many protein regions, and even entire proteins, do not adopt a defined three-dimensional structure under physiological conditions. These protein regions are known as intrinsically disordered regions (IDRs). The structural flexibility and plasticity of IDRs provide a remarkable functional advantage that allows them not only to serve as linkers connecting various structured domains but also to participate in multiple functions [1]. Many IDRs bind a broad range of molecular partners, such as proteins, nucleic acids, small molecules, carbohydrates, lipids, or metal ions. The attractive intermolecular and intramolecular interactions between IDRs have been shown to contribute to the driving force for phase separation, leading to condensation of a protein-liquid phase [2]. Moreover, some IDRs display the ability to sense the local chemical environment [3]. This feature of IDRs is due to their unique biophysical and structural properties and the presence of posttranslational modifications [4, 5]. Accordingly, IDRs are over-represented in proteins involved in cell signaling and regulatory processes such as transcription regulation, cell cycle, and protein folding [4, 6].

IDRs are prevalent in eukaryotes, where more than 30% of proteins are predicted to have extended IDRs. In contrast, fewer proteins bearing similar IDRs have been reported in prokaryotes. Protein disorder in prokaryotes has not been systematically identified and characterized even in model organisms such as *Escherichia coli* and *Bacillus subtilis*. However, recent experimental evidence highlights the importance of prokaryotic IDRs in various aspects of cell physiology, including signaling pathways, protein sorting mechanisms, cell division, and liquid-liquid phase separation [7, 8, 9, 10, 11, 12, 13, 14]. In this study, we explored two important yet understudied properties of bacterial proteomes: disorder and protein glycosylation. Gram-positive bacteria *Streptococcus mutans*, *Streptococcus pyogenes*, and *Streptococcus pneumoniae* inhabit a variety of sites in the human body, from the mouth and nasopharynx to the skin, causing various infections. We reveal that these bacteria possess protein glycosylation mechanisms that add *O*-sugars on the IDRs of membrane and secreted proteins. The glycoproteins identified in our study are involved in critical cellular processes such as protein folding, cell wall biosynthesis, cell division, and signal transduction. *S. mutans* is the major contributor to human dental plaque formation, primarily living in biofilms on the tooth

surfaces [15]. We discover that *O*-glycosylation of the C-terminal IDR of a post-translocation secretion chaperone PrsA is required for biofilm formation of *S. mutans* under ethanol stress. Further investigations indicate that N-acetylgalactosamine moieties on the PrsA IDR protect this important region from proteolytic cleavage. Hence, our report not only reveals the mechanism utilized by *S. mutans* to modulate biofilm synthesis but also provides a roadmap for future functional studies of glycosylated IDRs in prokaryotes.

Results

GtrB and PgtC2 catalyze glucosylation of *S. pyogenes* membrane proteins

In bacteria, GT-A type glycosyltransferases typically transfer a sugar moiety from nucleotide-sugars to a lipid carrier, undecaprenyl phosphate (Und-P), which subsequently serves as sugar donor to a wide range of acceptor molecules, including proteins and polysaccharides [16, 17]. We recently demonstrated that in *S. mutans,* a GT-A type glycosyltransferase SccN is a UDP-glucose:Und-P glucosyl transferase, which synthesizes β-Glc-P-Und [18]. This glycolipid is required for modification of the cell wall serotype *c* rhamnopolysaccharide, SCC, with the α-glucose (Glc) side-chains [19]. Many bacteria, including streptococcal species, encode an SccN homolog, GtrB [16], of previously unknown function (Supplementary Fig. 1). To investigate the enzymatic activity of this enzyme in *S. pyogenes* (M5005_Spy_420, hereafter $GtrB^{Spy}$), we expressed the protein on a plasmid in the *E. coli* JW2347 strain which is devoid of glucosyl phosphoryl undecaprenol (Glc-P-Und) synthase activity [20]. When $GtrB^{Spy}$ activity was assayed in *E. coli* membrane fractions *in vitro* using UDP-[^{3}H]Glc and exogenously added Und-P [17], we observed robust time-dependent incorporation of [^{3}H]Glc into the glycolipid fraction. A kinetic analysis demonstrated that $GtrB^{Spy}$ possesses 8 μM and 51 μM affinity for UDP-Glc and Und-P, respectively, suggesting that $GtrB^{Spy}$ is an active Glc-P-Und synthase (Supplementary Fig. 2a, b). The TLC analysis of the lipids on silica gel detected a single peak of radioactivity, with chromatographic mobilities expected for [^{3}H] Glc-P-Und (Supplementary Fig. 2c). No incorporation of radioactivity into [^{3}H]glycolipid was noted when the reaction was conducted with the *E. coli* strain carrying an empty plasmid. To characterize the $GtrB^{Spy}$ activity in the *S. pyogenes* membranes, we constructed the GtrB-deficient mutant in the *S. pyogenes* MGAS2221 background (2221Δ*gtrB*) and the

2221Δ*gtrB* strain complemented with the plasmid expressing the WT copy of *gtrB*^{Spy} (2221Δ*gtrB*:p*gtrB*). When the membranes of these strains were incubated with UDP-[^{3}H]Glc and Und-P, two classes of glucolipids were detected in wild-type (WT) and 2221Δ*gtrB*:p*gtrB*: a major glucolipid product with properties consistent with a glucosyldiglyceride and a minor glucolipid, presumably Glc-P-Und (Fig. 1a, c), which co-migrates on TLC with the lipid accumulating in the *E. coli* pBAD33_GtrB^{Spy} membranes (Supplementary Fig. 2c). The major glucolipid product was also detected in 2221Δ*gtrB* (Fig. 1b). When glucosyldiglyceride was eliminated by mild alkaline de-acylation, the WT membranes contained a single peak of radioactivity corresponding to Glc-P-Und (Supplementary Fig. 3a). The membranes of 2221Δ*gtrB* did not catalyze the synthesis of Glc-P-Und (Supplementary Fig. 3b). The complementation of the mutant with the plasmid resulted in a robust increase in Glc-P-Und synthase activity (Supplementary Fig. 3c). The minor lipid product was stable to mild alkaline methanolysis, but sensitive to mild acid (0.1 N HCl, 50 °C, 50% isopropanol) [17], consistent with its identification as Glc-P-Und. Thus, these results implicate GtrB^{Spy} in the synthesis of Glc-P-Und.

Given that *S. pyogenes* cell wall polysaccharide does not contain Glc moieties [21], we hypothesized that GtrB^{Spy} is involved in the first step of protein glucosylation. In Gram-positive bacteria, protein glycosylation of surface proteins such as Serine-Rich Repeat Proteins adhesins, pili, flagellins and S-layer proteins have been reported [22]. However, *S. pyogenes* cells do not bind fluorescent concanavalin A (ConA) lectin, which selectively recognizes α-mannose/α-glucose residues [23, 24], suggesting that surface proteins do not carry Glc moieties. To test the role of GtrB^{Spy} in glucosylation of membrane proteins, we used sepharose-bound ConA to purify glycoproteins from detergent-solubilized extracts of the WT, 2221Δ*gtrB*, and 2221Δ*gtrB*:p*gtrB* membranes. Bound putative glycoproteins were eluted by the hapten sugar, methyl α-D-mannopyranoside. The SDS-PAGE analysis demonstrated that the glycoproteins are present in WT and 2221Δ*gtrB*:p*gtrB*, but not in 2221Δ*gtrB* (Fig. 1d). This result establishes that *S. pyogenes* membrane proteins undergo glucosylation that requires the participation of GtrB^{Spy}. The fact that glycoproteins are bound tightly by ConA confirms the presence of non-reducing terminal α-Glc units.

We hypothesized that the last step of protein glucosylation is catalyzed by an unknown GT-C fold glycosyltransferase, which transfers α-Glc from Glc-P-Und to membrane proteins. This enzyme family participates in asparagine-linked *N*-glycosylation and serine or threonine-linked *O*-glycosylation in higher-order eukaryotes and some Gram-negative bacteria [25]. *S. pyogenes* encodes two GT-C type glycosyltransferases, M5005_Spy_1794 and M5005_Spy_1862 (hereafter PgtC1 and $PgtC2^{Spy}$), with unknown functions. $PgtC2^{Spy}$ belongs to the YfhO protein family, which is universally present in Firmicutes (Supplementary Fig. 4). Interestingly, in *B. subtilis* and *Listeria monocytogenes*, YfhO homologs are involved in the transfer of sugar moieties to teichoic acid polymers [26]. We constructed the PgtC1 and $PgtC2^{Spy}$ deletion mutants (2221Δ*pgtC1* and 2221Δ*pgtC2*) and the corresponding complemented strains (2221Δ*pgtC1*:p*pgtC1* and 2221Δ*pgtC2*:p*pgtC2*) in *S. pyogenes*. We used these strains for purification of glycoproteins on ConA as described above. This approach revealed that glucosylation of membrane proteins requires the participation of $PgtC2^{Spy}$, but not PgtC1 (Fig. 1d and Supplementary Fig. 5).

S. pyogenes PrsA1, PBP1a, LCP1 and PknB are glycoproteins

Mass spectrometry (MS)-based proteomic analysis of *S. pyogenes* glycoproteins present in the gel bands (Fig. 1d, lane 1) identified a conserved post-translational chaperone lipoprotein of peptidyl-prolyl cis/trans isomerase family PrsA1 [27] and a class A penicillin-binding protein PBP1a as the most abundant proteins in the WT eluate. Additionally, the analysis detected LytR-Cps2A-Psr family protein (LCP1) involved in the attachment of polysaccharides to peptidoglycan (known as LytR in *S. pyogenes* [28]), serine/threonine (Ser/Thr) protein kinase PknB [29, 30] and a homolog of PrsA1, PrsA2. To validate the results of lectin affinity enrichment, we evaluated the expression of PrsA1, PrsA2, and PknB in WT, 2221Δ*gtrB*, and 2221Δ*gtrB*:p*gtrB* using protein-specific polyclonal antibodies. Deletion of $gtrB^{Spy}$ resulted in increased electrophoretic mobility of PrsA1 and PknB, but not PrsA2 (Fig. 1e, f and Supplementary Fig. 6). We suggested that MS analysis is likely misidentified PrsA2 because PrsA1 and PrsA2 have 65% amino acid identity. The electrophoretic behavior of PrsA1 and PknB in 2221Δ*gtrB* is consistent with an alteration of molecular weight due to the loss of Glc moieties. Expression of the WT copy of $gtrB^{Spy}$

or the *S. mutans sccN* on a plasmid in 2221Δ*gtrB* affected the electrophoretic behavior of PrsA1 and PknB (Fig. 1e, f). These experiments support the conclusion that the observed change in the molecular weight of PrsA1 and PknB in 2221Δ*gtrB* is due to the absence of α-Glc, donated by β-Glc-P-Und.

Streptococcal putative glycoproteins harbor Ser/Thr-rich IDRs

Multiple studies reported that ConA binds weakly to α-mannose/α-glucose monosaccharides, whereas the lectin exhibits a high affinity to oligosaccharides [23, 24]. A strong association of ConA with *S. pyogenes* glycoproteins implies the presence of clusters of α-Glc units that provide a tight binding site for ConA. When we analyzed the AlphaFold2 [31] structural models of PrsA1, PBP1a, PknB, and LCP1, we observed that their C-terminal regions lack secondary structure and have low pLDDT confidence scores (Fig. 1g and Supplementary Fig. 7). The low pLDDT scores correlate with predicted disorder and can serve as a predictor of intrinsic disorder [32, 33]. Visual inspection of the amino acid sequences of C-terminal regions revealed the presence of serine (Ser) and threonine (Thr) rich regions suggesting *O*-glucosylation of these sites (Fig. 2a). Notably, the ConA-enrichment strategy may fail to capture all glycoproteins because of the low abundance of some glycoproteins in the analyzed conditions and their high resistance to tryptic cleavage used for MS analysis. To identify other putative membrane or secreted glycoproteins in streptococci proteomes, we developed an automated pipeline that uses protein language-based models TMbed and DR-BERT to annotate membrane and secreted proteins, and disorder propensities for all proteins (Fig. 2b, Methods) [34, 35]. We analyzed *S. pyogenes* MGAS2221, *S. mutans* Xc, and *S. pneumoniae* TIGR4 proteomes to select membrane and secreted proteins with predicted IDRs containing at least 10 consecutive residues (Supplementary Dataset 1). The relative membrane topology of predicted IDRs in selected membrane proteins was assigned by TOPCONS [36]. We also analyzed AlphaFold2 [31] models to validate the disorder predictions. Remarkably, we found that all analyzed streptococci possess proteins with Ser/Thr-rich IDRs (Fig. 2a). Among the proteins harboring this type of IDRs, we identified PrsA and PBP1a in all streptococci, PknB and LCP1 in *S. pyogenes* and *S. mutans* (LCP1 is known as BrpA [37]), the extracytoplasmic regulator of c-di-AMP synthesis CdaR [38], cell division proteins MapZ

[39, 40, 41] and RodZ [42] in *S. pneumoniae*, and lipoteichoic acid synthase LtaS and a divisome protein FtsQ [43] in *S. mutans* (Supplementary Figs. 7, 8 and 9, Supplementary Table 1). Additionally, these Ser/Thr-rich regions were shown to be highly disordered by several commonly used disorder predictors assembled into the RIDAO platform [44]. In fact, in all the cases analyzed in this study, the Ser/Thr-rich regions are invariantly disordered and are often located within longer IDRs (Fig. 2c, Supplementary Fig. 10). This is an important but expected observation, since the amino acid sequences of many IDPs and IDRs are known to be characterized by the presence of repeats, and since it was indicated that the level of repeat perfection correlates with their tendency to be intrinsically disordered or unstructured, with more perfect repeats being less structured [45]. Interestingly, this type of IDRs is frequently located at the C-termini of the membrane-associated proteins. Furthermore, protein sequence alignments revealed the presence of Ser/Thr-rich IDRs in the homologs of the identified proteins from streptococcal species and other Firmicutes (Supplementary Fig. 11, 12). To understand if IDRs carry important functional features, we generated sequence logos (WebLogo [46]) for each subset of IDRs (Supplementary Fig. 13, 14). The results showed that the main feature of the PrsA, PBP1a, LtaS, PBP1b, PknB, and CdaR IDRs is the presence of poly-Ser/Thr tracks adjacent to a folded domain suggesting the functional importance of this region.

GtrB and PgtC2 glucosylate Ser/Thr-rich IDRs in *S. pneumoniae*

To test a hypothesis that extracytoplasmic Ser/Thr-rich IDRs are glycosylated, we first focused on identifying membrane glycoproteins in *S. pneumoniae.* We were able to purify the membrane glycoproteins from the *S. pneumoniae* serotype 2 D39W derivative (D39Δ*cps*; IU1824) and the *S. pneumoniae* serotype 4 TIGR4 strain using glycoprotein enrichment on ConA-bound sepharose (Fig. 3a and Supplementary Fig. 15a). Proteomic analysis of captured proteins revealed that both strains produce five putative glycoproteins: PrsA, PBP1a, MapZ, RodZ and CdaR. Additionally, we identified PBP1b and peptidoglycan hydrolase LytC [47] in D39Δ*cps*, and membrane protein SP_2093 in TIGR4. As outlined above, C-terminal Ser/Thr-rich IDRs are present in PrsA, PBP1a, PBP1b, and CdaR. MapZ, RodZ, and SP_2093 possess linker IDRs. In contrast, LytC has an N-terminal IDR enriched with Ser and Thr (Fig. 2a, 3b and Supplementary Fig. 9).

Interestingly, contrary to *S. pyogenes* PknB, the *S. pneumoniae* PknB homolog, known as StkP, lacks the C-terminal IDR (Supplementary Fig. 15b) and does not bind to ConA sepharose. We constructed GtrB (SPD_1431) deletion mutant in the D39Δ*cps* background (D39Δ*cps*Δ*gtrB*), and PgtC2 (SPD_2058) deletion mutants in the D39Δ*cps* and TIGR4 backgrounds (D39Δ*cps*Δ*pgtC2* and TIGR4Δ*pgtC2*, respectively) to examine the roles of these enzymes in protein glucosylation. The glycoproteins were absent in the ConA-bond fractions of the mutants, indicating that the protein glucosylation pathway in *S. pneumoniae* is similar to *S. pyogenes* (Fig. 3a, Supplementary Fig. 15a).

To confirm the results of the lectin-affinity enrichment, the electrophoretic behavior of PBP1b, PBP1a, PrsA, and StkP was assessed by immunoblot analysis in the D39W derivative (*cps*$^+$ WT; IU1781), its isogenic Δ*gtrB* mutant (*cps*$^+$Δ*gtrB*; IU18498) and two strains in which *gtrB*^{Spn} was ectopically expressed from a constitutive promoter in the backgrounds of *cps*$^+$ WT (*gtrB*$^+$; IU18715) and *cps*$^+$Δ*gtrB* (Δ*gtrB*//P-*gtrB*; IU18719). PBP1b, PBP1a, and PrsA revealed increased electrophoretic mobility in the *cps*$^+$Δ*gtrB* mutant, indicating a loss of sugar moieties. In contrast, no alteration in the electrophoretic behavior of the PBP1b, PBP1a, and PrsA variants lacking the C-terminal IDRs in the *cps*$^+$Δ*gtrB* background was observed in comparison to the *cps*$^+$ WT background, supporting the hypothesis that the IDRs are the sites of Glc attachment (Fig. 3c, d, e). As expected, the deletion of GtrB^{Spn} did not change the electrophoretic mobility of StkP (Supplementary Fig. 15f), providing evidence that this protein does not carry Glc residues.

Quantitative analysis of the expression of PBP1b, PBP1a, and PrsA and their IDR-less variants in the *S. pneumoniae* WT and the *gtrB* mutants established that glucosylation does not affect the expression of these proteins, but the IDR enhances the expression of PBP1a (Extended Data Fig. 1a, b, c and Supplementary Fig. 15c, d, e).

S. mutans **Pgf pathway glycosylates Ser/Thr-rich IDRs with GalNAc**

Surprisingly, no ConA-bound proteins were recovered from detergent-solubilized extracts of *S. mutans* Xc membranes despite the presence of GtrB (known as SccN) and PgtC2 (Smu_2160) homologs. However, *S. mutans* expresses the cell wall-anchored collagen-binding adhesins (CBAs) WapA, Cnm, CnaB, and CbpA harboring Ser/Thr-rich regions. WapA and Cnm are *O*-glycosylated with *O*-N-acetylhexosamines, presumably N-

acetylgalactosamine (GalNAc) and N-acetylglucosamine (GlcNAc), by the Pgf gene cluster encoded enzymes: the GT-A type glycosyltransferase PgfS, an epimerase PgfE, and two GT-C type glycosyltransferases, PgfM1 and PgfM2, that also belong to YfhO protein family [48, 49]. This gene cluster also encodes Cnm, CnaB, and CbpA in *S. mutans* Xc (Fig. 4a) [49]. We hypothesized that the Pgf pathway glycosylates extracytoplasmic Ser/Thr-rich IDRs of membrane proteins. To test this idea, we first focused on the identification of sugars attached to IDRs. We constructed the Xc∆*pgfS*, Xc∆*pgfM1*, Xc∆*pgfM2*, and Xc∆*pgtC2* mutants in *S. mutans* Xc and used these strains in a co-sedimentation assay with the fluorescent GalNAc and GlcNAc-specific lectins. We observed that the GalNAc-specific lectin, *Wisteria floribunda* agglutinin (WFA), binds to WT, Xc∆*pgfM2*, and Xc∆*pgtC2*, but not to Xc∆*pgfS*, Xc∆*pgfM1*, and the mutant lacking the adhesins encoded by the Pgf gene cluster, Xc∆*cnaB-cbpA-cnm* (Extended Data Fig. 2a). Furthermore, *S. mutans* WT cells, but not the Xc∆*pgfS*, Xc∆*pgfM1* and *S. pyogenes* cells (used as a negative control) associate with other GalNAc-specific lectins, *Helix pomatia* agglutinin (HPA) and *Griffonia simplicifolia* lectin (BS-I) (Extended Data Fig. 2b, c). Binding of WFA to the WT cells was completely inhibited by the hapten sugar, GalNAc (Fig. 4b). In contrast to WFA, the GlcNAc-specific lectin, wheat-germ agglutinin (WGA), associated with the WT and Xc∆*pgfS* cells, and the addition of the WGA hapten sugar, GlcNAc, did not reduce the binding, indicating a non-specific interaction of WGA with *S. mutans* cells (Extended Data Fig. 2d). Fluorescent microscopy of *S. mutans* strains confirmed that WFA binds the cell surface of WT and Xc∆*pgfS* complemented with *pgfS* on a plasmid (Xc∆*pgfS*:p*pgfS*), but not Xc∆*pgfS* (Fig. 4c). Thus, these experiments strongly suggest that the enzymes of the Pgf pathway, PgfS and PgfM1, transfer GalNAc on the cell wall-anchored adhesins encoded by *cnaB-cbpA-cnm*.

Next, to identify *S. mutans* membrane proteins decorated with GalNAc, we employed WFA agarose for glycoprotein enrichment as we described for *S. pyogenes* and *S. pneumoniae*. Bound proteins were eluted with GalNAc, separated on the SDS-PAGE gel, and identified by MS analysis. Two major WFA-binding proteins, PrsA and PBP1a, were detected in the membranes of WT, Xc∆*pgfM2* and Xc∆*pgtC2*, but not Xc∆*pgfS* and Xc∆*pgfM1* (Fig. 4d). Furthermore, increased electrophoretic mobility of PrsA

was observed in Xc∆*pgfM2*. These data strongly suggest that the Pgf pathway enzymes glycosylate the membrane proteins with GalNAc.

Besides PrsA and PBP1a, MS analysis of the WT eluate revealed the presence of peptides derived from PknB and the cell division protein FtsQ. These proteins were predicted to possess the C-terminal Ser/Thr-rich IDRs (Fig. 2a, c, 4e and Supplementary Fig. 8). We confirmed glycosylation of PknB and FtsQ by comparing the electrophoretic mobilities of the proteins in the cell extracts of *S. mutans* WT, Xc∆*pgfS* and Xc∆*pgfS*:p*pgfS* using protein-specific polyclonal antibodies as we outlined for *S. pyogenes* and *S. pneumoniae* glycoproteins (Fig. 4f, g).

To validate the glycosylation of *S. mutans* PrsA and PBP1a and to identify the location of glycosites, we constructed the 3×FLAG-tagged PrsA and PBP1a variants at their native chromosomal loci in the WT and Xc∆*pgfS* backgrounds. We fused the tag with the C-termini of the full-lengths and the IDR-truncated variants of the proteins, creating PrsA-3×FLAG, PBP1a-3×FLAG, PrsA∆IDR-3×FLAG and PBP1a∆IDR-3×FLAG. Immunoblot analysis with the anti-FLAG antibodies revealed that deletion of *pgfS* increased electrophoretic mobilities of the full-length proteins but not the IDR-less variants (Fig. 4h, i). Taken together, these experiments conclude that PrsA, PBP1a, PknB, and FtsQ are glycosylated with GalNAc, and the sugars are attached to extracytoplasmic Ser/Thr-rich IDRs.

S. mutans **biofilm depends on glycosylation of the PrsA IDR**

Previous studies have shown that *S. mutans* exhibits defective biofilm formation due to the deletion of PrsA, PBP1a, PknB, or BrpA [50, 51, 52, 53]. To understand the role of protein glycosylation in *S. mutans*, we analyzed the biofilm biomasses of WT, Xc∆*pgfS*, and Xc∆*pgfS*:p*pgfS* grown overnight in a medium supplemented with sucrose or glucose. *S. mutans* grown in a sucrose medium produces an exopolysaccharide-based biofilm. However, when sucrose is replaced with glucose, the bacterium forms a protein-based biofilm mediated by protein–protein interactions of the surface proteins [54]. The biofilm assay showed that glycosylation is dispensable for exopolysaccharide- or protein-biofilm development in normal conditions (Supplementary Fig. 16). However, deletion of *pgfS*

significantly impaired biofilm formation when glucose medium was supplemented with 3.5% ethanol, which is known to cause proteotoxic stress (Fig. 5a).

To identify a glycoprotein linked to this phenotype, we constructed the PrsA, PknB, and BrpA mutants lacking C-terminal IDRs (*prsA*ΔIDR, *pknB*ΔIDR and *brpA*ΔIDR), and the ΔCBA mutant deficient in all four collagen-binding adhesins, WapA, CnaB, CbpA, Cnm, that are highly likely glycosylated by the Pgf pathway. Additionally, we analyzed the *pbp1a*ΔIDR-3×FLAG mutant which lacks the PBP1a IDR. Only *prsA*ΔIDR lacking the C-terminal 33 amino acid residues was biofilm-defective in the ethanol-stressed conditions (Fig. 5b). The *prsA*ΔIDR and Xc*Δpgf*S strains were not attenuated for growth in the presence of ethanol (Supplementary Fig. 17), indicating that a significant decline of the biofilm biomass of these mutants is not due to the change in cell growth. The isogenic *prsA* deletion mutant, Xc*Δprs*A, was also biofilm-defective in these conditions (Fig. 5c). However, this mutant displayed a growth defect (Supplementary Fig. 17), indicating that PrsA is necessary to support cell growth under ethanol stress. Furthermore, deletion of the C-terminal 8 or 23 amino acid residues of PrsA (*prsA*ΔIDR_8 and *prsA*ΔIDR_23, respectively) also resulted in biofilm deficiency (Fig. 5e). Based on these results, we concluded that the full-length PrsA IDR is critical for *S. mutans* biofilm development. Interestingly, a more prominent biofilm defect was observed when *pgfS* was deleted in the *prsA*ΔIDR_8, *prsA*ΔIDR_23, and *prsA*ΔIDR strains, implying that *O*-glycosylation maintains biofilm synthesis not only via the PrsA IDR (Fig. 5e).

It is noteworthy that extra 24 amino acid residues from 3×FLAG at the C-terminus of the full-length or the IDR-less variant of PrsA in the WT or Xc*Δpgf*S background did not influence biofilm formation compared to that of their original strains (Supplementary Fig. 18), indicating that a C-terminal 3×FLAG-tag disordered region does not functionally substitute the PrsA IDR. To understand a link between the amino acid composition of the PrsA IDR (hereafter IDR^{PrsA_Smu}) and biofilm formation, we replaced IDR^{PrsA_Smu} from its *S. mutans* chromosomal locus with a) Ser/Thr-rich region from WapA (hereafter IDR^{WapA}) which has been shown to be *O*-glycosylated [48], and b) the *S. pyogenes* PrsA1 IDR (hereafter IDR^{PrsA1_GAS}). IDR^{PrsA_Smu} shares a very low sequence homology with IDR^{WapA} and IDR^{PrsA1_GAS} (Fig. 5d). Strikingly, strains expressing the chimeras produced the

protein-based biofilm under ethanol stress. However, the deletion of *pgfS* in these strain backgrounds resulted in no biofilm (Fig. 5f). Taken together, these data suggest that PrsA function in biofilm development depends on glucans attached to Ser/Thr-rich IDRs.

Glycosylation protects the IDRs from proteolytic cleavage

Remarkably, examination of PrsA expression in *S. mutans* using the anti-PrsA polyclonal antibodies revealed that PrsA migrates as a broad band at around 43-47 kDa in WT and Xc∆*pgfS*:p*pgfS*, but as two bands at around 38 and 40 kDa in Xc∆*pgfS* (Fig. 6a, lanes 1-3). We hypothesized that the 40 kDa band (lane 2, red arrow) is a non-glycosylated form of PrsA and the 38 kDa band (lane 2, black circle) is a proteolytically processed form of PrsA which was not detected in the 3×FLAG-tagged strains by the anti-FLAG antibodies (Fig. 4h). Upon exposure of cells to 3.5% ethanol, heterogeneity of PrsA was significantly increased in WT and Xc∆*pgfS*:p*pgfS*. The protein migrated between 38 and 47 kDa (Fig. 6a, lanes 4 and 6). Bands corresponding to the non-glycosylated and proteolytically processed forms of PrsA were also detected in these cells, implying that ethanol inhibits glycosylation. These observations were further corroborated by the analysis of PrsA expression in the *S. mutans* strains harboring the PrsA-3×FLAG variant (Fig. 6b). Immunoblot analysis with the anti-FLAG antibodies demonstrated that the ethanol-subjected WT cells produced an additional low molecular weight band (Fig. 6b, lane 3) which co-migrates with the non-glycosylated form of PrsA-3×FLAG (Fig. 6b, lane 2, red arrow). When the expression of the PrsA-3×FLAG variant in the cells grown in normal conditions was assessed using anti-PrsA, the protein migrated as two bands in Xc∆*pgfS* (Fig. 6c, lane 2), consistent with the idea that the smallest fragment (black circle) is formed by proteolytic cleavage. Upon exposure of the cells to ethanol stress, the PrsA-3×FLAG cleavage product was also detected in WT (Fig. 6c, lane 3), which aligns with the hypothesis that ethanol inhibits protein glycosylation. Because the smallest fragment of the PrsA-3×FLAG variant is not detected by the anti-FLAG antibodies, it confirms that this fragment is a proteolytically processed form of PrsA lacking the 3×FLAG tag. These observations also argue that the cleavage site for an unknown protease is present on the PrsA IDR and *O*-glycosylation protects this region from proteolysis.

Further investigation of the expression of the PrsAΔIDR variant and two PrsA-IDR^{WapA} and PrsA-IDR^{PrsA1_GAS} chimeras using anti-PrsA revealed that PrsAΔIDR migrated as a single sharp band (Fig. 6d, lane 1), but the chimeras migrated as multiple smeared bands (Fig. 6d, lanes 3 and 5). The protein heterogeneity was especially remarkable for PrsA-IDR^{WapA} (Fig. 6d, lane 3). Importantly, in the absence of glycosylation, the electrophoretic mobility and expression of the PrsAΔIDR variant were not altered (Fig. 6d, lane 2), confirming the vulnerability of the PrsA IDR for proteolysis. In contrast, the non-glycosylated forms of both chimeras demonstrated multiple bands of greater mobility (Fig. 6d, lanes 4 and 6), suggesting a proteolytic cleavage of the IDRs. These data provide evidence that the chimeric IDRs are glycosylated and sensitive to proteolysis when *O*-glycans are removed. Considering these observations, we examined the expression of other *S. mutans* glycoproteins in the WT and Xc∆*pgfS* backgrounds. Additional lower-molecular weight PknB-specific bands were detected in Xc∆*pgfS*, suggesting the function of *O*-glycosylation in protecting PknB from proteolysis (Fig. 4f).

The efficiency of PrsA glycosylation and expression depends on the IDR length

We considered that the chaperone function of PrsA likely safeguards extracellular proteins from misfolding and aggregation in proteotoxic conditions when ethanol is present in the growth medium. Therefore, we investigated whether the PrsA IDR promotes folding and expression of PrsA. We first explored the possibility that the IDR affects the secondary structure and the overall thermal stability of PrsA. We expressed in *E. coli* two recombinant soluble variants of PrsA that lack the signal peptide: the full-length version, ePrsA-FL, and the IDR-less version, ePrsAΔIDR. Both variants displayed similar behavior in protein thermal stability assays conducted in a wide range of buffers and salts, including 3.5% ethanol (Supplementary Fig. 19a, b), indicating that the IDR does not affect the thermal stability of PrsA. Moreover, the circular dichroism spectra of ePrsA-FL and ePrsAΔIDR did not reveal any prominent structural changes in the secondary structure content between these variants (Supplementary Fig. 19c).

We next examined the possibility that the IDR facilitates folding and expression of PrsA in *S. mutans* periplasm. We quantified the expression of the FLAG-tagged PrsA variants in the WT and Xc∆*pgfS* strains revealing that the PrsA IDR enhances the protein

expression (Fig. 4h, Extended Data Fig. 3a). Notably, similar results were also obtained when we compared the expression of PBP1a in the PBP1a-3×FLAG and PBP1aΔIDR-3×FLAG strains using the anti-FLAG antibodies (Fig. 4i, and Extended Data Fig. 3b) and the expression of PknB in the *S. mutans* WT and *pknB*ΔIDR using the anti-PknB antibodies (Supplementary Fig. 20). These data together with the analysis of expression of *S. pneumoniae* PrsA, PBP1a and PBP1b (Extended Data Fig. 1) establish that several C-terminal IDRs work as the enhancers of protein expression.

To understand the biofilm-deficient phenotype of the *prsA*ΔIDR_8 and *prsA*ΔIDR_23 and *prsA*ΔIDR mutants, we compared the expression of the full-length PrsA, PrsAΔIDR_8, PrsAΔIDR_23, and PrsAΔIDR variants in the WT and XcΔ*pgfS* backgrounds using the anti-PrsA antibodies. Strikingly, PrsAΔIDR_8 and PrsAΔIDR_23 produced the major and minor bands of approximately 37 kDa in both strain backgrounds (Fig. 6e, lanes 3-6, black circle), indicating that even the removal of 8 amino acid residues from the IDR significantly inhibits the efficiency of glycosylation leading to increased proteolysis of PrsA. Moreover, the glycosylated form of PrsAΔIDR_8 is more smeared than the full-length PrsA, supporting the notion that this mutant has altered glycosylation (Fig. 6e, lanes 3). Quantification of expression of all PrsA species in the WT, *prsA*ΔIDR_8, *prsA*ΔIDR_23, and *prsA*ΔIDR strains revealed that shortening of the IDR progressively reduces PrsA expression (Extended Data Fig. 4). Despite a modest effect of the deletion of 8 or 23 amino acid residues on PrsA expression, these variants are defective in biofilm formation.

Based on the above results, we propose a model to explain the biofilm phenotype of these mutants (Fig. 6f). The C-terminal disordered region of PrsA might be necessary for the function of PrsA in preventing misfolding and aggregation of protein clients in proteotoxic stress conditions. Sugar decorations on the PrsA IDR protect this important region from degradation by an unknown protease and might also stimulate the chaperone function of PrsA. The optimal length of the IDR is essential for efficient glycosylation and expression of PrsA.

Discussion

Our glycobiological studies, in combination with genome-wide identification of IDRs, uncovered a unique class of IDRs linked to the extracytoplasmic proteins in streptococci (Supplementary Table 1). A distinct feature of these IDRs is the presence of poly-Ser/Thr residue tracks. In *S. pyogenes* and *S. pneumoniae*, we elucidated a novel Glc-type *O*-linked glycosylation mechanism in which GtrB works in pair with PgtC2 to transfer α-Glc to Ser/Thr-rich IDRs. Surprisingly, *S. mutans* does not utilize the GtrB/PgtC2 mechanism for IDR glycosylation. The glycosyltransferases of Pgf protein glycosylation pathway — PgfS and PgfM1 — transfer GalNAc to Ser/Thr-rich IDRs tethered to membrane proteins in *S. mutans* (Supplementary Fig. 21). The lectin-affinity glycoprotein enrichment strategy impartially revealed a strong correlation between this subset of IDRs and *O*-glycosylation. The presence of glycans on IDRs was further confirmed by comparing the differences in the electrophoretic mobilities of glycoproteins and their IDR-less variants in the WT and *O*-glycosylation-deficient backgrounds. Tight binding of *S. pyogenes* and *S. pneumoniae* glycoproteins to ConA lectin, as well as altered electrophoretic mobility of glycoproteins lacking the sugar decorations, argue that the numerous Ser/Thr residues on IDRs are heavily glycosylated, similar to mucin proteins produced by epithelial cells [55].

Strikingly, the identified IDRs tend to localize at the C-termini of membrane proteins involved in critical cellular processes such as protein folding, cell wall biosynthesis, cell division, and signal transduction. Among the proteins harboring this type of IDRs, we identified a conserved post-translational chaperone lipoprotein PrsA, class A penicillin-binding proteins PBP1a and PBP1b, LytR-CpsA-Psr glycopolymer transferase LCP involved in the attachment of polysaccharides to peptidoglycan, serine/threonine protein kinase PknB, the extracytoplasmic regulator of c-di-AMP synthesis CdaR, and a divisome protein FtsQ. Importantly, sequence alignments of identified glycoproteins with the bacterial orthologs provide evidence that this class of IDRs occurs universally in streptococci and is present in other Firmicutes, suggesting that the IDRs may modulate the respective functions of the folded domains.

To understand the biological function of these IDRs, we focused on *S. mutans*, a major etiological agent of human dental caries. Dental caries progression involves the attachment of *S. mutans* to the tooth surface and the development of virulent dental plaque. Surface proteins play a key role in these processes. We found that under ethanol-

stressed conditions, the formation of protein-based biofilm is dependent on protein glycosylation, and this phenotype is linked to the PrsA IDR. While the PrsA IDR is dispensable for *S. mutans* biofilm formation in the absence of stress, it becomes critical upon ethanol exposure. In Gram-positive bacteria, PrsA plays a crucial role in the export and maturation of membrane and secreted proteins, including numerous virulence factors [27, 50, 56, 57, 58, 59, 60]. The crystal structures of *B. subtilis* and *Listeria monocytogenes* PrsA homologs demonstrate that the enzymes possess two functional subdomains — the chaperone-like and the peptidyl-prolyl isomerase domain — suggesting that PrsA has a dual role in protein folding by providing a chaperone function to control the aggregation of protein clients and catalyzing the cis/trans isomerization of peptide bonds with proline residues [61, 62]. The chaperone function of PrsA is likely critically important in ethanol-stressed cells when misfolding of proteins is increased. By generating the PrsA variants in which the PrsA IDR is swapped with the IDRs from *S. mutans* WapA or *S. pyogenes* PrsA1, we determined that the C-terminal disorder and *O*-glycosylation are critical for the functionality of PrsA. It is currently unknown how this region contributes to PrsA activity. Interestingly, IDRs are abundant and functionally important in different folding chaperones including *E. coli* DnaK, GroEL, HtpG, Hsp33, and HdeA [63]. The disordered regions of some folding chaperones enable them to bind multiple clients preventing toxic aggregation. However, since our experiments with the swapped IDR variants reveal that the function of the PrsA IDR in the biofilm synthesis mechanism is *O*-glycosylation-specific rather than amino acid sequence-specific, if this region binds protein clients, the interaction should involve the sugar residues.

Another possible function of the IDR is to increase protein stability and folding of PrsA through steric hindrance acting as a so-called “entropic bristle” by entropically excluding large neighboring molecules [64, 65]. This property of some disordered proteins has been used to design the fusion tags to enhance the soluble expression of recombinant proteins [64, 66]. Our data demonstrating that the IDRs enhance the expression of *S. mutans* PrsA, PBP1a, PknB, and *S. pneumoniae* PBP1a support this idea. Furthermore, there is growing evidence that protein N-glycosylation in eukaryotes positively affects protein folding and stability by hindering hydrophobic aggregation-prone regions and reducing the conformational entropy of proteins [67, 68]. Thus, the dual effect of

disorder and glycosylation might substantially enhance the folding of adjacent domains. Because PrsA is a folding chaperone, it can be expected that the PrsA IDR also slows the aggregation of protein clients, thereby promoting their correct folding.

Our work also showed that the absence of *O*-glycosylation in *S. mutans* causes proteolytic degradation of the PrsA IDR, including the chimeric IDRs, supporting the idea that *O*-glycans shield proteins from membrane and extracellular proteases. The efficiency of PrsA glycosylation drops with the reduction of the IDR length, leading to increased degradation of the IDR and, as a result, to a biofilm formation defect. This effect of glycosylation is likely due to obstructing proteases through steric hindrance. Thus, *S. mutans* have evolved a sophisticated mechanism to regulate the function of PrsA by the C-terminal IDR, which carries diverse features, including the cleavage sites for an unknown protease, the *O*-glycosites, and the optimal length of the IDR. Furthermore, these features are not restricted to PrsA because *S. mutans* PknB lacking GalNAc decorations also displays increased proteolytic processing.

Another key finding of our study is that glycosylation of the PrsA IDR is suppressed in the ethanol-stressed *S. mutans* cells, probably because the last step of glycan transfer, which occurs in the periplasmic space, is highly vulnerable to small toxic molecules such as ethanol that can pass through the porous peptidoglycan. Interestingly, in *B. subtilis*, the major penicillin-binding protein PBP1 also harbors a C-terminal IDR required for efficient cell wall biosynthesis [7]. PBP1 has been detected as two isoforms, PBP1a and PBP1b, which are different by molecular weight but encoded by the same gene, *ponA* [69]. Intriguingly, an alternative form of PBP1 is formed by post-translational processing of the PBP1 IDR by an unknown mechanism [69]. Although the *B. subtilis* PBP1 IDR lacks the poly-Ser/Thr, it has Ser/Thr residues and a large number of asparagine residues, (Supplementary Fig. 12), implying that PBP1a is generated by *O*- or *N*-glycosylation of PBP1b. The rapid transition of PBP1a to a small isoform PBP1b was observed upon alkaline treatment of *B. subtilis* cells [70], raising the possibility that the protein glycosylation mechanism is repressed in these conditions. Thus, this example and our studies with *S. mutans* PrsA highlight the tunable nature of IDRs to external stimuli and pinpoint a direct mechanism Firmicutes use to modulate the production of extracytoplasmic proteins in response to specific environments. Future investigation of the C-terminal

extracytoplasmic IDRs identified in our study may provide fundamentally new insights into how these regions sense and respond to environmental changes to boost the survival of bacteria. Finally, current advances in deep neural network-based methods have presented the opportunity to systematically evaluate IDRs in other prokaryotes. We anticipate that these approaches may expand our observation to multiple bacterial species and lead to the discovery of novel functional motifs in the procaryotic IDRs.

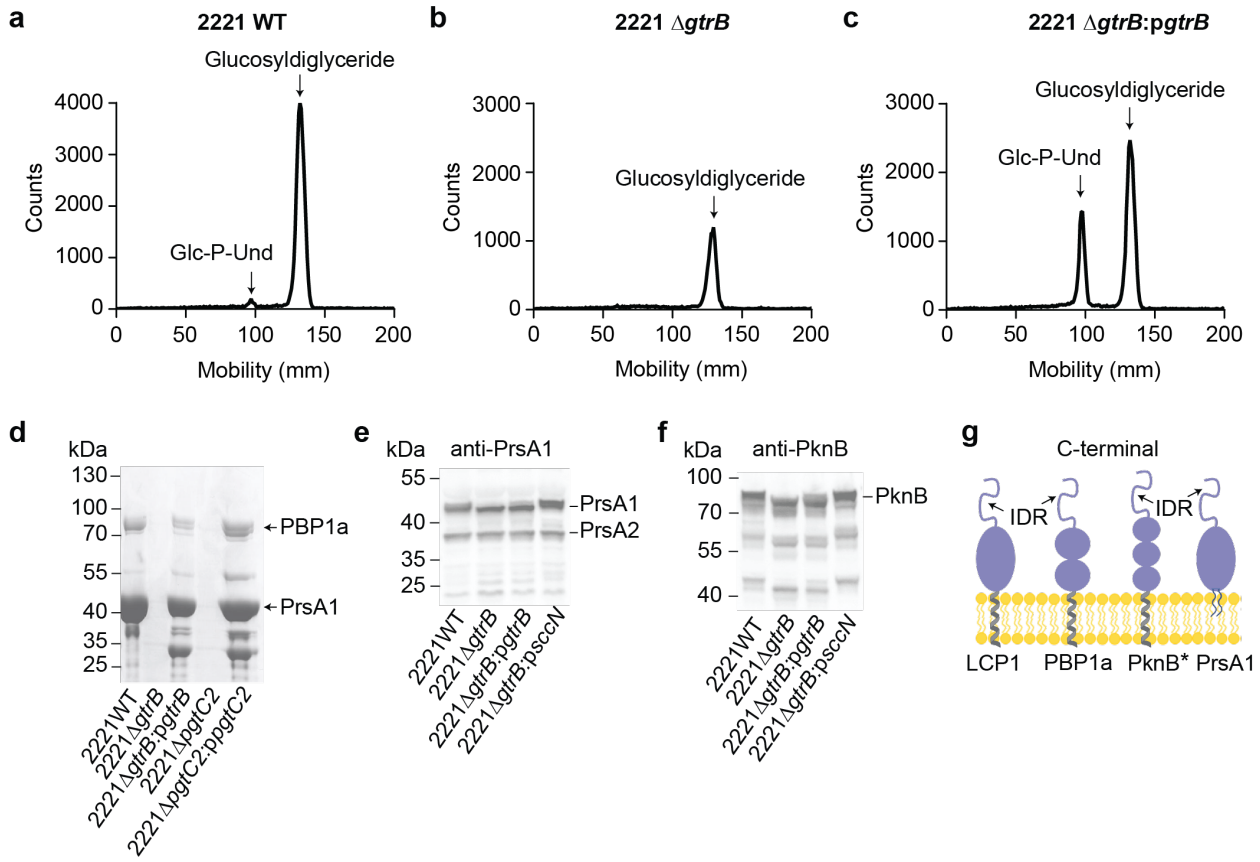

Fig. 1 | Glucosylation of *S. pyogenes* membrane proteins

a, **b**, **c**, Representative TLC analysis of [^{3}H]Glc-lipids extracted from *in vitro* incubations of UDP-[^{3}H]Glc with the membrane fractions isolated from *S pyogenes* WT (**a**), Δ*gtrB* (**b**), and Δ*gtrB*:p*gtrB* (**c**). **d**, Coomassie-stained gel of *S. pyogenes* membrane proteins purified by conA affinity chromatography. **e**, **f**, Immunoblot analysis of PrsA1 (**e**) and PknB (**f**) in *S. pyogenes* WT, Δ*gtrB*, Δ*gtrB*:p*gtrB*, and Δ*gtrB*:p*sccN* using specific antibodies. Anti-PrsA1 antibodies also recognize PrsA2. **g**, Topology of extracytoplasmic domains of *S. pyogenes* conA-bound proteins. * The cytoplasmic domain of PknB is omitted for clarity. The experiments were performed independently three times in **a**, **b**, **c**, **d**, **e**, and **f**, yielding the same results. A representative image from one experiment is shown.

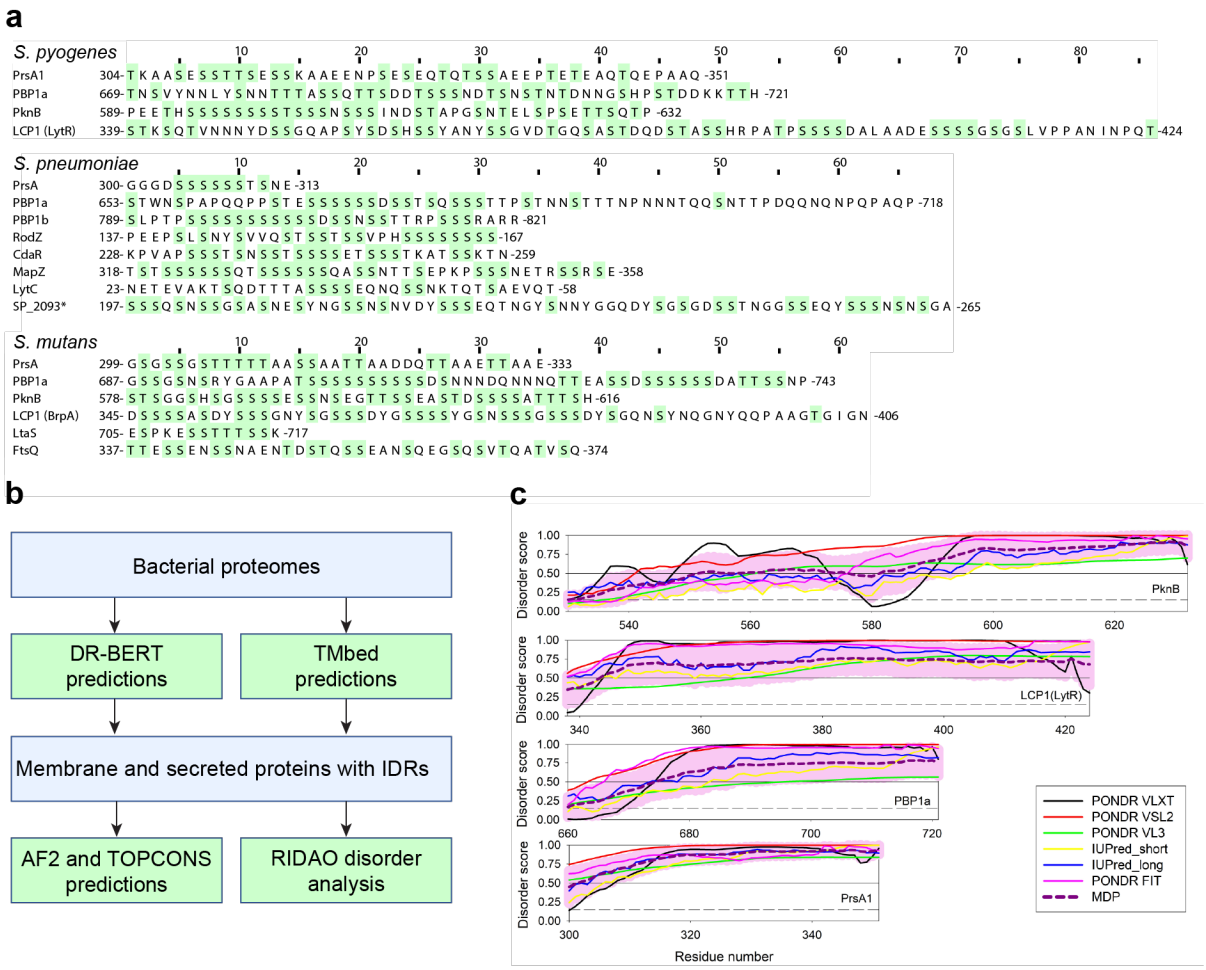

Fig. 2 | Features of Ser/Thr-rich IDRs identified in streptococcal extracytoplasmic proteins

a, IDR sequences of extracytoplasmic proteins. Ser and Thr residues are highlighted in green. Numbers correspond to amino acid residue starting and ending the IDR sequence. A * symbol denotes a TIGR4-specific protein. **b**, A flowchart outlining a genome-wide analysis of IDRs in streptococci. **c**, Evaluation of the intrinsic disorder status of the Ser/Thr regions in *S. pyogenes* proteins analyzed in this study. Plots represent the per-residue disorder profiles generated for the indicated proteins by the Rapid Intrinsic Disorder Analysis Online (RIDAO) platform [44]. RIDAO integrates six established disorder predictors, such as PONDR® VLS2, PONDR® VL3, PONDR® VLXT, PONDR® FIT, IUPred-Long, and IUPred-Short into a single, unified platform. The mean disorder profiles (MDP) for these proteins were calculated for each protein by averaging the disorder profiles of individual predictors. The light pink shade represents the MDP error

distribution. The thin black line at the disorder score of 0.5 is the threshold between order and disorder, where residues/regions above 0.5 are disordered, and residues/regions below 0.5 are ordered. The dashed line at the disorder score of 0.15 is the threshold between order and flexibility, where residues/regions above 0.15 are flexible, and residues/regions below 0.15 are mostly ordered. Analysis of *S. pneumoniae* and *S. mutans* proteins is shown in Supplementary Fig. 11.

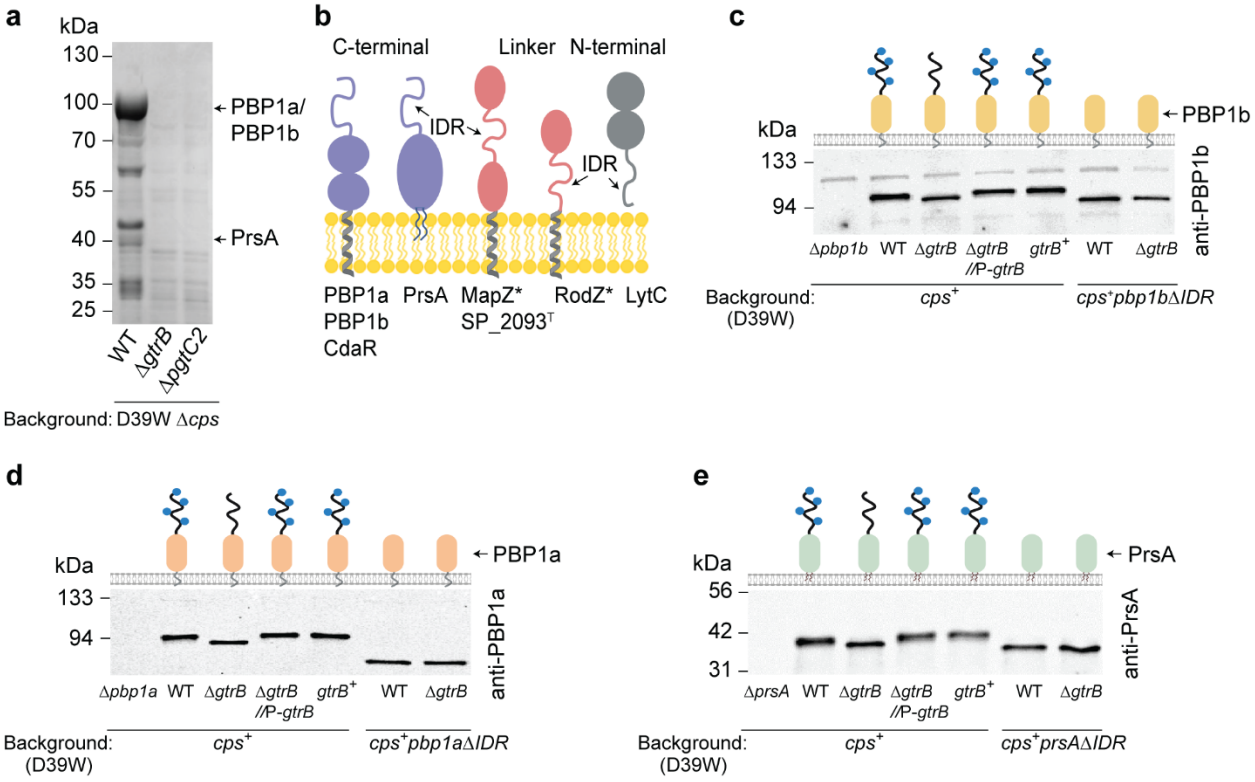

Fig. 3 | Glucosylation of *S. pneumoniae* extracytoplasmic proteins

a, Coomassie-stained gel of *S. pneumoniae* membrane proteins obtained from D39W Δ*cps* (IU1824), the isogenic Δ*gtrB* mutant (IU18454), and Δ*pgtC2* mutant (IU18454) purified by conA affinity chromatography. **b**, Topology of extracytoplasmic domains of *S. pneumoniae* conA-bound proteins. * The cytoplasmic domains of RodZ and MapZ are omitted for clarity. ^{T} denotes TIGR4 protein. **c**, **d**, **e**, Immunoblot analysis of PBP1b (**c**), PBP1a (**d**), and PrsA (**e**) in *S. pneumoniae cps*$^{+}$ WT (IU1781), and the isogenic strains, Δ*gtrB*, Δ*gtrB*//P-*gtrB*, *gtrB*$^{+}$, *pbp1b*ΔIDR, *pbp1b*ΔIDRΔ*gtrB,* *pbp1a*ΔIDR, *pbp1a*ΔIDRΔ*gtrB*, *prsA*ΔIDR, and *prsA*ΔIDR Δ*gtrB*. Samples obtained from *pbp1a*, *pbp1b,* and *prsA* deletion strains (left lanes) were used to demonstrate the specificity of the antibodies. In **c**, **d**, and **e**, schematic representations of PBP1b, PBP1a, and PrsA are shown (top panel). Each of the single blob of PBP1b and PBP1a represents two extracellular domains. Glc is depicted as blue circles. The experiments were performed independently two times in **a** and three times in **c**, **d**, and **e**. A representative image from one experiment is shown. Quantifications of protein loads and immunoblot band

intensities of PBP1a, PBP1b, and PrsA are provided in Supplementary Fig. 15c, d, e, and Extended Data Fig. 1.

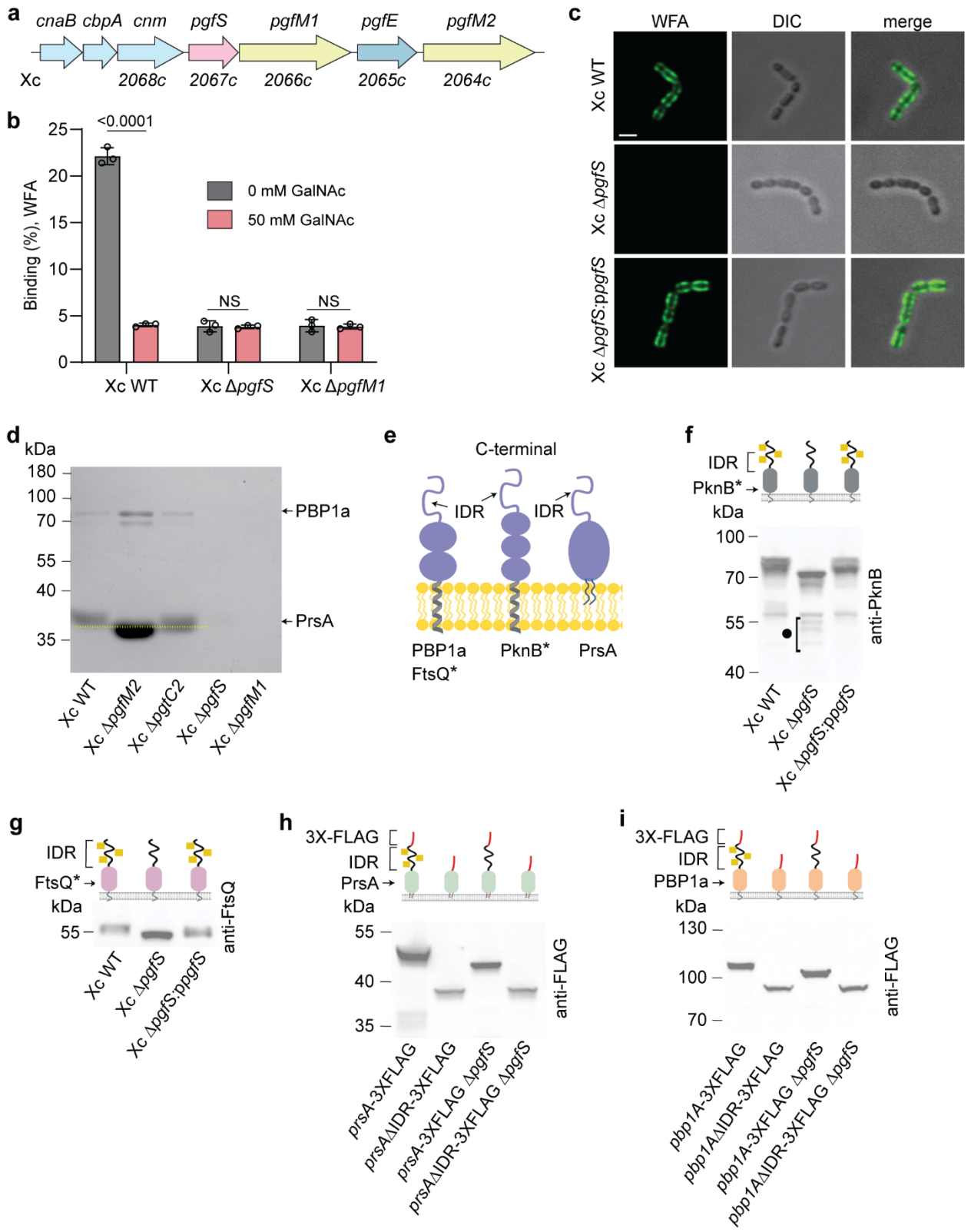

Fig. 4 | The *pgf* operon glycosylates *S. mutans* membrane proteins with GalNAc

a, Schematic representation of the *pgf* operon in *S. mutans* Xc. **b**, Co-sedimentation of FITC-conjugated WFA with the *S. mutans* strains. 0 or 50 mM GalNAc was added to the binding buffer. Data were recorded as a total fluorescence *vs* a pellet fluorescence. Columns and error bars represent the mean and S.D., respectively (n = 3 biologically independent replicates). A two-way ANOVA with Bonferroni's multiple comparisons test was performed to determine *P* values. **c**, Fluorescence and DIC images of bacteria bond to FITC-conjugated WFA. The scale bar is 1 μm. **d**, Coomassie-stained gel of *S. mutans* membrane proteins purified on the WFA-conjugated agarose column. **e**, Topology of extracytoplasmic domains of *S. mutans* WFA-bound proteins. * The cytoplasmic domains of PknB and FtsQ are omitted for clarity. **f**, **g**, Immunoblot analysis of PknB (**f**) and FtsQ (**g**) in *S. mutans* cell lysates using anti-PknB and anti-FtsQ antibodies. In **f** and **g**, schematic representations of PknB and FtsQ are shown (top panel). Each of the single blobs of PknB and FtsQ represents three- and two extracellular domains, respectively. * The cytoplasmic domains are not shown in the schematic. **h**, **i**, Immunoblot analysis of 3×FLAG-tagged variants of PrsA (**h**) and PBP1a (**i**) in the WT and Xc∆*pgfS* backgrounds examined by anti-FLAG antibodies. Schematic of PrsA and PBP1a variants is shown in **h** and **i** top panels, respectively. Each of the single blobs of PBP1a represents two extracellular domains. GalNAc is depicted as yellow squares. The experiments in **c**, **f**, **g**, **h,** and **i** were performed independently three times yielding the same results. A representative image from one experiment is shown. Quantifications of immunoblot band intensities of PrsA (**h**) and PBP1a (**i**) are provided in Extended Data Fig. 3a, b.

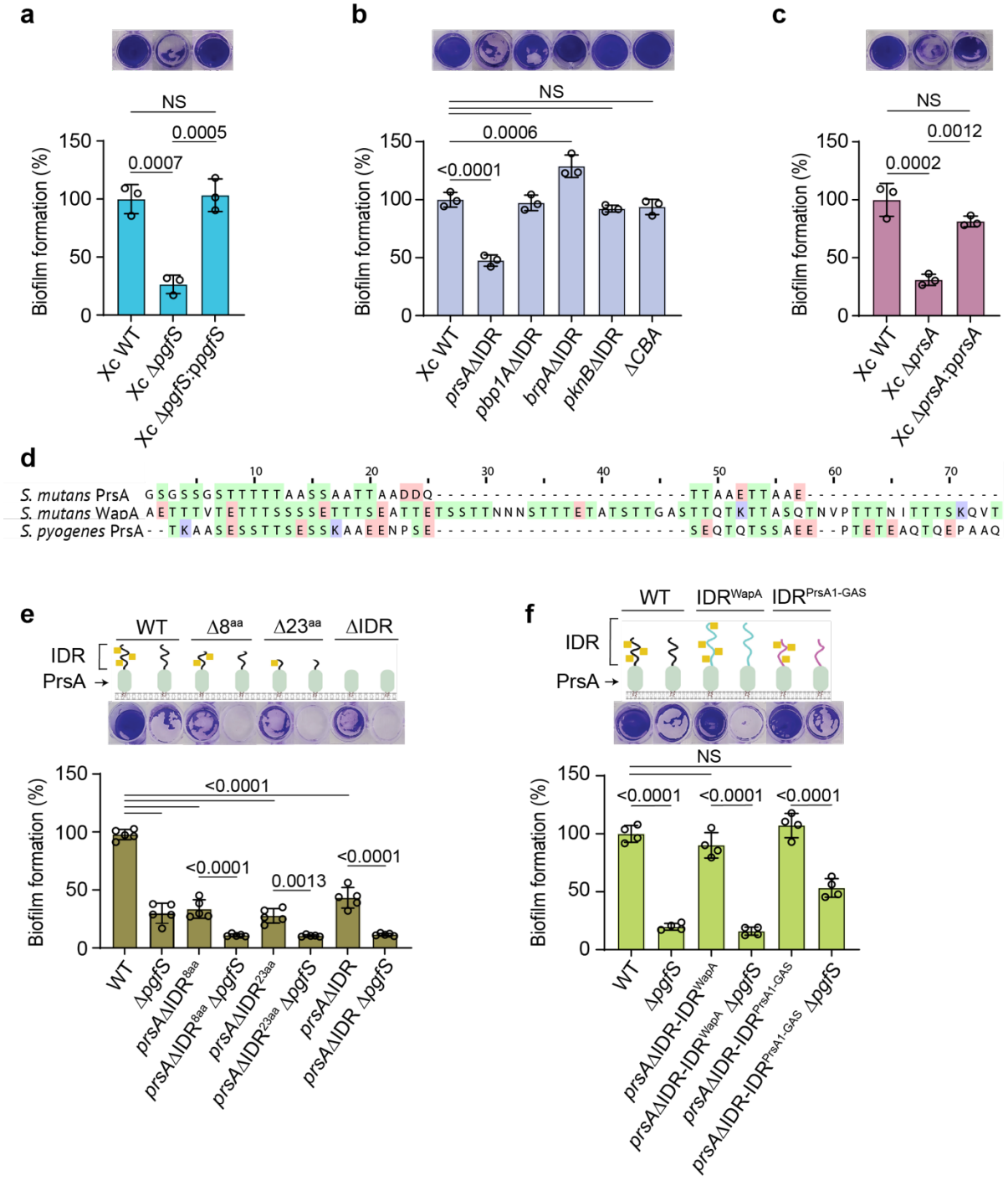

Fig. 5 | PrsA IDR regulates *S. mutans* biofilm formation under ethanol stress

a, Biofilm formation of *S. mutans* WT, Xc∆*pgfS,* and Xc∆*pgfS*:p*pgfS*. **b**, Biofilm formation of *S. mutans* strains expressing the IDR-less variants of PrsA, PknB, BrpA, PBP1a (*prsA*∆IDR, *pknB*∆IDR, *brpA*∆IDR and *pbp1a*∆IDR) or deficient in four cell wall-associated adhesins WapA, Cnm, CnaA and CnaB (∆CBA). The *pbp1a*∆IDR mutant is the

*pbp1a*Δ IDR-3×FLAG strain. **c**, Biofilm formation of *S. mutans* WT, XcΔ*prsA,* and XcΔ*prsA*:p*prsA*. **d**, A sequence alignment of IDRs of *S. mutans* PrsA and WapA, and *S. pyogenes* PrsA1. Ser and Thr residues are highlighted in green, Arg and Lys in blue, and Asp and Glu in red. **e**, Biofilm phenotype of *S. mutans* strains expressing PrsA with truncations in the C-terminal IDR. **f**, Biofilm formation of *S. mutans* strains expressing PrsA variants with the chimeric IDRs. In **e** and **f**, schematic representations of PrsA variants are shown (top panel). GalNAc is depicted as yellow squares. Representative biofilms stained with crystal violet are shown in **a**, **b**, **c**, **e**, and **f**. Quantifications of biofilm formation are shown in the bottom panels. Columns and error bars represent the mean and S.D., respectively (n = 3 in **a**, **b**, and **c;** n = 4 in **f** and n = 5 in **d** biologically independent replicates. *P* values were calculated by one-way ANOVA with Tukey's (**a**, **c**, **e**, **f**) and Dunnett's (**b**) multiple comparisons test.

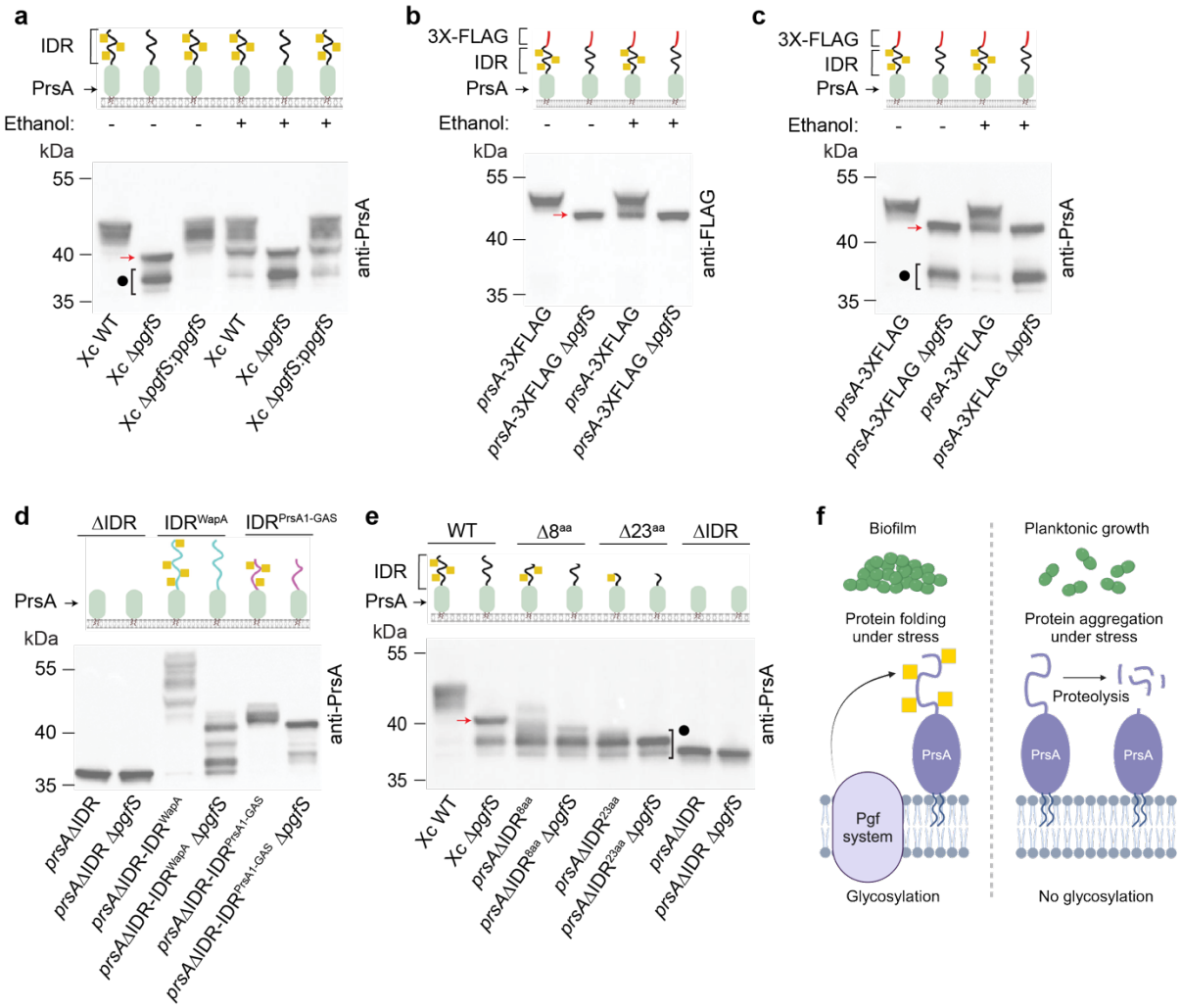

Fig. 6 | Biogenesis of PrsA is regulated by *O*-glycosylation of the IDR

a, Immunoblot analysis of PrsA in WT, Xc∆*pgfS,* and Xc∆*pgfS*:p*pgfS* grown in normal or ethanol-stressed conditions. PrsA is probed with anti-PrsA antibodies. **b**, **c**, Immunoblot analysis of 3×FLAG-tagged variant of PrsA in the WT and Xc∆*pgfS* backgrounds grown in normal or ethanol-stressed conditions. PrsA is probed with anti-FLAG antibodies in (**b**) and anti-PrsA antibodies in (**c**). **d**, **e**, Immunoblot analysis of PrsA variants harboring the chimeric IDRs (**d**) or the truncated IDRs (**e**) in the WT and Xc∆*pgfS* backgrounds grown in normal conditions. PrsA is probed with anti-PrsA antibodies. Schematic of PrsA variants is shown in **a**, **b**, **c**, **d**, and **e** top panels. GalNAc is depicted as yellow squares. In **a**, **b**, **c**, **d**, and **e**, the red arrow indicates a non-glycosylated form of PrsA and the black circle denotes the proteolytic products of PrsA. The experiments were performed independently three times in **a**, **b**, **d**, **e**, and two times in **c**, and they yielded the same

results. A representative image from one experiment is shown. Quantifications of immunoblot band intensities of PrsA carrying the truncated IDRs (**e**) are provided in Extended Data Fig. 4. **f**, The PrsA C-terminal IDR plays a functional role in *S. mutans* biofilm formation in ethanol-stressed cells. Pgf protein *O*-glycosylation machinery covalently attaches GalNAc moieties to the C-terminal IDR of a post-translocation secretion chaperone PrsA. This functional form of PrsA prevents misfolding of protein clients in ethanol-stressed conditions supporting the development of a protein-based biofilm. Loss of *O*-glycosylation exposes the IDR to proteolytic attack, resulting in a non-functional PrsA, which does not aid in biofilm formation.

Methods

Bacterial strains, growth conditions, and media

All plasmids, strains, and primers used in this study are listed in Supplementary Tables 2-6. *S. pyogenes* MGAS 2221*, S. mutans* Xc*,* and *S. pneumoniae* strains D39 and TIGR4 were used in this study. *S. pneumoniae* D39 strains were derived from unencapsulated strains IU1824 (D39 Δ*cps rpsL1*), which were derived from the encapsulated serotype-2 D39W progenitor strain IU1690 [71, 72]. Unless otherwise indicated, *S. pyogenes* and *S. mutans* were grown in BD Bacto Todd-Hewitt broth supplemented with 1% yeast extract (THY) without aeration at 37 °C. *S. pneumoniae* was cultured statically in Becton-Dickinson brain heart infusion (BHI) broth at 37 °C in an atmosphere of 5% CO_2, and growth was monitored by OD_{620} as described before [73]. In all experiments, cells were inoculated from frozen glycerol stocks into BHI broth, serially diluted, and incubated for 12–15 h. The next day, cultures at OD_{620} ~0.1–0.4 were diluted to OD_{620} ~0.005 in BHI broth. *S. pneumoniae* D39 cultures were sampled for western analysis at OD_{620} ~0.15–0.2 (early to mid-exponential phase). *E. coli* strains were grown in Lysogeny Broth (LB) medium or on LB agar plates at 37 °C. When required, antibiotics were included at the following concentrations: erythromycin at 500 μg mL^{-1} for *E. coli* and 5 μg mL^{-1} for streptococci; chloramphenicol at 10 μg mL^{-1} for *E. coli* and 4 μg mL^{-1} for streptococci; spectinomycin at 200 μg mL^{-1} for *E. coli* and 300 μg mL^{-1} for streptococci; kanamycin at 50 μg mL^{-1} for *E. coli* and 300 μg mL^{-1} for streptococci.

Construction of mutants in streptococci

To construct 2221Δ*gtrB*, we used a PCR overlapping mutagenesis approach, as previously described [21]. Briefly, 600-700 bp fragments both upstream and downstream of the gene of interest were amplified from *S. pyogenes* 2221 chromosomal DNA with designed primers that contained 16-20 bp extensions complementary to a nonpolar kanamycin resistance cassette which was PCR-amplified from pOSKAR [74]. Two fragments flanking the gene of interest and kanamycin cassette were purified using the QIAquick PCR purification kit (Qiagen) and fused by Gibson Assembly (SGA-DNA). A PCR was then performed on the Gibson Assembly sample using the primers listed in Supplementary Table 5 to amplify the fused fragments. The assembled DNA fragments

for mutagenesis of GAS *gtrB* were cloned into pMBsacB [75] to generate the plasmid pMBsacB-gtrB. The plasmid was electroporated into the GAS 2221 cells, and the transformants were selected on THY agar containing erythromycin. A two-step temperature-dependent selection process was used to isolate the double-crossover mutants that are sensitive to erythromycin but resistant to kanamycin [75]. The *S. mutans* mutants were constructed using a similar PCR overlapping mutagenesis approach. The primers used to amplify 600-700 bp fragments for Gibson Assembly are listed in Supplementary Table 5, and the antibiotic resistance cassettes are listed in Supplementary Table 2. The assembled DNA fragments were directly transformed into the *S. mutans* Xc cells by electroporation. The transformants were selected on THY agar containing the corresponding antibiotic. Double-crossover recombination in the *S. pyogenes* and *S. mutans* mutant clones was confirmed by PCR and Sanger sequencing using the primers listed in Supplementary Table 5. Some deletion mutants were verified by whole-genome sequencing (Supplementary Table 7). For complementation of *S. pyogenes* and *S. mutans* mutants, the promoterless genes were amplified from the corresponding bacterial strains using the primers listed in Supplementary Table 5. The PCR products were ligated into pDC123 vector. The resultant plasmids (Supplementary Table 2) were transformed into the corresponding mutants by electroporation. Chloramphenicol-resistant single colonies were picked and checked for the presence of the transformed gene using PCR and Sanger sequencing.

The *S. pneumoniae* mutants (Supplementary Table 3) were constructed using the background strains, primers and DNA templates listed in Supplementary Tables 4 and 6. *S. pneumoniae* strains containing antibiotic markers were constructed by transformation of CSP1-induced competent pneumococcal cells with linear DNA amplicons synthesized by overlapping fusion PCR [73, 76, 77]. Strains containing markerless alleles in native chromosomal loci were constructed using allele replacement via the P_c-[*kan-rpsL*$^+$] (Janus cassette) [78]. Bacteria were grown on plates containing trypticase soy agar II (modified; Becton-Dickinson), and 5% (vol/vol) defibrinated sheep blood (TSAII-BA). TSAII-BA plates used for selection were incubated at 37 °C in an atmosphere of 5% CO_2 and contained antibiotics at concentrations described previously [73, 77]. Ectopic expression *gtrB* for complementation was achieved with a constitutive P_{ftsA} promoter in the ectopic *bgaA*

site [79]. All constructs were confirmed by PCR and DNA sequencing of chromosomal regions corresponding to the amplicon region used for transformation.

Construction of the plasmids for protein expression in *E. coli* LOBSTR(DE3) cells

The genes encoding the extracellular domain of *S. mutans* PrsA (hereafter ePrsA-FL) and the IDR truncated version of the extracellular domains of *S. mutans* PrsA, FtsQ and PknB (hereafter ePrsAΔIDR, eFtsQΔIDR and ePknBΔIDR^{Smu}) were PCR-amplified from *S. mutans* Xc. The gene encoding the IDR truncated version of the extracellular domain of *S. pyogenes* PknB (hereafter ePknBΔIDR^{Spy}) was PCR-amplified from *S. pyogenes* 2221. The corresponding primers used for cloning are listed in Supplementary Table 5. The PCR products were digested by the corresponding restriction enzymes and subcloned into a pRSF-NT vector. The resultant plasmids (Supplementary Table 2) express genes fused at the N-terminus with a His-tag followed by a TEV protease recognition site. They were transferred into competent *E. coli* LOBSTR(DE3) (Kerafast)

Construction of the plasmid for expression of GtrB^{Spy} in *E. coli* JW2347

To create a vector for expression of GtrB^{Spy}, the gene was amplified from *S. pyogenes* 2221 chromosomal DNA using the primer pair listed in Supplementary Table 5. The PCR product was digested by the corresponding restriction enzymes and subcloned into a pBAD33 vector. The resultant plasmid, pBAD33_GtrB^{Spy} was transferred into competent *E. coli* JW2347 strain that has a deletion of the *gtrB* gene [20].

Overexpression and purification of recombinant proteins from *E. coli*

To purify recombinant proteins, cells were grown to exponential phase (OD_{600} ~0.6) at 37 °C in an LB medium supplemented with kanamycin and chloramphenicol. Protein overexpression was induced by adding 0.25 mM isopropyl β-*D*-1-thiogalactopyranoside (IPTG), and growth was continued for approximately 18 h at 18 °C. Cells were harvested by centrifugation and lysed in 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 1× bacterial protease inhibitor cocktail (Pierce Chemical Co.), 1 μg mL^{-1} DNase, and 25 μg mL^{-1} lysozyme using a microfluidizer cell disrupter. The soluble fractions were purified by Ni-NTA chromatography as previously reported for ePplD [80]. The eluate was dialyzed overnight

into 20 mM Tris-HCl pH 7.5 and 300 mM NaCl to remove imidazole. Protein concentration was determined by Pierce™ BCA protein assay kit (Thermo Fisher Scientific) and purity was estimated by SDS-PAGE.

Preparation of *S. pyogenes* and *E. coli* JW2347 membrane fractions to assess $GtrB^{Spy}$ activity

S. pyogenes strains grown in THY medium to an OD_{600} ~0.8 were collected by centrifugation at 18,000 g for 10 min. Cells were resuspended in 50 mM HEPES pH 7.4, 0.6 M sucrose, recombinant PlyC (50 μg mL^{-1}), and incubated at 37 °C for 1 hr. The cell suspensions were rapidly diluted with 100 mL ice-cold 10 mM HEPES pH 7.4, 2 mM $MgCl_2$, 1× bacterial protease inhibitor cocktail, 5 μg mL^{-1} DNAse, 5 μg mL^{-1} RNAse, and incubated on ice for 10 min. The suspension was homogenized with a Dounce homogenizer followed by adjustment of sucrose and EDTA in the suspension to a final concentration of 0.25 M and 10 mM, respectively. The suspension was centrifuged (2,500 g, 10 min, 4 °C) followed by centrifugation of the supernatant (40,000 g, 30 min, 4 °C). The resulting pellet that contained the membrane fraction was resuspended in 10 mM HEPES pH 7.4, 0.25 M sucrose, 1 mM EDTA.

For expression of $GtrB^{Spy}$, *E. coli* JW2347 cells carrying pBAD33_$GtrB^{Spy}$ plasmid were grown in LB at 37 °C, and induced with 13 mM L-arabinose at OD_{600} ~0.8, followed by 3 hours of growth at 25 °C. The cells were lysed in 20 mM Tris-HCl pH 7.5, 300 mM NaCl with two passes through a microfluidizer cell disrupter. The lysate was centrifuged (1000 g, 15 minutes, 4 °C), followed by centrifugation of the supernatant (40,000 g, 60 min, 4 °C) to isolate the membrane fraction.

***In vitro* analysis of Glc-P-Und synthase activity**

Reactions for the assay of Glc-P-Und synthase activity contained 50 mM HEPES pH 7.4, 20 mM $MgCl_2$, 1 mM DTT, 1 mM sodium orthovanadate, the indicated amounts of UDP-[^{3}H]Glc (50-500 cpm/pmol), Und-P, CHAPS and bacterial enzyme (10-25 μg membrane protein from either *S. pyogenes* or *E. coli*) in a total volume of 0.02 mL. Where indicated, Und-P was added as a sonicated dispersion in 1% CHAPS, with a final reaction concentration of 0.35% CHAPS. Following incubation at 37 °C for 10-20 min, reactions

were stopped by the addition of 2 mL $CHCl_3/CH_3OH$ (2:1). Reactions were freed of unincorporated radioactivity by Folch partitioning as described previously [17] and analyzed for radioactivity by scintillation spectrometry or by TLC on silica gel G developed in $CHCl_3/CH_3OH/H_2O/NH_4OH$ (65:30:4:1). Reactions with membrane fractions from *S. pyogenes* were subjected to mild alkaline de-acetylation with 0.1 N KOH in methanol/toluene (3:1) for 60 min at 0 °C, to destroy [^{3}H]glucosylglycerolipids and freed of released radioactivity by partitioning as described previously [17].

Co-sedimentation assay with fluorescent lectins

S. mutans cells grown overnight (~17 hours) were harvested by centrifugation and washed three times with 1× phosphate-buffered saline (PBS). The cell pellet was resuspended in 0.5 mL of PBS (OD_{600} ~ 9). 10-50 μg mL^{-1} of FITC-conjugated lectin from *Bandeiraea simplicifolia* (Sigma-Aldrich), Alexa Fluor™ 488-conjugated lectin from *Helix pomatia* (Thermo Fisher Scientific), or FITC-conjugated lectin from *Wisteria floribunda* (WFA) (Thermo Fisher Scientific) were added to bacterial suspension. The suspension was incubated at room temperature for 1 h with agitation. Sample aliquots were assayed to record the total fluorescence. The incubations were then centrifuged (16,000 g, 3 min), washed with PBS three times to remove unbound lectin, and resuspended in 0.5 mL of PBS. Sample aliquots were assayed to determine the pellet fluorescence. Data are presented as a percentage of the fluorescence of the pellet normalized to the total fluorescence of the sample.

Fluorescent and differential interference contrast (DIC) microscopy

Fluorescent and differential interference contrast (DIC) microscopy was performed as described previously [19]. Briefly, exponential phase cells (OD_{600} ~0.3) were fixed with paraformaldehyde (4% final concentration), pipetted onto poly-L-lysine coated microscope-slide-cover-glasses (high performance, D=0.17 mm, Zeiss), and allowed to settle for one hour at room temperature. Cells were then incubated with FITC-conjugated WFA (15.6 μg mL^{-1}) for 15 min at room temperature. The samples were dried at room temperature after four washes with PBS and subsequently mounted on a microscope slide with ProLong Glass Antifade Kit (Invitrogen). Samples were imaged on a Leica SP8

microscope with 100×, 1.44 N.A. objective and DIC optics. Images were processed with Huygens Professional software.

Isolation and purification of membrane glycoproteins

Streptococcal strains (2 L) grown in THY (OD_{600} ~0.8) were harvested by centrifugation at 18,000 g at 4 °C and stored at -80 °C until processing. Cell pellets were resuspended in 50 mL of lysis buffer containing 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 200 μg mL^{-1} mutanolysin, 250 μg mL^{-1} lysozyme, 1× protease inhibitor cocktail (Sigma-Aldrich) and 2 μg mL^{-1} DNase. To lyse *S. mutans* and *S. pyogenes*, recombinant 50 μg mL^{-1} zoocin A [81] and PlyC [82] were added to the lysis buffer, respectively. Suspension was homogenized and incubated for 1 h at room temperature. *S. mutans* and *S. pneumoniae* were lysed by bead beating (0.1 mm disruptor beads) followed by incubation of the homogenized samples at room temperature for 30 min on a spin rotator. The cell lysate was centrifuged at 5,000 g for 10 min at 4 °C to remove beads and cell debris. The resulting supernatant was centrifuged at 104,000 g for 1 h at 4 °C. The pellet was solubilized in 50 mL of the membrane-solubilization buffer (20 mM Tris-HCl pH 7.5, 300 mM NaCl, 1% Triton X-100, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 1 mM $MnCl_2$, and 1× bacterial protease inhibitor cocktail). The membrane proteins were separated by centrifugation at 104,000 g for 1 h at 4 °C and loaded onto a column prefilled with either 2 mL of agarose-bound WFA (Vector Laboratories) for *S. mutans* sample or 2 mL of sepharose-bound concanavalin A from *Canavalia ensiformis* (Millipore Sigma) for *S. pyogenes* and *S. pneumoniae* samples. The columns were washed with 10 volumes of the membrane-solubilization buffer. The proteins were eluted with 50 mM GalNAc from agarose-bound WFA and 300 mM methyl α-D-mannopyranoside from sepharose-bound concanavalin. The eluent was further concentrated to a final volume of 1 mL using an Amicon Ultra-10 mL (10,000 MWCO) spin filter followed by methanol-chloroform precipitation. The resulting protein pellet was resuspended in the SDS-PAGE loading buffer and separated by SDS-PAGE followed by Coomassie blue staining.

LC-MS/MS analysis of proteins

Proteins were identified by LC–MS/MS proteomics analysis by the Taplin Biological Mass Spectrometric Facility, Harvard Medical School, Boston, MA (Orbitrap mass spectrometers, Thermo Fisher Scientific). Briefly, Coomassie-stained protein bands were excised and subjected to a modified in-gel trypsin digestion procedure [83]. Peptides were reconstituted in 2.5% acetonitrile plus 0.1% formic acid, loaded *via* a Thermo EASY-LC (Thermo Fisher Scientific) and eluted with 90% acetonitrile plus 0.1% formic acid. Subsequently, peptides were subjected to electrospray ionization and loaded into either an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific) or a Velos Orbitrap Pro ion-trap mass spectrometer (Thermo Fisher Scientific). Peptide sequences were determined by matching protein databases with the acquired fragmentation pattern by Sequest (Thermo Fisher Scientific). All databases include a reversed version of all the sequences, and the data was filtered to between a one and two percent peptide false discovery rate.

SDS-PAGE and immunoblot analysis of *S. mutans* and *S. pyogenes* cell lysates

S. mutans and *S. pyogenes* grown to an OD_{600} ~0.5 were pelleted by centrifugation at 3,200 g for 15 min at 4 °C. The cell pellet was resuspended in 1 mL of lysis buffer containing 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1% Triton X-100, 1% SDS, 0.5% DOC, 1× bacterial protease inhibitor cocktail, and 2 μg mL^{-1} DNase. Cells were disrupted by bead beating for 1 h at room temperature and centrifuged at 16,000 g for 25 min. The total protein content in the supernatant fraction was determined by Pierce™ BCA protein assay kit (Thermo Fisher Scientific). An equivalent amount of total proteins was loaded in each lane and resolved on a 15% SDS-PAGE gel or 4-12% SurePAGE™ Bis-Tris gel (Genscript), transferred to nitrocellulose by electroblotting, and subjected to immunoblot analysis using as primary antibodies either the polyclonal produced in rabbits or anti-FLAG® antibody (Millipore Sigma). Rabbit polyclonal antibodies against the *S. pyogenes* PrsA1 and PrsA2 homologs were provided by Dr. Yung-Chi Chang [27]. Rabbit polyclonal antibodies against the *S. pyogenes* PknB homolog, and the *S. mutans* PrsA, FtsQ and PknB homologs were produced by Cocalico Biologicals. Purified recombinant PrsAΔIDR, FtsQΔIDR, $PknB\Delta IDR^{Smu}$, and $PknB\Delta IDR^{Spy}$ were used to produce the antibodies. Anti-sera were validated by immunoblot assay using streptococcal strains expressing the WT

and the mutants deficient in the corresponding tested protein. The validation is shown in Supplementary Fig 22.

SDS-PAGE and immunoblot analysis of *S. pneumoniae* cell lysates

Cell lysate preparations using SEDS lysis buffer [0.1% deoxycholate (vol/vol), 150 mM NaCl, 0.2% SDS (vol/vol), 15 mM EDTA pH 8.0] and immunoblotting was performed as previously described [42, 84]. Briefly, bacteria were grown exponentially in 5 ml BHI broth to an OD_{620} ~ 0.15–0.2. 1.8 mL of culture were centrifuged at 4 °C at 16,100 g for 5 min, and washed once with PBS at 4 °C. Pellets were frozen on dry ice for more than 15 min, then thawed, and resuspended in volumes of SEDS lysis buffer proportional to the harvest OD_{620} values. Frozen pellets collected from 1.8 mL of cultures at OD_{620} ~0.16 were suspended in 80 μL of SEDS lysis buffer followed by incubation at 37 °C for 15 min with shaking. Proteins were denatured by addition of 2× Laemmeli SDS loading buffer (Bio-Rad) plus β-mercaptoethanol (5% vol:vol) at 95 °C for 10 min. 8 μL of lysate were loaded in each lane and resolved on a 7.5% Mini-PROTEAN® TGX™ precast protein gel (Biorad 4561025) for PBP1a and PBP1b samples, or 10% precast protein gels (Biorad 4561035) for PrsA and StkP samples. For quantification, 4, 8, and 12 μL of WT samples were also loaded in additional lanes. The proteins were transferred to nitrocellulose, stained with Totalstain Q-NC reagent from Azure biosystems, and imaged as described by the manufacturer. Signal intensities from Totalstain Q-NC stain were used to normalize antibody signals. Rabbit polyclonal antibodies used were anti-StkP (1:7,000) [85, 86], anti-$PBP1a^{Spn}$ (1:7,000) [42], anti-$PBP1b^{Spn}$ (Dr. Cécile Morlot, 1:7,000), and anti-$PrsA^{Spn}$ (Dr. Marien de Jonge, 1:7,000). Anti-sera were validated by immunoblot assays using WT and the Δ*pbp1a*, Δ*pbp1b* or Δ*prsA* strains. The secondary antibody used was Licor IR Dye800 CW goat anti-rabbit (926–32,211, 1:14,000). IR Signal intensities were detected with Azure biosystem 600 and normalized with total protein stain signals for each lane using Totalstain Q-NC reagent. Measurements of band intensities were performed with AzureSpot software.

Biofilm assay

S. mutans cells grown overnight (17 h) were used as an inoculum at a 100-fold dilution in UFTYE (2.5% tryptone and 1.5% yeast extract containing) medium supplemented with 1% (wt/vol) sucrose or glucose. Bacterial biofilms were grown in 24-well plates (Corning) at 37 °C in the presence of 5% CO_2 for 24 h. For ethanol stress, 3.5% ethanol was added to the growth medium. After 24 h of growth, cell suspension in each well was removed, and biofilm was washed three times with water. The biofilm was stained with crystal violet (0.2%) followed by washes and resuspension in a de-staining solution (10% methanol, 7.5% acetic acid). Finally, the absorbance of the de-staining solution was measured at 540 nm.

Thermal protein stability assay

The protein thermal stability measurements were performed using Tycho NT.6 (NanoTemper Technologies). The purified protein samples were diluted in the test condition buffers to a range of concentrations from 1 to 2 mg mL^{-1}. The samples were heated from 35 to 95 °C at a rate of 30 °C min^{-1}. Protein unfolding was monitored by measuring the intrinsic fluorescence of tryptophan and tyrosine residues at emission wavelengths of 330 and 350 nm. The inflection temperature (Ti) was determined by calculating the maximum of the first derivative of the temperature-dependent change in fluorescence ratio F_{350}/F_{330}. The data were processed by Tycho NT.6 software.

Circular dichroism (CD)

Purified recombinant proteins ePrsA-FL and ePrsAΔIDR were buffer exchanged into 20 mM sodium phosphate pH 7.4, and 150 mM sodium fluoride. CD spectra were documented at a protein concentration of 0.80 mg/mL over a 200–250 nm wavelength range with a scan rate of 100 nm/min using a JASCO J-810 spectropolarimeter at 25 °C. The secondary structure content of ePrsA-FL and ePrsAΔIDR was calculated using the BeStSel server [87].

Prediction of extracytoplasmic IDRs

Predictions were run and aggregated using a custom Python script we developed. It takes as input and parses a genome in GenBank file format, extracting individual protein

sequences for forwarding to the predictors, handles switching conda environments, and includes logic for resuming incomplete runs. After both predictors are run, it concatenates the results as both a residue map and provides a list of every gene that contains predicted transmembrane regions, signal peptides, and disordered regions. Transmembrane regions and signal peptides were predicted using serial runs of the protein language model TMbed on the coding regions of *S. pyogens* MGAS2221 (GenBank: CP043530.1), *S. mutans* Xc (GenBank: PRJNA1071625), *S. pneumoniae* D39 (GenBank: CP000410.2) and *S. pneumoniae* TIGR4 (https://genomes.atcc.org/genomes/641decb736d446cc) with default batch size [34]. Disordered region prediction was performed using the compact protein language model DR-BERT using the pretrained weights [88]. The membrane topology of the subset of membrane and secreted proteins possessing IDRs (Supplementary Dataset 1) was predicted using TOPCONS [36]. Extracytoplasmic IDRs in the subset of membrane and secreted proteins were also validated by analysis of structural models generated by AlphaFold2 prediction pipeline [31]. Selected proteins were additionally analyzed using protein disorder predictors PONDR® VLS2, PONDR® VL3, PONDR® VLXT, PONDR® FIT, IUPred-Long, and IUPred-Short as implemented in the Rapid Intrinsic Disorder Analysis Online (RIDAO) platform [44].

Statistical analysis

Unless otherwise indicated, statistical analysis was performed using GraphPad Prism version 10 on pooled data from at least three independent biological experiments. A two-tailed Student's *t*-test and one-way ANOVA with Tukey's multiple comparisons test were performed as described for individual experiments. A *P*-value equal to or less than 0.05 was considered statistically significant.

Data availability

All data generated during this study are included in the article and supplementary information files.

Code availability

Code for automated running of TMbed and DR-BERT has been released on GitHub at: https://github.com/ctch225/ScanDis.

Acknowledgments

The authors thank Dr. Yung-Chi Chang (National Taiwan University) for the gift of anti-*S. pyogenes* PrsA1 and PrsA2 antibodies, Dr. Cécile Morlot (Institue de Biologie Structurale, Grenoble) for the gift of anti-*S. pneumoniae* PBP1b antibodies, Dr. Marien de Jonge (Radboud University Medical Center) for the gift of anti-*S. pneumoniae* PrsA antibodies and Dr. Catalina Velez-Ortega (University of Kentucky) for access to the Leica SP8 confocal microscope. We thank William de Souza and Joseph Mougous for their comments on the manuscript.

This work was supported by NIH grants R01DE028916 from the NIDCR, R21AI149366 from the NIAID (to NK), and R35GM131767 from the NIGMS (to MEW). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

MMR, SZ, CTC, JSR, H-CTT, MEW, KVK, and NK designed the experiments. MMR, SZ, CTC, JSR, CWK, YMH, H-CTT, KVK, and NK performed functional and biochemical experiments. SZ performed microscopy analysis. CTC developed code. VNU conducted bioinformatics analysis. H-CTT, SZ, KVK, and NK constructed streptococci and *E. coli* strains. MMR, SZ, CTC, JSR, H-CTT, MEW, KVK, and NK analyzed the data. NK wrote the manuscript with contributions from all authors. All authors reviewed the results and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Uversky VN. Functional unfoldomics: Roles of intrinsic disorder in protein (multi)functionality. *Adv Protein Chem Struct Biol* **138**, 179-210 (2024).

2. Schiano Lomoriello I, Sigismund S, Day KJ. Biophysics of endocytic vesicle formation: A focus on liquid-liquid phase separation. *Curr Opin Cell Biol* **75**, 102068 (2022).

3. Moses D, Ginell GM, Holehouse AS, Sukenik S. Intrinsically disordered regions are poised to act as sensors of cellular chemistry. *Trends Biochem Sci* **48**, 1019-1034 (2023).

4. Bondos SE, Dunker AK, Uversky VN. Intrinsically disordered proteins play diverse roles in cell signaling. *Cell Commun Signal* **20**, 20 (2022).

5. Peng Z*, et al.* Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci* **72**, 137-151 (2015).

6. Cermakova K, Hodges HC. Interaction modules that impart specificity to disordered protein. *Trends Biochem Sci* **48**, 477-490 (2023).

7. Brunet YR, Habib C, Brogan AP, Artzi L, Rudner DZ. Intrinsically disordered protein regions are required for cell wall homeostasis in Bacillus subtilis. *Genes Dev* **36**, 970-984 (2022).

8. El Rayes J*, et al.* Disorder is a critical component of lipoprotein sorting in Gram-negative bacteria. *Nature chemical biology* **17**, 1093-1100 (2021).

9. Boradia V, Frando A, Grundner C. The Mycobacterium tuberculosis PE15/PPE20 complex transports calcium across the outer membrane. *PLoS Biol* **20**, e3001906 (2022).

10. Lasker K*, et al.* The material properties of a bacterial-derived biomolecular condensate tune biological function in natural and synthetic systems. *Nat Commun* **13**, 5643 (2022).

11. Wang H*, et al.* Rubisco condensate formation by CcmM in beta-carboxysome biogenesis. *Nature* **566**, 131-135 (2019).

12. Wittwer M, Luo Q, Kaila VR, Dames SA. Oxidative Unfolding of the Rubredoxin Domain and the Natively Disordered N-terminal Region Regulate the Catalytic Activity of Mycobacterium tuberculosis Protein Kinase G. *J Biol Chem* **291**, 27062-27072 (2016).

13. Gardner KA, Moore DA, Erickson HP. The C-terminal linker of Escherichia coli FtsZ functions as an intrinsically disordered peptide. *Mol Microbiol* **89**, 264-275 (2013).

14. Mitrea DM, Kriwacki RW. Phase separation in biology; functional organization of a higher order. *Cell Commun Signal* **14**, 1 (2016).

15. Lemos JA*, et al.* The Biology of Streptococcus mutans. *Microbiology spectrum* **7**, (2019).

16. Ardiccioni C*, et al.* Structure of the polyisoprenyl-phosphate glycosyltransferase GtrB and insights into the mechanism of catalysis. *Nat Commun* **7**, 10175 (2016).

17. Rush JS*, et al.* The molecular mechanism of N-acetylglucosamine side-chain attachment to the Lancefield group A carbohydrate in *Streptococcus pyogenes*. *J Biol Chem* **292**, 19441-19457 (2017).

18. Rush JS*, et al.* Structure and mechanism of biosynthesis of Streptococcus mutans cell wall polysaccharide. *bioRxiv*, (2024).

19. Zamakhaeva S*, et al.* Modification of cell wall polysaccharide guides cell division in Streptococcus mutans. *Nature chemical biology* **17**, 878-887 (2021).

20. Baba T*, et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006 0008 (2006).

21. Edgar RJ*, et al.* Discovery of glycerol phosphate modification on streptococcal rhamnose polysaccharides. *Nature chemical biology* **15**, 463-471 (2019).

22. Schaffer C, Messner P. Emerging facets of prokaryotic glycosylation. *FEMS Microbiol Rev* **41**, 49-91 (2017).

23. Williams BA, Chervenak MC, Toone EJ. Energetics of lectin-carbohydrate binding. A microcalorimetric investigation of concanavalin A-oligomannoside complexation. *J Biol Chem* **267**, 22907-22911 (1992).

24. Mandal DK, Kishore N, Brewer CF. Thermodynamics of lectin-carbohydrate interactions. Titration microcalorimetry measurements of the binding of N-linked carbohydrates and ovalbumin to concanavalin A. *Biochemistry* **33**, 1149-1156 (1994).

25. Alexander JAN, Locher KP. Emerging structural insights into C-type glycosyltransferases. *Current opinion in structural biology* **79**, 102547 (2023).

26. Rismondo J, Percy MG, Grundling A. Discovery of genes required for lipoteichoic acid glycosylation predicts two distinct mechanisms for wall teichoic acid glycosylation. *J Biol Chem* **293**, 3293-3306 (2018).

27. Wu ZY*, et al.* Unique virulence role of post-translocational chaperone PrsA in shaping Streptococcus pyogenes secretome. *Virulence* **12**, 2633-2647 (2021).

28. van Hensbergen VP*, et al.* Streptococcal Lancefield polysaccharides are critical cell wall determinants for human Group IIA secreted phospholipase A2 to exert its bactericidal effects. *PLoS Pathog* **14**, e1007348 (2018).

29. Mikkat S, Kreutzer M, Patenge N. Dynamic Protein Phosphorylation in Streptococcus pyogenes during Growth, Stationary Phase, and Starvation. *Microorganisms* **12**, (2024).

30. Bugrysheva J, Froehlich BJ, Freiberg JA, Scott JR. Serine/threonine protein kinase Stk is required for virulence, stress response, and penicillin tolerance in Streptococcus pyogenes. *Infect Immun* **79**, 4201-4209 (2011).

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

32. Tunyasuvunakool K*, et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590-596 (2021).

33. Alderson TR, Pritisanac I, Kolaric D, Moses AM, Forman-Kay JD. Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. *Proc Natl Acad Sci U S A* **120**, e2304302120 (2023).

34. Bernhofer M, Rost B. TMbed: transmembrane proteins predicted through language model embeddings. *BMC bioinformatics* **23**, 326 (2022).

35. Nambiar A, Forsyth JM, Liu S, Maslov S. DR-BERT: A protein language model to annotate disordered regions. *Structure*, (2024).

36. Tsirigos KD, Peters C, Shu N, Kall L, Elofsson A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res* **43**, W401-407 (2015).

37. De A*, et al.* Deficiency of RgpG causes major defects in cell division and biofilm formation, and deficiency of LytR-CpsA-Psr family proteins leads to accumulation of cell wall antigens in culture medium by *Streptococcus mutans*. *Appl Environ Microbiol* **83**, (2017).

38. Gibhardt J*, et al.* An extracytoplasmic protein and a moonlighting enzyme modulate synthesis of c-di-AMP in Listeria monocytogenes. *Environ Microbiol* **22**, 2771-2791 (2020).

39. Fleurie A*, et al.* MapZ marks the division sites and positions FtsZ rings in Streptococcus pneumoniae. *Nature* **516**, 259-262 (2014).

40. Holeckova N*, et al.* LocZ is a new cell division protein involved in proper septum placement in Streptococcus pneumoniae. *MBio* **6**, e01700-01714 (2014).

41. Manuse S*, et al.* Structure-function analysis of the extracellular domain of the pneumococcal cell division site positioning protein MapZ. *Nat Commun* **7**, 12071 (2016).

42. Lamanna MM*, et al.* Roles of RodZ and class A PBP1b in the assembly and regulation of the peripheral peptidoglycan elongasome in ovoid-shaped cells of Streptococcus pneumoniae D39. *Mol Microbiol* **118**, 336-368 (2022).

43. Radler P, Loose M. A dynamic duo: Understanding the roles of FtsZ and FtsA for Escherichia coli cell division through in vitro approaches. *Eur J Cell Biol* **103**, 151380 (2024).

44. Dayhoff GW, 2nd, Uversky VN. Rapid prediction and analysis of protein intrinsic disorder. *Protein Sci* **31**, e4496 (2022).

45. Jorda J, Xue B, Uversky VN, Kajava AV. Protein tandem repeats - the more perfect, the less structured. *FEBS J* **277**, 2673-2682 (2010).

46. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190 (2004).

47. Guiral S, Mitchell TJ, Martin B, Claverys JP. Competence-programmed predation of noncompetent cells in the human pathogen Streptococcus pneumoniae: genetic requirements. *Proc Natl Acad Sci U S A* **102**, 8710-8715 (2005).

48. Andresen S*, et al.* Involvement of the Streptococcus mutans PgfE and GalE 4-epimerases in protein glycosylation, carbon metabolism, and cell division. *Glycobiology* **33**, 245-259 (2023).

49. Aviles-Reyes A*, et al.* Characterization of the pgf operon involved in the posttranslational modification of Streptococcus mutans surface proteins. *Scientific reports* **8**, 4705 (2018).

50. Hsu CC*, et al.* Streptococcus mutans PrsA mediates AtlA secretion contributing to extracellular DNA release and biofilm formation in the pathogenesis of infective endocarditis. *Virulence* **13**, 1379-1392 (2022).

51. Wen ZT, Bitoun JP, Liao S. PBP1a-deficiency causes major defects in cell division, growth and biofilm formation by Streptococcus mutans. *PLoS One* **10**, e0124319 (2015).

52. Bitoun JP, Liao S, Xie GG, Beatty WL, Wen ZT. Deficiency of BrpB causes major defects in cell division, stress responses and biofilm formation by Streptococcus mutans. *Microbiology* **160**, 67-78 (2014).

53. Banu LD, Conrads G, Rehrauer H, Hussain H, Allan E, van der Ploeg JR. The Streptococcus mutans serine/threonine kinase, PknB, regulates competence development, bacteriocin production, and cell wall metabolism. *Infect Immun* **78**, 2209-2220 (2010).

54. Zheng T*, et al.* Regulatory mechanisms of exopolysaccharide synthesis and biofilm formation in Streptococcus mutans. *J Oral Microbiol* **15**, 2225257 (2023).

55. Schjoldager KT, Narimatsu Y, Joshi HJ, Clausen H. Global view of human protein glycosylation pathways and functions. *Nat Rev Mol Cell Biol* **21**, 729-749 (2020).

56. Willett JLE, Robertson EB, Dunny GM. The Phosphatase Bph and Peptidyl-Prolyl Isomerase PrsA Are Required for Gelatinase Expression and Activity in Enterococcus faecalis. *J Bacteriol* **204**, e0012922 (2022).

57. Roch M, Lelong E, Panasenko OO, Sierra R, Renzoni A, Kelley WL. Thermosensitive PBP2a requires extracellular folding factors PrsA and HtrA1 for Staphylococcus aureus MRSA beta-lactam resistance. *Commun Biol* **2**, 417 (2019).

58. Hyyrylainen HL*, et al.* Penicillin-binding protein folding is dependent on the PrsA peptidyl-prolyl cis-trans isomerase in Bacillus subtilis. *Mol Microbiol* **77**, 108-127 (2010).

59. Williams RC*, et al.* Production of Bacillus anthracis protective antigen is dependent on the extracellular chaperone, PrsA. *J Biol Chem* **278**, 18056-18062 (2003).

60. Cahoon LA, Freitag NE. Listeria monocytogenes virulence factor secretion: don't leave the cell without a chaperone. *Frontiers in cellular and infection microbiology* **4**, 13 (2014).

61. Jakob RP*, et al.* Dimeric Structure of the Bacterial Extracellular Foldase PrsA. *J Biol Chem* **290**, 3278-3292 (2015).

62. Cahoon LA, Freitag NE, Prehna G. A structural comparison of Listeria monocytogenes protein chaperones PrsA1 and PrsA2 reveals molecular features required for virulence. *Mol Microbiol* **101**, 42-61 (2016).

63. Bardwell JC, Jakob U. Conditional disorder in chaperone action. *Trends Biochem Sci* **37**, 517-525 (2012).

64. Santner AA, Croy CH, Vasanwala FH, Uversky VN, Van YY, Dunker AK. Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression. *Biochemistry* **51**, 7250-7262 (2012).

65. Hoh JH. Functional protein domains from the thermally driven motion of polypeptide chains: a proposal. *Proteins* **32**, 223-228 (1998).

66. Tang NC*, et al.* Synthetic intrinsically disordered protein fusion tags that enhance protein solubility. *Nat Commun* **15**, 3727 (2024).

67. Duran-Romana R*, et al.* N-glycosylation as a eukaryotic protective mechanism against protein aggregation. *Sci Adv* **10**, eadk8173 (2024).

68. Hebert DN, Lamriben L, Powers ET, Kelly JW. The intrinsic and extrinsic effects of N-linked glycans on glycoproteostasis. *Nature chemical biology* **10**, 902-910 (2014).

69. Popham DL, Setlow P. Cloning, nucleotide sequence, and mutagenesis of the Bacillus subtilis ponA operon, which codes for penicillin-binding protein (PBP) 1 and a PBP-related factor. *J Bacteriol* **177**, 326-335 (1995).

70. Mitchell SL, Kearns DB, Carlson EE. Penicillin-binding protein redundancy in Bacillus subtilis enables growth during alkaline shock. *Appl Environ Microbiol* **90**, e0054823 (2024).

71. Lanie JA*, et al.* Genome sequence of Avery's virulent serotype 2 strain D39 of Streptococcus pneumoniae and comparison with that of unencapsulated laboratory strain R6. *Journal of bacteriology* **189**, 38-51 (2007).

72. Slager J, Aprianto R, Veening JW. Deep genome annotation of the opportunistic human pathogen Streptococcus pneumoniae D39. *Nucleic acids research* **46**, 9971-9989 (2018).

73. Tsui HC*, et al.* Suppression of a deletion mutation in the gene encoding essential PBP2b reveals a new lytic transglycosylase involved in peripheral peptidoglycan synthesis in Streptococcus pneumoniae D39. *Molecular microbiology* **100**, 1039-1065 (2016).

74. Le Breton Y, McIver KS. Genetic manipulation of *Streptococcus pyogenes* (the Group A Streptococcus, GAS). *Current protocols in microbiology* **30**, Unit 9D 3 (2013).

75. Hooven TA, Bonakdar M, Chamby AB, Ratner AJ. A Counterselectable Sucrose Sensitivity Marker Permits Efficient and Flexible Mutagenesis in Streptococcus agalactiae. *Appl Environ Microbiol* **85**, (2019).

76. Ramos-Montanez S*, et al.* Polymorphism and regulation of the spxB (pyruvate oxidase) virulence factor gene by a CBS-HotDog domain protein (SpxR) in serotype 2 Streptococcus pneumoniae. *Molecular microbiology* **67**, 729-746 (2008).

77. Tsui HT*, et al.* Pbp2x localizes separately from Pbp2b and other peptidoglycan synthesis proteins during later stages of cell division of Streptococcus pneumoniae D39. *Molecular microbiology* **94**, 21-40 (2014).

78. Sung CK, Li H, Claverys JP, Morrison DA. An rpsL cassette, janus, for gene replacement through negative selection in Streptococcus pneumoniae. *Applied and environmental microbiology* **67**, 5190-5196 (2001).

79. Zheng JJ, Perez AJ, Tsui HT, Massidda O, Winkler ME. Absence of the KhpA and KhpB (JAG/EloR) RNA-binding proteins suppresses the requirement for PBP2b by overproduction of FtsA in Streptococcus pneumoniae D39. *Molecular microbiology* **106**, 793-814 (2017).

80. Rush JS*, et al.* PplD is a de-N-acetylase of the cell wall linkage unit of streptococcal rhamnopolysaccharides. *Nat Commun* **13**, 590 (2022).

81. Gargis SR*, et al.* Zif, the zoocin A immunity factor, is a FemABX-like immunity protein with a novel mode of action. *Appl Environ Microbiol* **75**, 6205-6210 (2009).

82. Nelson D, Schuch R, Chahales P, Zhu S, Fischetti VA. PlyC: a multimeric bacteriophage lysin. *Proc Natl Acad Sci U S A* **103**, 10765-10770 (2006).

83. Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **68**, 850-858 (1996).

84. Cleverley RM*, et al.* The cell cycle regulator GpsB functions as cytosolic adaptor for multiple cell wall enzymes. *Nat Commun* **10**, 261 (2019).

85. Beilharz K, Novakova L, Fadda D, Branny P, Massidda O, Veening JW. Control of cell division in Streptococcus pneumoniae by the conserved Ser/Thr protein kinase StkP. *Proc Natl Acad Sci U S A* **109**, E905-913 (2012).

86. Rued BE*, et al.* Suppression and synthetic-lethal genetic relationships of DeltagpsB mutations indicate that GpsB mediates protein phosphorylation and penicillin-binding protein interactions in Streptococcus pneumoniae D39. *Mol Microbiol* **103**, 931-957 (2017).

87. Micsonai A*, et al.* BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res* **46**, W315-W322 (2018).

88. Nambiar A, Forsyth JM, Liu S, Maslov S. DR-BERT: A Protein Language Model to Annotate Disordered Regions. *bioarxiv*, (2023).