Disorder regulates homeostasis of extracytoplasmic 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], 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

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

Short title: Disorder regulates homeostasis of extracytoplasmic proteins in streptococci

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

Abstract

Proteins harboring intrinsically disordered regions (IDRs) that lack regular secondary or tertiary structure are abundant across three domains of life. Here, using a deep neural network (DNN)-based method we predict IDRs in the extracytoplasmic proteome of *Streptococcus mutans*, *Streptococcus pyogenes* and *Streptococcus pneumoniae*. We identify a subset of the serine/threonine-rich IDRs and demonstrate that they are *O*-glycosylated with glucose by a GtrB-like glucosyltransferase in *S. pyogenes* and *S. pneumoniae*, and N-acetylgalactosamine by a Pgf-dependent mechanism in *S. mutans*. Loss of glycosylation leads to a defect in biofilm formation under ethanol-stressed conditions in *S. mutans*. We link this phenotype to a C-terminal IDR of peptidyl-prolyl isomerase PrsA which is protected from proteolytic degradation by *O*-glycosylation. The IDR length attenuates the efficiency of glycosylation and expression of PrsA. Taken together, our data support a model in which extracytoplasmic IDRs function as dynamic switches 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 IDRs not only to serve as linkers connecting different structured domains, but also actively participate in binding to a broad range of molecular partners such as proteins, nucleic acids, small molecules, carbohydrates, lipids or metals. Moreover, some IDRs display the ability to sense the local chemical environment [1]. This feature of IDRs is due to their unique biophysical and structural characteristics and the presence of posttranslational modifications [2, 3]. Accordingly, IDRs are over-represented in proteins involved in cell signaling and regulatory processes such as transcription regulation, cell cycle and protein folding [2, 4].

IDRs are prevalent in eukaryotes where more than 30% of proteins are predicted to have long IDRs. In contrast, fewer proteins bear long IDRs in prokaryotes, and the functional features of these IDRs are poorly understood. Recent experimental evidence highlights the importance of prokaryotic IDRs in different aspects of cell physiology including signaling pathways, protein sorting mechanisms, cell division and the liquid-liquid phase separation [5, 6, 7, 8, 9, 10, 11, 12]. The current advances in deep neural network (DNN)-based methods have presented the opportunity to evaluate IDRs in different organisms. Here we took advantage of deep learning algorithms to identify with high accuracy the IDRs in the extracytoplasmic proteome of *Streptococcus mutans*, *Streptococcus pyogenes* and *Streptococcus pneumoniae*. These bacteria inhabit a variety of sites in the human body from the mouth and nasopharynx to skin causing various infections. Proteins that function in the external milieu play a crucial role in microbial growth and pathogenesis by maintaining cell envelope integrity, providing protection from various environmental stimuli, facilitating the uptake of nutrients and secretion of waste products, and participating in interactions with a host. We report that streptococci possess a subset of proteins harboring extracytoplasmic IDRs that are *O*-glycosylated. Functional characterization of the C-terminal IDR of peptidyl-prolyl isomerase PrsA in *S. mutans* reveals its critical function in biofilm formation under ethanol-stressed conditions. We show that PrsA expression is positively regulated by

the IDR and *O*-glycosylation which shields the IDR from proteolytic degradation. Our findings provide a roadmap for future functional studies of the extracytoplasmic IDRs in bacteria.

Results

Serine/threonine-rich IDRs in streptococcal extracytoplasmic proteins

To better understand the role of IDRs in prokaryotes, we used the deep learning algorithm DR-BERT to predict IDRs of at least 10 consecutive residues in the proteome of *S. pyogenes* MGAS2221, *S. mutans* Xc and *S. pneumoniae* TIGR4. Prediction of membrane and secreted proteins was conducted employing the deep learning-based model TMBed. A machine-learning protein structure prediction algorithm AlphaFold2 [13] was employed to validate the predictions. Membrane topology of predicted IDRs in the subset of membrane and secreted proteins was assigned by TOPCONS [14] (Supplementary Dataset 1). Visual inspection of predicted extracytoplasmic IDRs revealed that all analyzed streptococci possess a subset of IDRs containing extended regions enriched in serine (Ser) and threonine (Thr) residues (Fig. 1a). Among the proteins harboring this type of IDRs, we identified the homologs of a conserved post-translational chaperone lipoprotein with peptidyl-prolyl cis/trans isomerase activity PrsA [15, 16] and a class A penicillin-binding protein PBP1a in all streptococci; serine/threonine (Ser/Thr) protein kinase PknB [17] and LytR-Cps2A-Psr family protein (LCP1) involved in the attachment of polysaccharides to peptidoglycan in *S. pyogenes* (LCP1 is known as LytR [18]) and *S. mutans* (LCP1 is known as BrpA [19]); a class A penicillin binding protein PBP1b, the extracytoplasmic regulator of c-di-AMP synthesis CdaR [20], cell division proteins MapZ [21, 22, 23] and RodZ [24] in *S. pneumoniae*; and lipoteichoic acid synthase LtaS and a divisome protein FtsQ [25] in *S. mutans* (Supplementary Fig. 1, 2, 3, and Supplementary Table 1). Interestingly, Ser/Thr-rich IDRs are frequently located at the C-termini of the membrane-associated proteins. Furthermore, protein sequence alignments revealed the presence of Ser/Thr-rich IDRs in many homologs of identified proteins from streptococcal species (Supplementary Fig. 4). Sequence logos for each subset of IDRs were generated by WebLogo [26] to understand if the IDRs carry important functional features (Supplementary Fig. 5, 6). The predicted structural model of *S.*

mutans PrsA and the sequence logo of PrsA IDR are shown in Fig 1b, c. We observed that the main feature of the PrsA, PBP1a, LtaS, PBP1b, PknB and CdaR IDRs is the presence of large clusters of Ser and Thr residues adjacent to a folded domain suggesting the functional importance of this region (Supplementary Fig. 5).

Proteins with Ser/Thr-rich IDRs are glucosylated in *S. pyogenes*

To test a hypothesis that the extracytoplasmic Ser/Thr-rich IDRs are glycosylated, we focused on the identification of the enzymes involved in protein glycosylation. All sequenced genomes of *S. pyogenes* encode GtrB-type glycosyltransferase M5005_Spy_420 (hereafter $GtrB^{Spy}$) with an unknown function. These enzymes attach a sugar moiety to a lipid carrier, undecaprenyl phosphate (Und-P) which serves as a sugar donor for a wide range of acceptor molecules, including proteins [27, 28]. A BLAST search identifies that $GtrB^{Spy}$ homologs are universally present in lactic acid bacteria including *S. pneumoniae* (SPD_1431) and *S. mutans* (SccN) (Supplementary Fig. 7). In *S. mutans*, SccN is required for modification of the cell wall serotype *c* rhamnopolysaccharide, SCC, with the glucose (Glc) side-chains [29], suggesting that $GtrB^{Spy}$ is a glucosyltransferase. To determine $GtrB^{Spy}$ activity, we expressed the protein on a plasmid in the *E. coli* JW2347 background creating the *E. coli* pBAD33_$GtrB^{Spy}$ strain. *E. coli* JW2347 is devoid of glucosyl phosphoryl undecaprenol (Glc-P-Und) synthase activity [30]. $GtrB^{Spy}$ activity was assayed in *E. coli* membrane fractions *in vitro* using UDP-[^{3}H]Glc and exogenously added Und-P [28]. We observed the incorporation of radioactivity in the lipid fraction. A kinetic analysis demonstrated that $GtrB^{Spy}$ possesses ~ 8 μM and 51 μM affinity for UDP-Glc and Und-P, respectively, suggesting that the [^{3}H]glucolipid product is Glc-P-Und (Supplementary Fig. 8a, b). The TLC analysis of the lipids on silica gel detected a single peak of radioactivity (Supplementary Fig. 8c). No incorporation of radioactivity 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 $GtrB^{Spy}$-deficient mutant in *S. pyogenes* MGAS2221, 2221Δ*gtrB*, and 2221Δ*gtrB* complemented with the plasmid expressing the WT copy of $gtrB^{Spy}$, 2221Δ*gtrB*:p$gtrB^{Spy}$. When the membranes of *S. pyogenes* wild type (WT), 2221Δ*gtrB*, and 2221Δ*gtrB*:p$gtrB^{Spy}$ were incubated with UDP-[^{3}H]Glc and Und-P,

two classes of glucolipids were detected in WT and the complemented strain: a major glucolipid product with properties consistent with a glucosyldiglyceride and a minor glucolipid, presumably Glc-P-Und, which co-migrates on TLC with the lipid accumulating in the *E. coli* pBAD33_GtrB^{Spy} membranes. A major glucolipid product was also detected in the 2221Δ*gtrB* (Supplementary Fig. 9a, b, c). When glucosyldiglyceride was eliminated by mild alkaline de-acylation, the WT produced a single peak of radioactivity corresponding to Glc-P-Und (Fig. 2a). The membranes of 2221Δ*gtrB* did not catalyze the synthesis of Glc-P-Und (Fig. 2b). The complementation of the mutant with the plasmid resulted in a robust increase in Glc-P-Und synthase activity (Fig. 2c). The detected lipid was stable to mild alkaline methanolysis, but sensitivity to mild acid (0.1 N HCl, 50 °C, 50% isopropanol) [28], consistent with its identification as Glc-P-Und. Thus, these results implicate GtrB^{Spy} in the synthesis of Glc-P-Und.

To test a hypothesis that GtrB^{Spy} is required for glucosylation of the Ser/Th-rich IDRs, we used sepharose-bound concanavalin A (ConA), which recognizes α-mannose/α-glucose, to purify glycoproteins from the membranes of WT, 2221Δ*gtrB* and 2221Δ*gtrB*:p*gtrB*^{Spy}. The bound putative glycoproteins were eluted by the hapten sugar, methyl α-D-mannopyranoside. The SDS-PAGE analysis in Fig. 2d shows that the glycoproteins are present in WT and 2221Δ*gtrB*:p*gtrB*^{Spy}, but not in 2221Δ*gtrB*. Mass spectrometry (MS)-based proteomic analysis of the gel bands identified PrsA1 [15] and PBP1a as the most abundant proteins in the WT eluate. Additionally, the analysis detected LCP1, PknB and a homolog of PrsA1, PrsA2. Among the identified proteins, PrsA1, PBP1a, PknB and LCP1 possess the C-terminal Ser/Thr-rich IDRs (Fig. 1a and 2e). To validate the results of a lectin enrichment, we evaluated the expression of PrsA1, PrsA2 and PknB in WT, 2221Δ*gtrB* and 2221Δ*gtrB*:p*gtrB*^{Spy} using the specific polyclonal antibodies. Deletion of *gtrB*^{Spy} resulted in increased electrophoretic mobility of PrsA1 and PknB (Fig. 2f, g), but not PrsA2 (Fig. 2f and Supplementary Fig. 10). This electrophoretic behavior of PrsA1 and PknB in the mutant is consistent with an alteration of molecular weight due to the loss of the sugars. Interestingly, expression of the WT copy of *gtrB*^{Spy} in 2221Δ*gtrB* did not completely restore the electrophoretic behavior of PrsA1 and PknB (Fig. 2f, g). To support the hypothesis that the observed change in the molecular weight of PrsA1 and PknB in 2221Δ*gtrB* is due to the absence

of Glc, provided by Glc-P-Und, we expressed the *S. mutans sccN*, a Glc-P-Und synthase which we have previously characterized [29], on a plasmid in 2221Δ*gtrB* generating 2221Δ*gtrB*:p*sccN*. Expression of *sccN* restored the electrophoretic mobility of PrsA1 and PknB to that observed in the WT strain (Fig. 2f, g). Hence, GtrB participates in the attachment of Glc to *S. pyogenes* extracytoplasmic proteins possessing the Ser/Thr-rich IDRs.

Ser/Thr-rich IDRs are glucosylated in *S. pneumoniae*

To identify membrane glycoproteins in the *S. pneumoniae* serotype 2 D39W derivative (D39Δ*cps*; IU1824) and the *S. pneumoniae* serotype 4 TIGR4 strain, we used enrichment of glycoproteins on ConA-bound sepharose followed by proteomic analysis of protein bands sliced from the SDS-PAGE gel (Fig. 3a and Supplementary Fig. 11a). MS analysis revealed five putative glycoproteins in both strains: PrsA, PBP1a, MapZ, RodZ and CdaR. Additionally, we identified PBP1b and peptidoglycan hydrolase LytC [31] in D39Δ*cps*, and membrane protein SP_2093 in TIGR4. The C-terminal IDRs are present in PrsA, PBP1a, PBP1b and CdaR. MapZ, RodZ and SP_2093 possess the linker IDRs and LytC has the N-terminal IDR (Fig. 1a, 3b and Supplementary Fig. 2). The glycoproteins were absent in the eluate fractions of the isogenic Δ*gtrB* mutant (D39Δ*cps*Δ*gtrB*; IU18454), indicating the function of $GtrB^{Spn}$ in protein glycosylation (Fig. 3a). To validate the lectin-affinity enrichment, the electrophoretic behavior of PBP1b, PBP1a and PrsA 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) (Fig. 3c, d, e). The proteins demonstrated increased electrophoretic mobility in the cps^+Δ*gtrB* mutant, indicating a loss of sugar moieties. In contrast, no differences in the electrophoretic behavior of the PBP1b, PBP1a and PrsA variants lacking their 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). Interestingly, contrary to *S. pyogenes* PknB, the *S. pneumoniae* PknB homolog, known

as StkP, is not a ConA-bound protein, and it lacks the C-terminal IDR. Deletion of GtrB^{Spn} also did not change the electrophoretic mobility of StkP (Supplementary Fig. 11e), suggesting the absence of glycans attached to the protein.

Additionally, the quantitative analysis of PBP1b, PBP1a and PrsA expression in the *cps*$^{+}$ WT, *cps*$^{+}$Δ*gtrB, gtrB*$^{+}$ and Δ*gtrB*//P-*gtrB* cells did not reveal the effect of glycosylation on protein expression. Truncation of the PBP1b and PrsA IDRs also did not impact the production of these proteins. In contrast, the expression of the IDR-less PBP1a was significantly reduced in comparison to the full-length PBP1a, indicating the importance of the PBP1a IDR for the production of this protein (Supplementary Fig. 12a, b, c).

Pgf pathway glycosylates PrsA, PBP1a, FtsQ and PknB with GalNAc in *S. mutans*

Contrary to *S. pyogenes* and *S. pneumoniae*, no ConA-bound proteins were recovered from solubilized extracts of the *S. mutans* membranes. However, *S. mutans* possesses a protein glycosylation machinery that decorates the cell wall-associated collagen-binding adhesins WapA and Cnm with *O*-N-acetylhexosamines, presumably N-acetylgalactosamine (GalNAc) and N-acetylglucosamine (GlcNAc) [32, 33]. *O*-Glycosylation of WapA and Cnm requires the participation of the Pgf gene cluster encoded enzymes: the GtrB-type glycosyltransferase PgfS, an epimerase PgfE, and two GT-C fold glycosyltransferases PgfM1 and PgfM2. The gene cluster also encodes Cnm and Ser/Thr-rich surface proteins CnaB and CbpA which might be glycosylated by the Pgf pathway (Fig. 4a) [33]. Fluorescent microscopy of *S. mutans* revealed that the GalNAc-specific lectin, Wisteria Floribunda lectin (WFA), binds to *S. mutans* Xc cells, but not to its isogenic Δ*pgfS* mutant (Fig. 4b). Additionally, we used a co-sedimentation assay that exploits the property of WFA to associate with *S. mutans* strains. As expected, WFA binds to *S. mutans* WT, and also to Δ*pgfM2* and Δ*pgtM3* (the mutant deleted in an additional GT-C fold glycosyltransferase present in the *S. mutans* genome) but not to Δ*pgfS*, Δ*pgfM1* and the mutant lacking the adhesins encoded by the Pgf gene cluster, Δ*cnaB-cbpA-cnm* (Supplementary Fig. 13a). *S. mutans* WT cells, but not the Δ*pgfS*, Δ*pgfM1* and *S. pyogenes* cells (used as a negative control) also associate with other GalNAc-specific lectins, Helix pomatia agglutinin (HPA)

(Supplementary Fig. 13b) and Griffonia simplicifolia Lectin (BS-I) (Supplementary Fig. 13c). Binding of WFA to the WT cells was completely inhibited by the hapten sugar, GalNAc (Fig. 4c). In contrast to WFA, binding of the GlcNAc-specific lectin, wheat-germ agglutinin (WGA), to the WT and Δ*pgfS* cells were not significantly affected by the addition of its hapten sugar, GlcNAc, in the buffer (Supplementary Fig. 13d). Thus, this analysis strongly suggests that the enzymes of the Pgf pathway transfer GalNAc to cell wall-associated collagen-binding adhesins encoded by *cnaB-cbpA-cnm*.

To identify *S. mutans* membrane glycoproteins in an unbiased manner, we purified proteins from the detergent-solubilized *S. mutans* membranes on WFA agarose and eluted with GalNAc. MS analysis identified two major WFA-binding proteins, PrsA and PBP1a in the membranes of WT, Δ*pgfM2* and Δ*pgtM3*, but not Δ*pgfS* and Δ*pgfM1* (Fig. 4d). Interestingly, PrsA demonstrated increased electrophoretic mobility in Δ*pgfM2* suggesting a role of PgfM2 in *O*-GalNAc attachment to PrsA (Fig. 4d). Additionally, the proteomic analysis of the WT eluate revealed the presence of peptides derived from PknB and cell division protein FtsQ (Fig. 4e). The C-terminal IDRs of PrsA, PBP1a and PknB harbor poly-Ser/Thr regions (Fig. 1a). However, the C-terminal IDR of FtsQ lacks the poly-Ser/Thr, but it is Ser/Thr-rich (Fig. 1a). We validated the glycosylation of PknB and FtsQ by comparing the electrophoretic mobilities of the proteins in the cell extracts of *S. mutans* WT, Δ*pgfS* and the complemented strain (Δ*pgfS*:p*pgfS*) using the specific polyclonal antibodies as we outlined for *S. pyogenes* and *S. pneumoniae* glycoproteins (Fig. 4f, g). Importantly, loss of protein glycosylation resulted in the appearance of the additional PknB-specific bands suggesting that *O*-glycosylation protects PknB from proteolysis (Fig. 4f). Thus, the Pgf-encoded enzymes PgfS and PgfM1 *O*-glycosylate specific membrane proteins with GalNAc.

S. mutans **biofilm development depends on the PrsA IDR**

When *S. mutans* grows on sucrose as a carbon source, it 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 [34]. To understand the role of protein glycosylation in *S. mutans*, we analyzed the development of the exopolysaccharide- and protein-based biofilms in WT,

Δ*pgfS* and Δ*pgfS*: p*pgfS* grown in normal conditions and the presence of 3.5% ethanol which causes proteotoxic stress. Loss of protein glycosylation significantly impaired the protein-based biofilm in cells subjected to ethanol stress (Fig. 5a). To identify a glycoprotein linked to this phenotype, we constructed the PrsA, PknB and LCP1 (known as BrpA) mutants lacking the C-terminal IDRs (*prsA*ΔIDR, *pknB*ΔIDR and *brpA*ΔIDR), the PBP1a mutant in which the C-terminal IDR was replaced with 3XFLAG (*pbp1a*ΔIDR-3XFLAG), and the mutant deficient in four cell wall-associated adhesins WapA, Cnm, CnaA and CnaB that are glycosylated by the Pgf pathway (Δ*wapA*Δ*cnaB-cbpA-cnm*). Only *prsA*ΔIDR which lacks the C-terminal 33 amino acid residues was biofilm-defective in the ethanol-stressed conditions (Fig. 5b). Importantly, the biofilm phenotype of the isogenic Δ*prsA* mutant was identical to Δ*pgfS* and *prsA*ΔIDR, and the Δ*prsA* phenotype was reversed by the complementation with *prsA* on an expression plasmid (Fig. 5c). Deletion of *pgfS* in the *prsA*ΔIDR background did not change the biofilm deficient phenotype of the mutant (Fig. 5d). Furthermore, deletion of the last 8 or 23 amino acid residues of PrsA (*prsA*ΔIDR_8 and *prsA*ΔIDR_23, respectively) in the WT and Δ*pgfS* backgrounds resulted in biofilm deficiency (Fig. 5d). In contrast, the addition of 3XFLAG at the C-terminus of PrsA in the WT and the Δ*pgfS* backgrounds did not affect biofilm phenotypes of the WT and Δ*pgfS* strains (Supplementary Fig. 14). Hence, the full-length PrsA IDR (hereafter IDR^{PrsA_Smu}) is critical for *S. mutans* biofilm development.

To understand a link between the amino acid composition of IDR^{PrsA_Smu} and biofilm formation, we replaced IDR^{PrsA_Smu} from its *S. mutans* chromosomal locus with a) the Ser/Thr-rich region from WapA (hereafter IDR^{WapA}) which has been shown to be *O*-glycosylated [32], 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} (Supplementary Fig. 15). Strikingly, strains expressing the chimeras produced the protein-based biofilms under ethanol stress. However, the deletion of *pgfS* in these strain backgrounds resulted in no biofilm (Fig. 5e). Taken together, these data argue that the PrsA function in biofilm development depends on the presence of glycans in the IDR.

S. mutans PrsA, PBP1a and PknB IDRs enhance protein expression

To test a hypothesis that the Ser/Thr-rich IDRs are glycosylated, we fused the C-termini of the full lengths and the IDR-truncated variants of *S. mutans* PrsA and PBP1a with 3XFLAG creating PrsA-3XFLAG, PBP1a-3xFLAG, PrsAΔIDR-3XFLAG and PBP1aΔIDR-3xFLAG. We expressed these fusions from their chromosomal locus in the WT and Δ*pgfS* backgrounds. Immunoblot analysis with the anti-FLAG antibodies revealed increased electrophoretic mobilities of PrsA-3XFLAG and PBP1a-3xFLAG in the Δ*pgfS* backgrounds compared to WT, but not PrsAΔIDR-3XFLAG and PBP1aΔIDR-3xFLAG in the Δ*pgfS* background compared to WT, supporting the idea that sugars are attached to extracytoplasmic Ser/Thr-rich IDRs (Fig. 6a, b). Importantly, expression of the IDR-less PrsA and PBP1a was significantly reduced (Fig. 6a, b, bottom panels). Similar results were obtained when PknB expression was evaluated in *S. mutans* WT and *pknB*ΔIDR by immunoblot using the anti-PknB antibodies (Fig. 6c). These findings demonstrate that the *S. mutans* PrsA, PBP1a and PknB IDRs enhance protein expression by a currently unknown mechanism.

To explore the possibility that the C-terminal IDRs increase protein stability, we *E. coli*-produced 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 shift assays conducted in a wide range of buffers and salts including the addition of 3.5% ethanol (Supplementary Fig. 16a, b), indicating that the IDR does not affect the thermal stability of PrsA. Moreover, the CD spectra of ePrsA-FL and ePrsA did not reveal any prominent structural changes in the secondary structure content between the variants (Supplementary Fig.16c).

Glycosylation protects the IDRs from proteolytic cleavage

Remarkably, examination of PrsA expression in *S. mutans* using anti-PrsA polyclonal antibodies revealed that PrsA migrates as a broad band at around 43-47 kDa in WT and Δ*pgfS*:p*pgfS*, but as two bands at around 38 and 40 kDa in Δ*pgfS* (Fig. 7a, lanes 1-3). We hypothesize 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 is not detected in the flag-tagged strains by the anti-FLAG antibodies (Fig. 6a). Upon exposure of cells to 3.5% ethanol, heterogeneity of PrsA was significantly

increased in WT and Δ*pgfS*:p*pgfS*. The protein migrated between 38 and 47 kDa (Fig. 7a, 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-3XFLAG variant (Fig. 7b). Immunoblot analysis with the anti-FLAG antibodies demonstrated that the ethanol-subjected WT cells produced an additional low molecular weight band (Fig. 7b, lane 3) which co-migrates with the non-glycosylated form of PrsA-3XFLAG (Fig. 7b, lane 2, red arrow). When expression of the PrsA-3XFLAG variant in the cells grown in normal conditions was assessed using the anti-PrsA polyclonal antibodies, the protein migrated as two bands in Δ*pgfS* (Fig. 7c, lane 2), consistent with the idea that the smallest fragment (black circle) is formed by proteolytic cleavage. Upon exposure of cells to ethanol stress, the PrsA-3XFLAG cleavage product was also detected in WT (Fig. 7c, lane 3) in line with the hypothesis that ethanol inhibits protein glycosylation. Because the smallest fragment of the PrsA-3XFLAG variant is not detected by the anti-FLAG antibodies, it confirms the hypothesis that this fragment is a proteolytically processed form of PrsA lacking the flag tag. These observations also suggest that the cleavage site for an unknown protease is located at the C-terminus of PrsA and 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 the anti-PrsA polyclonal antibodies revealed that PrsAΔIDR migrated as a single sharp band (Fig. 7d, lane 1), but the chimeras migrated as multiple smeared bands (Fig. 7d, lanes 3 and 5). The protein heterogeneity was especially remarkable for PrsA-IDR^{WapA}. Importantly, loss of glycosylation did not affect the electrophoretic mobility and expression of the PrsAΔIDR variant (Fig. 7d, 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. 7d, 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.

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

To understand how the length of the PrsA IDR impacts PrsA biogenesis, we compared the expression of the PrsAΔIDR_8, PrsAΔIDR_23 and PrsAΔIDR variants with the WT PrsA (Fig. 7e). Strikingly, PrsAΔIDR_8 and PrsAΔIDR_23 produced the major and minor bands approximately 37 kDa in the WT and the *O*-glycosylation deficient backgrounds (Fig. 7e, lanes 3-6, black circle) indicating that the truncations negatively affect glycosylation of the IDRs leading to increased proteolysis. Additional smeared bands were also present in the WT background expressing these truncated variants confirming their glycosylation (Fig. 7e, lanes 3 and 5). PrsA is a lipoprotein that has a dual role by providing a protein chaperone function to control the aggregation of client proteins and catalyzing the cis/trans isomerization of peptide bonds with proline residues [16, 35]. Because these truncations of the PrsA IDR result in a biofilm formation defect under ethanol stress (Fig. 5d), the immunoblotting data suggest that the proteolytically processed forms of PrsA are functionally deficient.

Finally, quantification of the WT PrsA and the PrsAΔIDR_8, PrsAΔIDR_23 and PrsAΔIDR variants revealed that PrsA expression positively correlates with the increasing length of the IDR (Fig. 7e, bottom panel). In sum, these data indicate that the long IDR of PrsA facilitates efficient glycosylation and protein expression. This conclusion is further corroborated by the observation that the glycosylated form of the WT PrsA migrates as a smeared band (Fig. 7a and e, lane 1), in contrast to a sharper band of the glycosylated form of the PrsA-3XFLAG variant which has the additional 24 amino acids at the C-terminus (Fig. 6a, 7b and c, lane 1). The reduction in heterogeneity of the PrsA-3XFLAG variant is likely due to an increase in the glycosylation efficiency.

Collectively, our results support a model in which the intrinsic disorder of the PrsA C-terminal region plays a functional role in biofilm formation in ethanol-stressed cells (Fig. 8). To protect this important region from degradation by an unknown protease, the IDR contains stretches of Ser and Thr residues that the Pgf machinery modifies with GalNAc moieties. PrsA expression and the degree of glycosylation strongly depend on the length of the IDR.

Discussion

In this study, using machine learning algorithms we identify a unique class of IDRs linked to the extracytoplasmic proteins in streptococci (Supplementary Table 1). These IDRs are detected in proteins participating in cell wall biosynthesis, cell division, signal transduction and protein folding. A distinct feature of this class of IDRs is the presence of Ser/Thr-rich patches. In *S. pyogenes* and *S. pneumoniae*, we elucidated a novel Glc-type *O*-linked glycosylation mechanism in which glycosyltransferase GtrB initiates the synthesis of a Glc-modified bactoprenol-phosphate lipid molecule for the transfer of Glc to extracytoplasmic proteins harboring the Ser/Thr-rich IDRs. Previous studies have reported that *S. mutans* possesses the *O*-linked protein glycosylation system encoded by the Pgf gene cluster which covalently links N-acetylhexosamine sugars, presumably GalNAc and GlcNAc, with the cell wall-associated adhesins Cnm and WapA [32, 33]. Like GtrB in *S. pyogenes* and *S. pneumoniae*, the Pgf-encoded glycosyltransferase, PgfS, initiates the protein glycosylation pathway [27, 28]. We show that the Pgf machinery transfers GalNAc to the extracellular IDRs tethered to a subset of membrane proteins in *S. mutans*.

A lectin-affinity purification of glycoproteins reveals that the major membrane glycoproteins are PrsA and PBP1a in *S. pyogenes* and *S. mutans*, and PBP1a/PBP1b in *S. pneumoniae*. The presence of glycans on the Ser/Thr-rich IDRs identified by *in silico* analysis and the lectin-affinity enrichment 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. It is noteworthy that the enrichment strategy failed to capture all extracytoplasmic proteins carrying this class of IDRs (Supplementary Table 1). This limitation may be due to the low abundance of some glycoproteins in the analyzed conditions and the high resistance of glycoproteins to tryptic cleavage which is used for MS analysis. Nevertheless, the enrichment strategy impartially reveals a strong correlation between the Ser/Thr-rich IDRs and *O*-glycosylation. Moreover, sequence alignments of glycoproteins with the bacterial orthologs indicate that this class of IDRs occurs universally in streptococci suggesting that *O*-glycosylation of extracytoplasmic IDRs is widespread among these species.

To understand how glycosylation impacts the biological functions of IDRs, we focused on *S. mutans* which is a major etiological agent of human dental caries. Dental caries progression involves the binding 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 in the absence of stress, the PrsA IDR is dispensable for *S. mutans* biofilm formation, it becomes critical upon ethanol exposure. By generating the PrsA chimeras in which the PrsA IDR is replaced with the WapA or the *S. pyogenes* PrsA1 IDRs, we illustrate 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 [36]. The crystal structures of *Bacillus subtilis* and *Listeria monocytogenes* PrsA homologs demonstrate that the enzymes fold into two subdomains, the chaperone-like and the peptidyl-prolyl isomerase domain, suggesting that PrsA may act as a folding chaperone by preventing aggregation of secreted proteins [16, 37]. Many IDRs have been shown to engage in weak multivalent interactions with one another or with other IDRs [12]. Our findings raise the possibility that PrsA employs a highly dynamic disordered C-terminal tail to capture unfolded client proteins bringing them into proximity of the chaperone domain of PrsA. This chaperone function of PrsA is probably critically important in ethanol-stressed cells because ethanol is known to denature proteins.

Our data indicate that loss 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. Interestingly, 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. Thus, *S. mutans* have evolved sophisticated mechanisms to regulate the function and expression of PrsA by the C-terminal IDR which carries diverse features including the cleavage sites for an unknown protease, the *O*-glycosites and the length of the IDR. Furthermore, these features might not be restricted to *S. mutans* PrsA because the

expression of *S. mutans* PBP1a and PknB, and *S. pneumoniae* PBP1a is positively regulated by the IDRs, and *S. mutans* PknB lacking GalNAc decorations displays increased proteolytic processing. Interestingly, sequence alignments of the homologs of the identified glycoproteins reveal that in the C-terminal IDRs, the stretches of Ser/Thr residues are frequently adjacent to folded domains, followed by the long IDR tails. Such placement of Ser and Thr residues that serve as the *O*-glycosylation sites are probably evolved to improve glycosylation efficiency. Streptococcal O-glycosylation machinery may be coupled to protein translocation as has been proposed for N-linked glycosylation in bacteria [38] and directly demonstrated for N-glycosylation in the eukaryotic endoplasmic reticulum [39]. The long IDR tails may slow the release of proteins from the Sec translocon thereby providing a time window for efficient protein folding and modification with sugars.

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. This observation highlights the highly tunable nature of IDRs to external stimuli and pinpoints a direct mechanism used by *S. mutans* to modulate the production of important membrane 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, we anticipate that further examination of the prokaryotic proteome may expand our observation to other bacterial species, and also lead to the discovery of novel motifs in the procaryotic IDRs.

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 [40, 41]. 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 [42]. 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 $OD_{620} \approx 0.1–0.4$ were diluted to $OD_{620} \approx 0.005$ in BHI broth. *S. pneumoniae* D39 cultures were sampled for western analysis at $OD_{620} \approx 0.15–0.2$ (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 [43]. 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 [44]. 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 [45] 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 [45]. 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 [42, 46, 47]. Strains containing markerless alleles in native chromosomal loci were constructed using allele replacement via the P_c-[*kan-rpsL*$^+$] (Janus cassette) [48]. 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 [42, 47]. Ectopic expression *gtrB* for complementation was achieved with a constitutive P_{ftsA} promoter in the ectopic *bgaA* site [49]. 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 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 [30].

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

To purify recombinant proteins, cells were grown to exponential phase ($OD_{600} \approx 0.6$) 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 further 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, 1x 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 [50]. 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, USA) and purity was estimated by SDS-PAGE.

Preparation of membrane fractions from *S. pyogenes* and *E. coli* JW2347 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 $MgCl_2$, 1x bacterial protease inhibitor cocktail, 5 µg mL^{-1} DNAse, 5 µg mL^{-1} RNAse, and allowed to swell 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-OH 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 [28] 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 [28].

Co-sedimentation assay with fluorescent lectins

S. mutans cells grown overnight (≈17 hours) were harvested by centrifugation and washed three times with 1x phosphate-buffered saline (PBS). The cell pellet was resuspended in 0.5 mL of PBS at $OD_{600} \approx 9$. 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 [29]. Briefly, exponential phase cells ($OD_{600} \approx 0.3$) 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 equipped with 100X, 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 an 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 X protease inhibitor cocktail (Sigma-Aldrich) and 2 μg mL^{-1} DNase. To lyse *S. mutans* and *S. pyogenes*, recombinant 50 μg mL^{-1} zoocin and PlyC 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 1x 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 Scientific). Briefly, Coomassie-stained protein bands were excised and subjected to a modified in-gel trypsin digestion procedure [51]. Peptides were reconstituted in 2.5% acetonitrile plus 0.1% formic acid, loaded *via* a Thermo EASY-LC

(Thermo Fisher Scientific, Waltham, MA) 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, Waltham, MA) or a Velos Orbitrap Pro ion-trap mass spectrometer (Thermo Fisher Scientific, Waltham, MA). Peptide sequences were determined by matching protein databases with the acquired fragmentation pattern by Sequest (Thermo Fisher Scientific, Waltham, MA). 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 $OD_{600} \approx 0.5$ 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, 1x bacterial protease inhibitor cocktail, and 2 $\mu g\ mL^{-1}$ 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, USA). An equivalent amount of total proteins was loaded in each lane and resolved on a 15% (wt/vol) SDS-PAGE gel or 4-12% SurePAGE™ Bis-Tris gel (Genscript Corporation), 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 [15]. Rabbit polyclonal antibodies against the *S. pyogenes* PknB homolog, and the *S. mutans* PrsA, FtsQ and PknB homologs were produced by Cocalico Biologicals, Denver. Purified recombinant PrsAΔIDR, FtsQΔIDR, $PknB\Delta IDR^{Smu}$ and $PknB\Delta IDR^{Spy}$ were used for the production of 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 17.

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 western blotting was performed as previously described [24, 52]. Briefly, bacteria were grown exponentially in 5 ml BHI broth to an $OD_{620} \approx 0.15–0.2$. 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 $OD_{620} \approx 0.16$ 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 2x 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 for the normalization of antibody signals. Rabbit polyclonal antibodies used were anti-StkP (1:7,000) [17, 53], anti-PBP1a^{Spn} (1:7,000) [24], anti-PBP1b^{Spn} (Dr. Cécile Morlot, 1:7000), and anti-PrsA^{Spn} (Dr. Marien de Jonge, 1:7000). 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, NY) 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 Shift Assay

The protein thermal stability measurements were performed using Tycho NT.6 (NanoTemper Technologies, Germany). 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 wavelength range from 200–250 nm 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 [54].

Prediction of extracytoplasmic IDRs

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* (GenBank: CP000410.2) and *S. pneumoniae* TIGR4 (https://genomes.atcc.org/genomes/641decb736d446cc) with default batch size [55]. Disordered region prediction was performed using the compact protein language model

DR-BERT using the pretrained weights [56]. Predictions were run and aggregated using a custom Python script. The membrane topology of the subset of membrane and secreted proteins possessing IDRs (Supplementary Dataset 1) was predicted using TOPCONS [14]. Extracytoplasmic IDRs in the subset of membrane and secreted proteins were also validated by a machine-learning protein structure prediction algorithm AlphaFold2 [13].

Statistical analysis

Unless otherwise indicated, statistical analysis was performed using Graph Pad 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.

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 the access to Leica SP8 confocal microscope.

This work was supported by NIH grants R01 DE028916 from the NIDCR and R21 AI149366 from the NIAID (to NK) and R35 GM131767 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. 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. 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).

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

3. 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).

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

5. 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).

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

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

8. 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).

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

10. 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).

11. 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).

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

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

14. 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).

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

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

17. 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).

18. 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).

19. 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).

20. 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).

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

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

23. 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).

24. 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).

25. 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).

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

27. Mann E, Whitfield C. A widespread three-component mechanism for the periplasmic modification of bacterial glycoconjugates. *Canadian Journal of Chemistry* **94**, 883-893 (2016).

28. 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).

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

30. 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).

31. 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).

32. 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).

33. 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).

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

35. 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).

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

37. 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).

38. Silverman JM, Imperiali B. Bacterial N-Glycosylation Efficiency Is Dependent on the Structural Context of Target Sequons. *J Biol Chem* **291**, 22001-22010 (2016).

39. Braunger K*, et al.* Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. *Science* **360**, 215-219 (2018).

40. 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).

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

42. 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).

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

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

45. 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).

46. 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).

47. 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).

48. 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).

49. 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).

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

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

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

53. 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).

54. 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).

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

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

a

S. pyogenes	PrsA1-	304-TKAASESSTTSESSKAAEENPSESEQTQTSSAEEPTETEAQTQEPAAQ-350
	PBP1a-	674-NNLYSNNTTTASSQTTSDDTSSSNDTSNSTNTDNNGSHPSTDDKKTTH-721
	PknB-	593-HSSSSSSSSTSSSNSSSINDSTAPGSNTELSPSETTSQTP-632
	LCP1 (LytR)-	338-GSTKSQTVNNNYDSSGQAPSYSDSHSSYANYSSGVDTGQSASTDQDSTASSHRPATPSSSSDA LAADESSSSGSGSLVPPANINPQT-424
S. pneumoniae	PrsA-	300-GGDSSSSSSTSNE-313
	PBP1a-	666-STESSSSSSDSSTSQSSSTTPSTNNSTTTNPNNNTQQSNTTPDQQNQNPQPAQP-719
	PBP1b-	794-SSSSSSSSSSSSDSSNSSTTRPSSSRARR-821
	RodZ-	137-PEEPSLSNYSVVQSTSSTSSVPHSSSSSSSS-167
	CdaR-	228-KPVAPSSSTSNSSTSSSSETSSSTKATSSKTN-259
	MapZ-	279-VLDTNAKAKSDAKFTDIKTGNTELDKVLDKAISLGKSQQTSTSSSSSSQTSSSSSSQASSNTTSEP KPSSSNETRSSRSEVNMGLSSA-366
	LytC-	17-TSHVAANETEVAKTSQDTTTASSSSEQNQSSNKTQTS-53
	SP_2093*-	212-SSGSASNESYNGSSNSNVDYSSSEQTNGYSNNYGGQDYSGSGDSSTNGGSSEQYSSSNSNSG ANN-276
S. mutans	PrsA-	298-SGSGSSGSTTTTTAASSAATTAADDQTTAAETTAAE-333
	PBP1a-	666-SSSSSSSSSSDSNNNDQNNNQTTEASSDSSSSSSDATTSSNP-707
	PknB-	583-SHSGSSSSESSNSEGTTSSEASTDSSSSATTTSH-616
	LCP1 (BrpA)-	346-SSSSASDYSSSGNYSGSSSDYGSSSSYGSNSSSGSSSDYSGQNSYNQGNYQQPAAGTGIGN-406
	LtaS-	706-SSKESSTTTSSK-717
	FtsQ-	340-SSENSSNAENTDSTQSSEANSQEGSQSVTQATVSQ-374

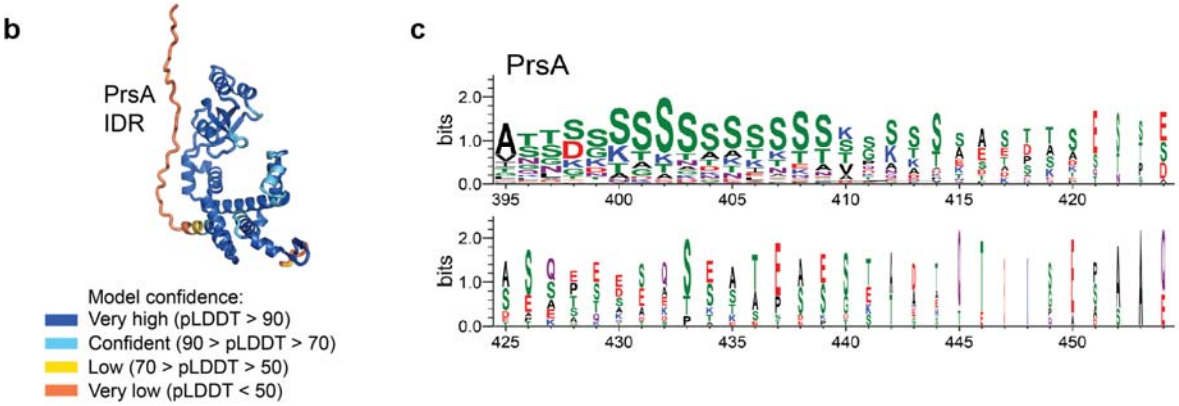

Fig. 1. Features of Ser/Thr-rich IDRs identified in streptococcal extracytoplasmic proteins

(**a**) IDR sequences of extracytoplasmic proteins. Ser and Thr residues are in orange. Numbers correspond to amino acid residue starting and ending the IDR sequence. * denotes TIGR4-specific protein. (**b**) Predicted AlphaFold2 structural model of *S. mutans* PrsA. Colors correspond to per-residue predicted local distance difference test (pLDDT) values. Protein segments with pLDDT < 50 usually correspond to unstructured regions. PrsA is a dimer; only one subunit is displayed. (**c**) A sequence logos of the PrsA IDR based on multiple sequence alignments of PrsA homologs from streptococci.

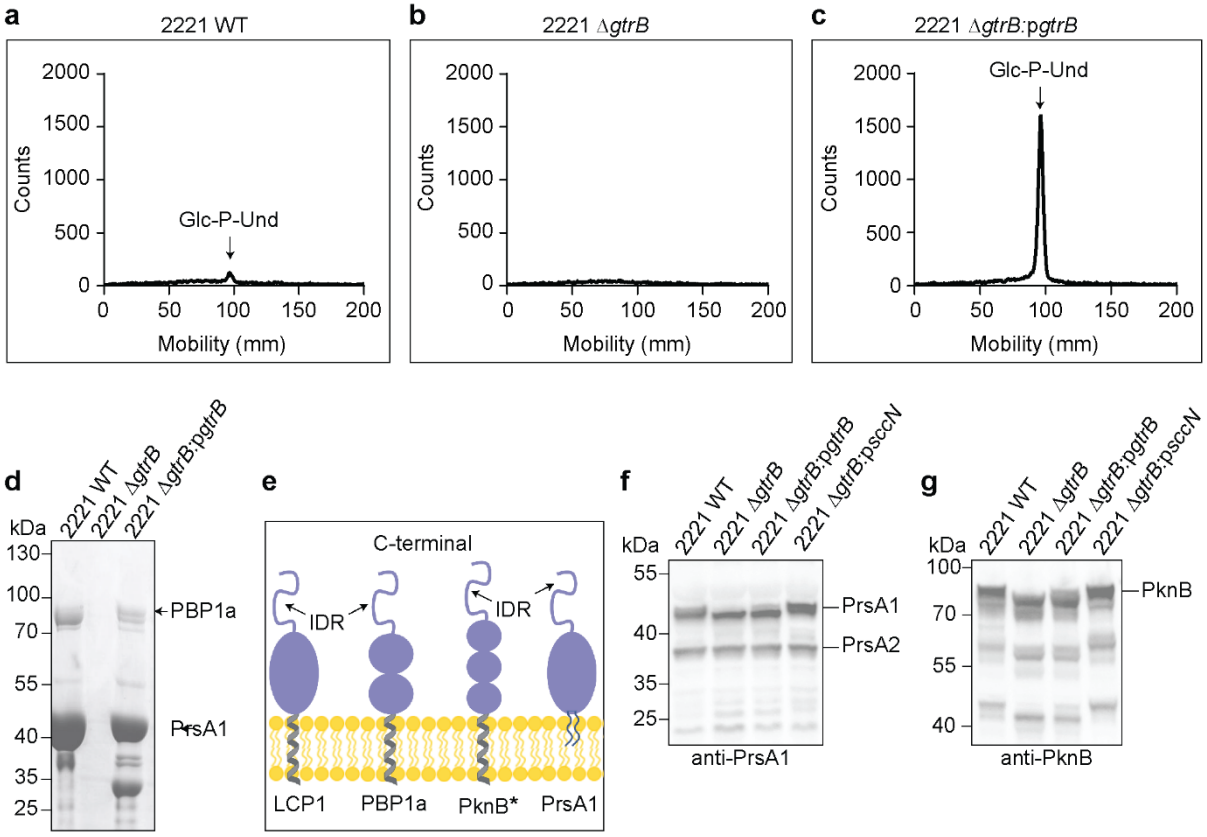

Fig. 2. Glucosylation of *S. pyogenes* extracytoplasmic proteins harboring IDRs

(**a, b** and **c**) TLC of [^{3}H]Glc-lipids from *in vitro* incubations of UDP-[^{3}H]Glc with membrane fractions isolated from *S. pyogenes* WT (**a**), Δ*gtrB* (**b**), and Δ*gtrB*:p*gtrB* (**c**). Membrane fractions were subjected to mild alkaline de-acetylation with 0.1N KOH to eliminate glucosyldiglyceride and incubated with UDP-[^{3}H]Glc. Extracted [^{3}H]Glc-lipids were analyzed by TLC. (**d**) Coomassie-stained gel of *S. pyogenes* membrane proteins purified by conA affinity chromatography. (**e**) Topology of extracytoplasmic domains of *S. pyogenes* conA-bound proteins. * the cytoplasmic domain of PknB is omitted for clarity. (**f** and **g**) Immunoblot analysis of PrsA1 (**f**) and PknB (**g**) in *S. pyogenes* WT, Δ*gtrB*, Δ*gtrB*:p*gtrB*, and Δ*gtrB*:p*sccN* using specific antibodies. Anti-PrsA1 antibodies also recognize PrsA2. The experiments were performed independently three times in **d**, **f**, and **g,** and two times in **a**, **b**, and **c**, and they yielded the same results. Representative image from one experiment is shown.

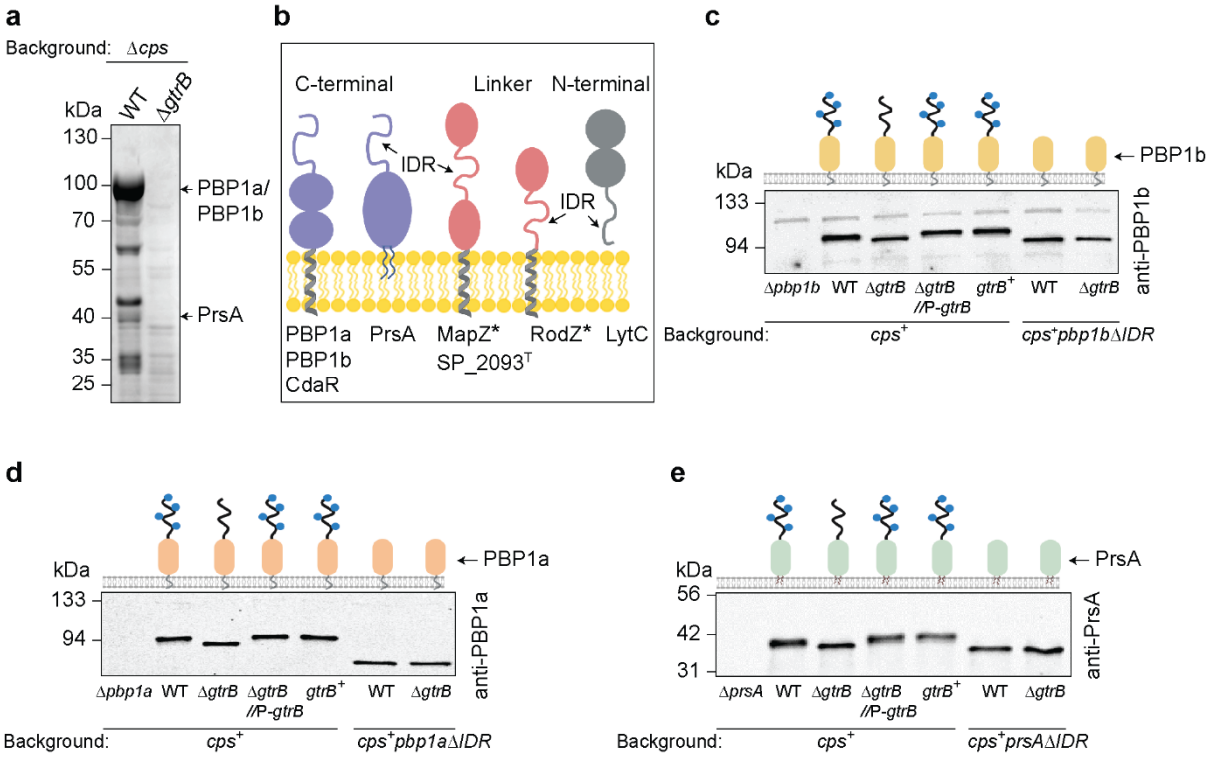

Fig. 3. Glucosylation of *S. pneumoniae* extracytoplasmic proteins harboring IDRs

(**a**) Coomassie-stained gel of *S. pneumoniae* membrane proteins obtained from D39W Δ*cps* (IU1824*)* and the isogenic Δ*gtrB* 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** and **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**. 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. 11b, c, d and Supplementary Fig. 12a, b, c.

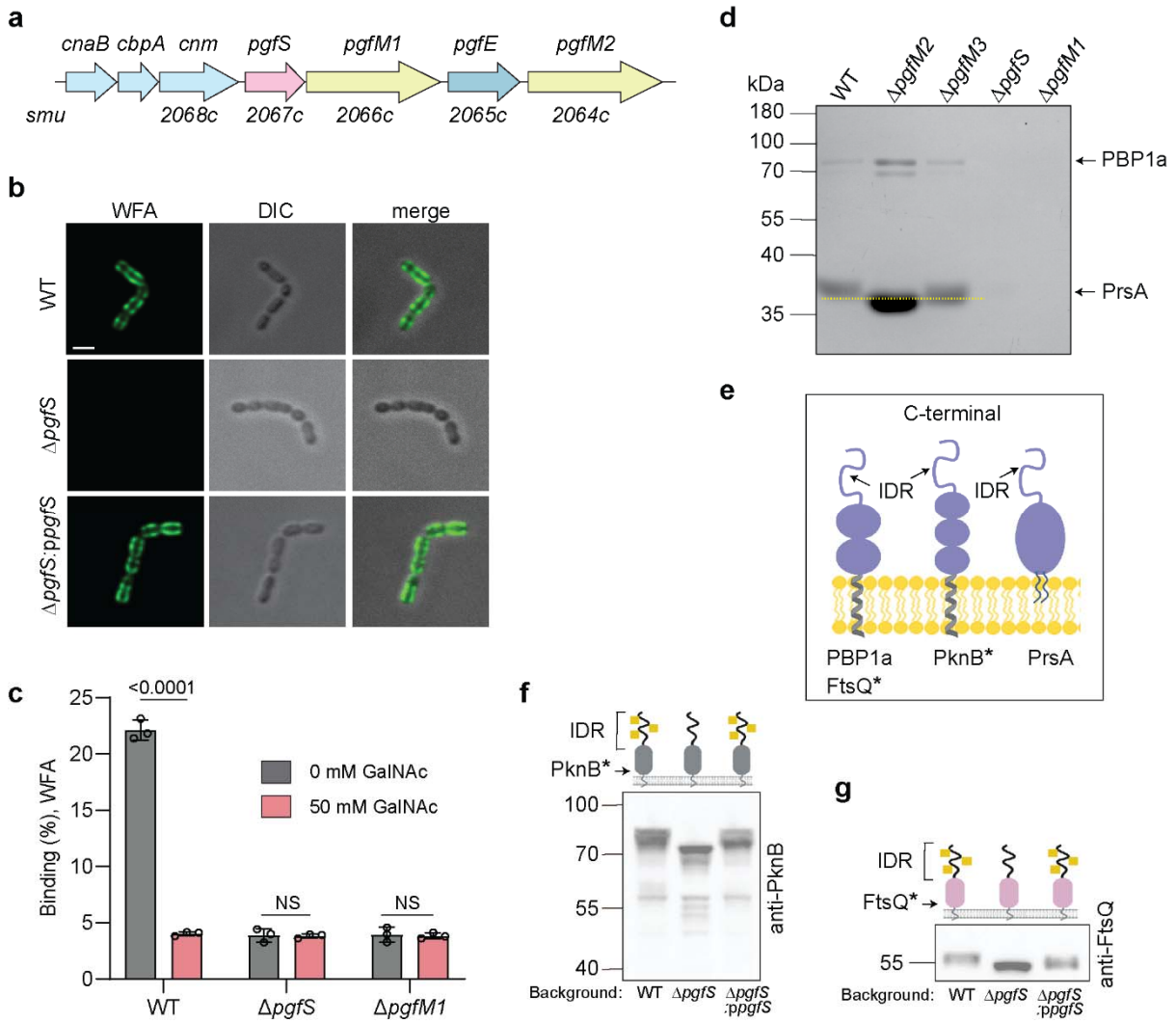

Fig. 4. The *pgf* operon glycosylates *S. mutans* membrane proteins with GalNAc
(**a**) Schematic representation of the *pgf* operon in *S. mutans* Xc. The *cnaB*, *cbpA*, and *cnm* genes encode surface adhesins that are present only in the Cnm$^+$ strains. (**b**) Fluorescence and DIC images of bacteria bond to FITC-conjugated WFA. The scale bar is 1 µm. (**c**) Co-sedimentation of FITC-conjugated WFA with the *S. mutans* strains. 0 or 50 mM GalNAc was added in 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. (**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** and **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 blob of PknB and FtsQ represents three- and two extracellular domains, respectively. * the cytoplasmic domains are not shown in the schematic. GalNAc is depicted as yellow squares. The experiments shown in **b**, **f**, and **g** were performed independently three times and yielded the same results. Representative image from one experiment is shown.

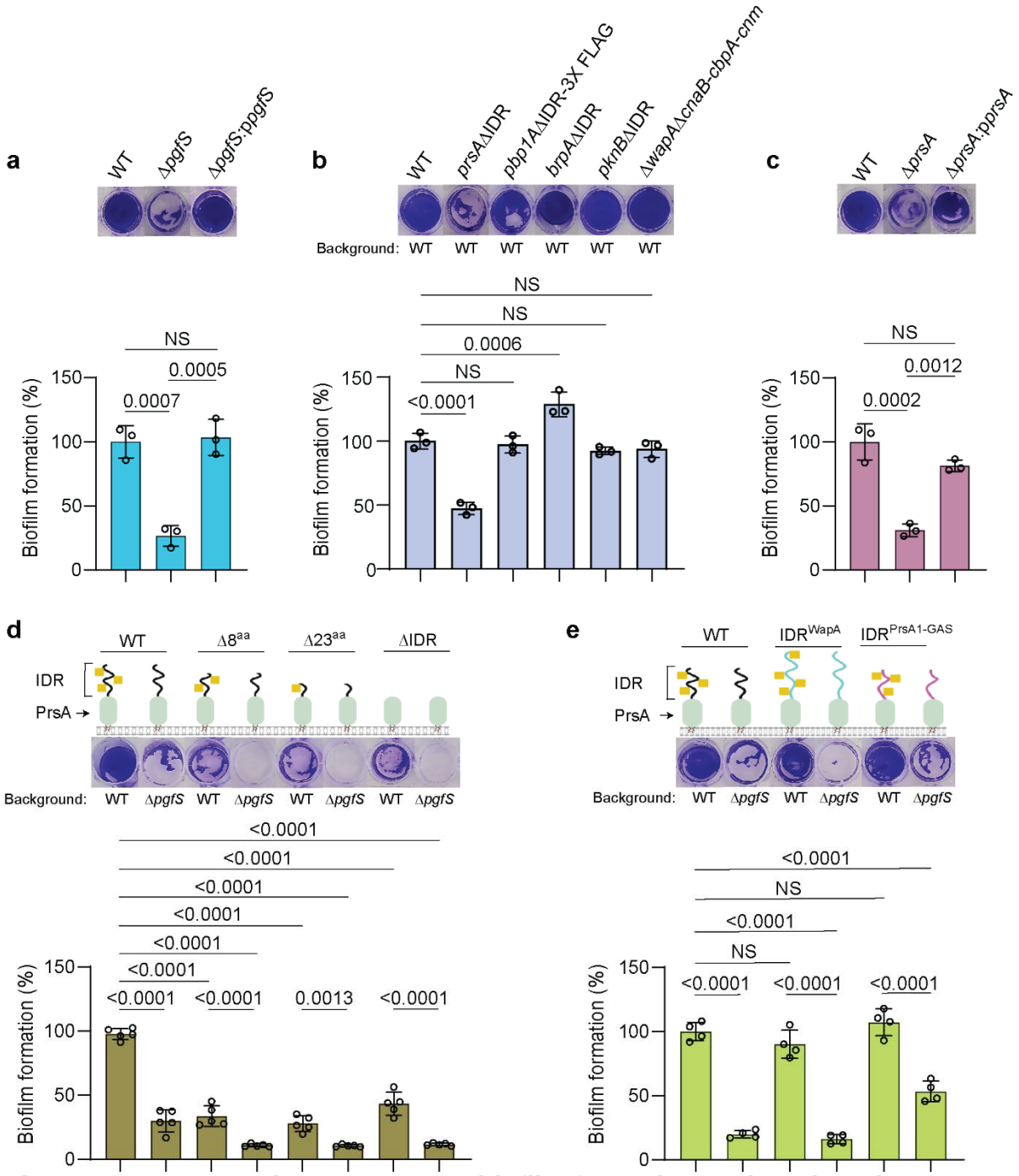

Fig. 5. PrsA IDR regulates *S. mutans* biofilm formation under ethanol stress

(**a**) Biofilm formation of *S. mutans* WT, Δ*pgfS,* and Δ*pgfS*:p*pgfS*. (**b**) Biofilm formation of *S. mutans* strains expressing the IDR-less variants of PrsA, PknB, LCP1 (known as BrpA), PBP1a or deficient in four cell wall-associated adhesins WapA, Cnm, CnaA and CnaB. **c**, Biofilm formation of *S. mutans* WT, Δ*prsA,* and Δ*prsA*:p*prsA*. (**d**) Biofilm phenotype of *S. mutans* strains expressing PrsA with truncations in the C-terminal IDR.

(**e**) Biofilm formation of *S. mutans* strains expressing PrsA variants with the chimeric IDRs. In **d** and **e**, schematic representations of PrsA variants are shown (top panel). GalNAc is depicted as yellow squares. In **a**, **b**, **c**, **d**, and **e**, representative biofilms stained with crystal violet are shown. Quantifications of biofilm formation are shown in bottom panels. Columns and error bars represent the mean and S.D., respectively (n = 3 in **a**, **b**, and **c;** n = 4 in **e** and n = 5 in **d** biologically independent replicates. *P* values were calculated by one-way ANOVA with Tukey's- (**a**, **c**, **d**, **e**) and Dunnett's (**b**) multiple comparisons test.

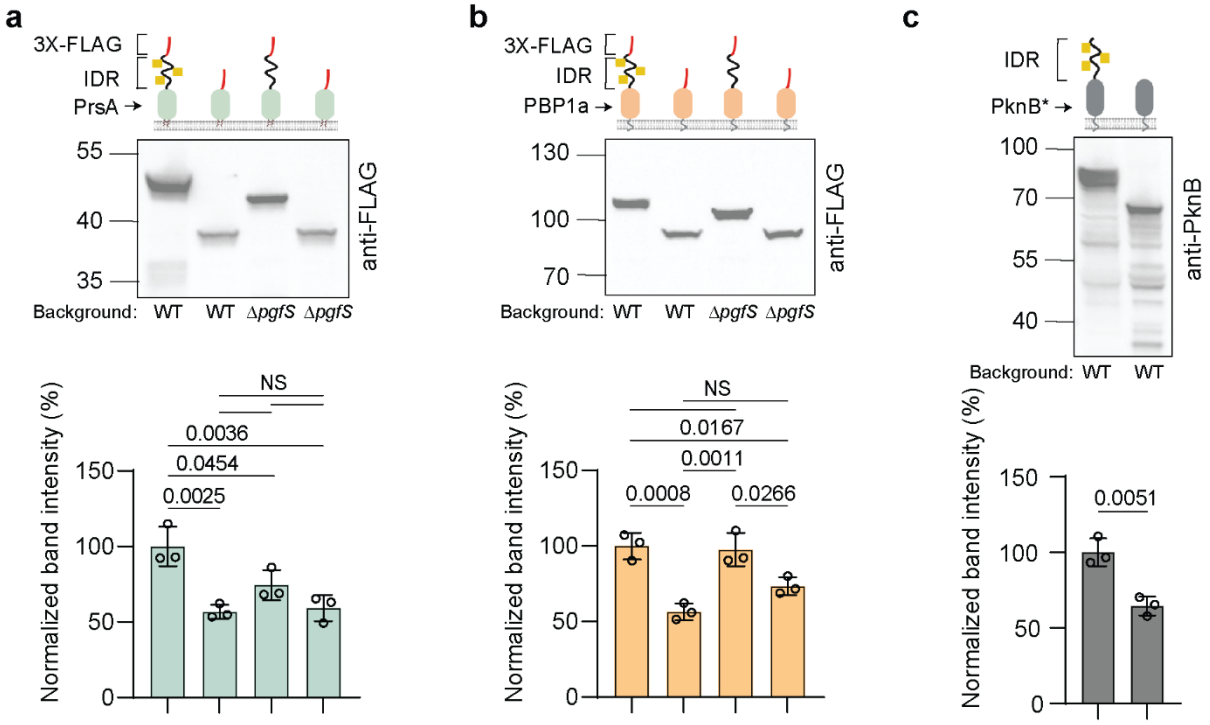

Fig. 6. PrsA, PBP1a and PknB IDRs enhance protein expression in *S. mutans*

(**a** and **b**) Immunoblot analysis of 3XFLAG-tagged variants of PrsA (**a**) and PBP1a (**b**) in the WT and Δ*pgfS* backgrounds examined by anti-FLAG antibodies. (**c**) Immunoblot analysis of PknB variants in the WT background assessed by anti-PknB antibodies. Schematic representations of PrsA, PBP1a and PknB variants are shown in **a, b** and **c** top panels, respectively. PknB is labeled as PknB* in **c** to show that the cytoplasmic region of this protein is omitted for clarity. Each of the single blob of PBP1a and PknB represents two- and three extracellular domains, respectively. GalNAc is depicted as yellow squares. The experiments were performed independently three times in **a**, **b** and **c**, and they yielded the same results. Representative image from one experiment is shown. Quantifications of immunoblot band intensities of PrsA, PBP1a and PknB are shown in **a**, **b** and **c** bottom panels. Band intensities from each lane of the immunoblot were calculated using ImageJ, and the data were normalized. Columns and error bars represent the mean and S.D., respectively (n = 3 biologically independent replicates). *P* values were determined by one-way ANOVA with Tukey's multiple comparisons test (**a** and **b**) or two-tailed Student's *t*-test (**c**).

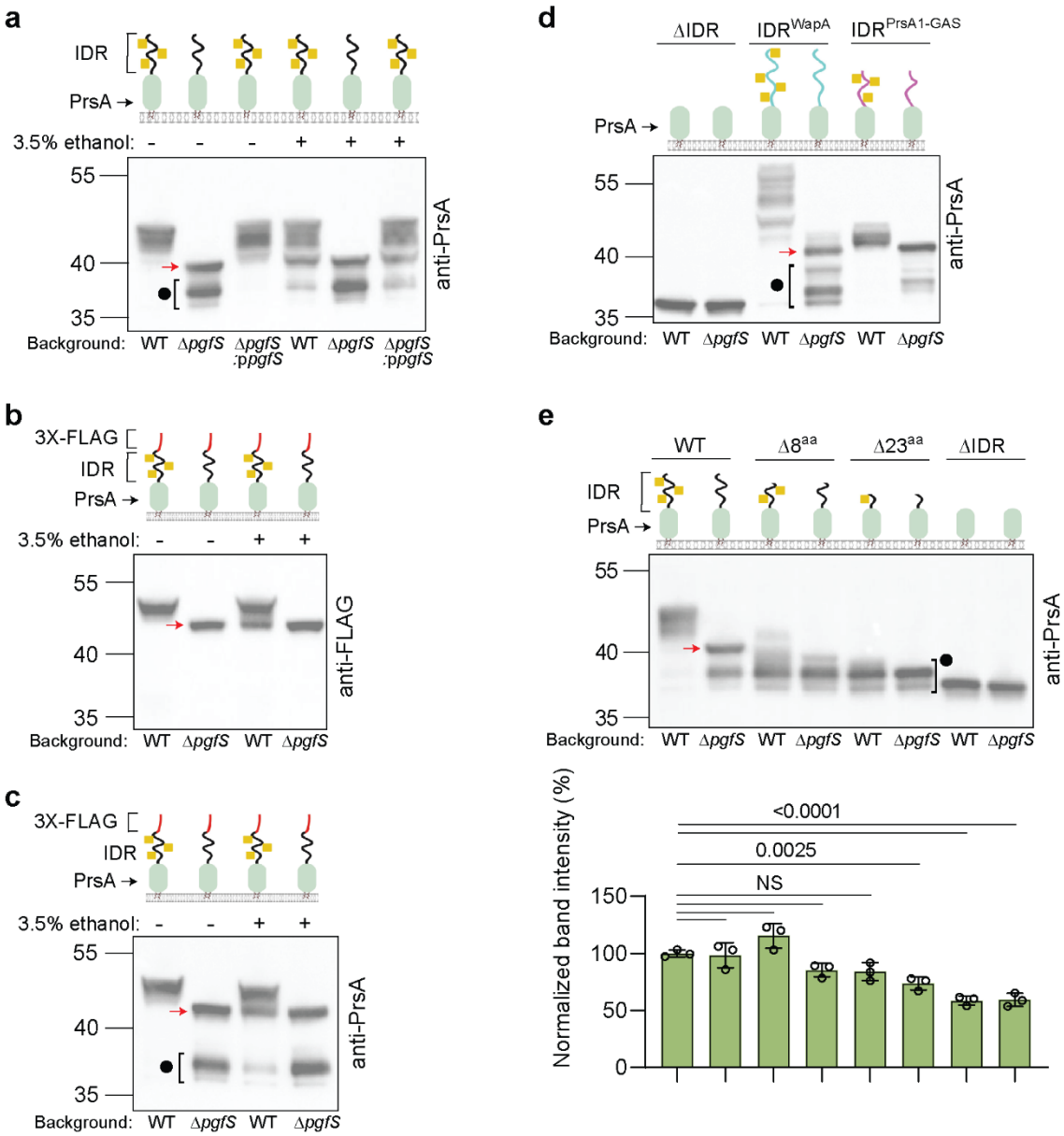

Fig. 7. IDRs regulates biogenesis of *S. mutans* extracytoplasmic proteins

(**a**) Immunoblot analysis of PrsA in the WT, Δ*pgfS,* and Δ*pgfS*:p*pgfS* backgrounds grown in normal or ethanol-stressed conditions. PrsA is assessed by anti-PrsA antibodies. (**b** and **c**) Immunoblot analysis of 3XFLAG-tagged variant of PrsA in the WT and Δ*pgfS* backgrounds grown in normal or ethanol-stressed conditions. PrsA is assessed by anti-FLAG (**b**) and anti-PrsA (**c**) antibodies. (**d** and **e**) Immunoblot analysis of PrsA variants carrying the chimeric IDRs (**d**) or truncated IDRs (**e**) in the WT and Δ*pgfS* backgrounds grown in normal conditions. PrsA is assessed by anti-PrsA

antibodies. Quantifications of immunoblot band intensities of PrsA carrying the truncated IDRs are provided in **e** (bottom panel). Total band intensities from each lane of blot in **e** were calculated using ImageJ, and the data were normalized. Columns and error bars represent the mean and S.D., respectively (n = 3 biologically independent replicates). *P* values were determined by one-way ANOVA with Dunnett's multiple comparisons test. Schematic representations of PrsA variants are 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. Representative image from one experiment is shown.

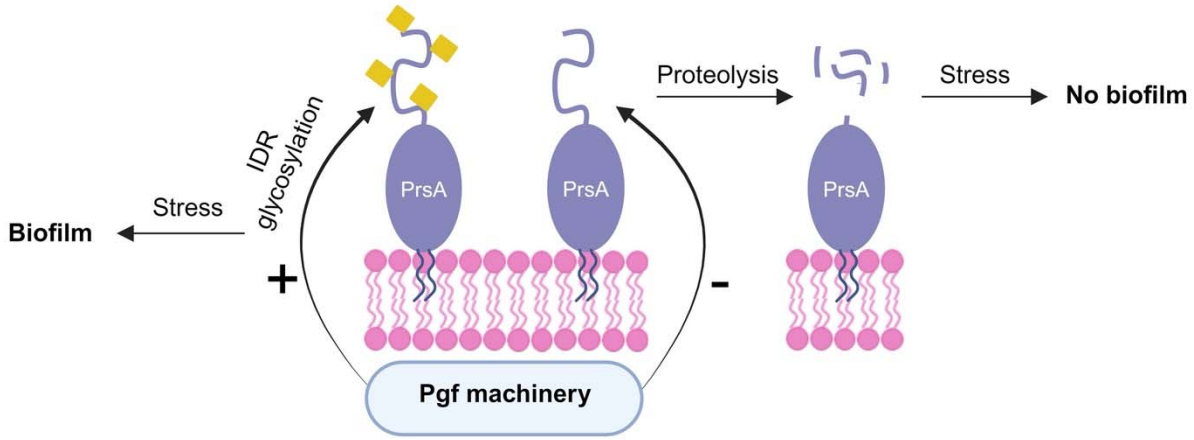

Fig. 8. The intrinsic disorder of the PrsA C-terminal region plays a functional role in biofilm formation in ethanol-stressed cells

Pgf protein *O*-glycosylation machinery covalently attaches GalNAc moieties to the C-terminal IDR of peptidyl-prolyl cis/trans isomerase PrsA. This functional form of PrsA is required for *S. mutans* biofilm formation in ethanol-stressed cells. Loss of *O*-glycosylation exposes the IDR to proteolytic attack resulting in a non-functional PrsA which does not aid in biofilm formation.