Structural foundation for the role of enterococcal PrgB in conjugation, biofilm formation and virulence

Wei-Sheng Sun[1,2], Lena Lassinantti[1], Michael Järvå[1], Andreas Schmitt[1], Josy ter Beek[1,2] and Ronnie P-A Berntsson[1,2]

[1] Department of Medical Biochemistry and Biophysics, Umeå University, SE-90187 Umeå, Sweden

[2] Wallenberg Centre for Molecular Medicine, Umeå University, Umeå, Sweden

[#] Correspondence should be addressed to R.P-A.B. (email: ronnie.berntsson@umu.se)

Abstract

Type 4 Secretion Systems are a main driver for the spread of antibiotic resistance genes and virulence factors in bacteria. In Gram-positives, these secretion systems often rely on surface adhesins to enhance cellular aggregation and mating pair formation. One of the best studied adhesins is PrgB from the conjugative plasmid pCF10 of *Enterococcus faecalis,* which has been shown to play major roles in conjugation, biofilm formation and importantly also in bacterial virulence. Since *prgB* orthologs exist on a large number of conjugative plasmids in various different species, this makes PrgB a model protein for this widespread virulence factor. Here we report structures for almost the entire PrgB, in the presence or absence of DNA, using a combination of X-ray crystallography and cryo-EM. These reveal that PrgB undergoes a large conformational change upon DNA-binding and that it contains four immunoglobulin-like domains. We re-evaluate previously studied variants and present new *in vivo* data where specific domains or conserved residues have been mutated. For the first time we can show a decoupling of cellular aggregation from biofilm formation and conjugation in *prgB* mutant phenotypes. Based on the presented data, we propose a new functional model to explain how PrgB mediates its different functions. We hypothesize that the Ig-like domains act as a rigid stalk that both protect the previously studied polymer adhesin domain from proteolysis, as well as presenting it at the right distance from the cell wall.

36 **Introduction**

37 *Enterococcus faecalis* (*E. faecalis*) is one of the leading causes of hospital acquired infections,
38 such as urinary tract infections and endocarditis[1,2]. These infections are difficult to treat as *E.*
39 *faecalis* has the tendency to form biofilms and is often resistant to various antibiotics. They are
40 also notorious for spreading antibiotic resistance and other fitness advantages by transfer of
41 mobile genetic elements (MGEs), which can be located on conjugative plasmids or in the
42 chromosome[3,4]. Conjugative plasmids usually also encode a Type 4 Secretion System (T4SS)
43 that mediates its transfer, via conjugation, from a donor cell to a recipient cell[5–8]. However,
44 conjugative plasmids and their T4SS have almost exclusively been studied in Gram-negative
45 model systems[8].

46

47 One of the few well-characterized Gram-positive conjugative plasmids is pCF10 from *E.*
48 *faecalis*[6,9,10]. This conjugative plasmid contains a ~27 kbp operon that is tightly regulated by
49 the P_Q promoter[11–14] and that encodes all proteins needed for conjugation. This operon also
50 encodes three cell-wall anchored proteins: PrgA, PrgB, and PrgC. PrgA is a conjugation
51 regulator that provides surface exclusion to prevent unwanted conjugation. We have previously
52 shown that PrgA consists of a protease domain that is presented far away from the cell wall via
53 a long stalk and that it's likely mediating the proteolytic cleavage of PrgB[15,16]. PrgC is a
54 virulence factor, but its function and structure remain unknown[17]. PrgB is the main adhesin
55 produced by pCF10 and has been studied for well over 3 decades. This protein is around a 140
56 kDa in size and possesses an N-terminal signal sequence and a C-terminal LPXTG cell wall
57 anchor motif[18]. PrgB distributes over the entire surface of the cell-wall and increases cellular
58 aggregation, biofilm formation and the efficiency of plasmid transfer[19,20]. Several mammalian
59 infection model systems have shown that PrgB is a strong virulence factor[18,21–26]. One reason
60 for this virulence is that PrgB mediates biofilm formation in an extracellular DNA (eDNA)

61 dependent manner[17]. Homologs of PrgB have been identified in many other conjugative
62 plasmids[16], suggesting that PrgB-like proteins confer important roles in a large number of
63 bacterial species[27–29].
64
65 PrgB was initially identified as one of the driving forces in cellular aggregation[19]. To
66 understand how it mediated this process, previous research has tried to identify the various
67 protein domains that are present in PrgB and evaluate their function(s). Two RGD (Arg-Gly-
68 Asp) motifs were identified (see figure 1A), and found to be important for vegetation and
69 biofilm formation in the host tissue environment[18,23]. The N-terminal half of PrgB was found
70 to be required for aggregation and to bind lipoteichoic acid (LTA)[29–31], which is a major
71 constituent of the cell-wall in Gram-positive bacteria. In 2018, we solved the structure of
72 $PrgB_{246-558}$ and showed that it has a lectin-like fold that was most similar to adhesins from
73 various oral Streptococci. As these adhesins are known to bind various polymers, we
74 subsequently referred to $PrgB_{246-558}$ as the polymer adhesin domain[16]. We have shown that this
75 domain can bind both LTA and eDNA in a competitive manner. Bound eDNA is thereby
76 strongly compacted as it is wrapped around the domains positively charged surface[32]. We
77 therefore proposed that PrgB could use eDNA to promote cell-to-cell contacts, as an alternative
78 to direct binding to the LTA from a recipient cell (Fig. 1B). As all described polymer adhesin
79 domains, PrgB has a central ridge with a conserved cation binding site[16,33]. In the homologous
80 GbpC, from *Streptococcus mutans*, this site has been suggested to bind glucans[34]. However, no
81 interaction with glucans has been observed for PrgB or any other homologs[35]. Thus, the
82 importance of this conserved motif remains an open question.
83
84 The polymer adhesin domain plays an important role in the function of PrgB, but it only
85 accounts for around a quarter of the entire protein. No structural data has been available for the

86 remainder of PrgB. Here we present the structure of almost full-length PrgB, based on both X-
87 ray crystallography and cryo-EM. This allows us to put the large amount of available
88 phenotypic data into a structural context and explain a lot of previous observations. We also
89 constructed several new mutants of *prgB* that better fitted the found domain organization and
90 analyzed their *in vivo* effects on cellular aggregation, biofilm formation and conjugation
91 efficiency. Based on our findings, we conclude with an updated model of how PrgB mediates
92 its different functions.
93

94 **Materials and Methods**

95 *Bacterial strains and growth conditions*

96 See Table S1 for a full list of all strains, plasmids and oligonucleotides used. *Escherichia coli*
97 Top10 was used in molecular cloning and grown in Lysogeny broth (LB). *E. coli* BL21 (DE3)
98 was used for recombinant protein expression and grown in Terrific Broth (TB). The *E. faecalis*
99 strains were cultured in Brain-Heart infusion broth (BHI) or Tryptic Soy broth without dextrose
100 (TSB-D) as indicated in each assay. Concentrations of antibiotics for *E. coli* selection were as
101 follows: ampicillin (100 µg/ml), kanamycin (50 µg/ml), spectinomycin (50 µg/ml), and
102 erythromycin (150 µg/ml). In *E. faecalis* cultures, antibiotics were used as the following
103 concentrations: tetracycline (10 µg/ml), fusidic acid (25 µg/ml), erythromycin (20 µg/ml for
104 chromosome-encoded resistance; 100 µg/ml for plasmid-encoded resistance), spectinomycin
105 (250 µg/ml for chromosome-encoded resistance; 1000 µg/ml for plasmid-encoded resistance),
106 streptomycin (1000 µg/ml). Plasmids were transformed to *E. coli* by heat-shock transformation,
107 whereas *E. faecalis* strains were transformed by electroporation[36].
108

109 *Cloning and mutagenesis*

110 To insert a multiple cloning site in pMSP3545S, DNA oligos of MCS_fwd and MCS_rev were
111 resuspended in miliQ to 100 µM, mixed 1:1, denatured at 95°C for 15 min and slowly cooled
112 to room temperature. The annealed MCS oligo was ligated into the gel-purified backbone
113 fragment from pMSP3545S-*prgK* vector[37] digested with NcoI and XbaI (removing the *prgK*
114 insert). This was done in a 30 µL ligation mixture with 90 ng of the digested vector and a 9
115 times molar access of insert. 5 µL of this ligation mixture was transformed into Top10
116 competent cells and plated on LB agar plates with spectinomycin (50 µg/ml), and erythromycin
117 (150 µg/ml). The constructed plasmid was analyzed by restriction digestion and sequenced to
118 confirm the correct insertion of the multiple cloning site and is further called pMSP3545S-
119 MCS.
120
121 pMSP3545S-*prgB* was constructed by PCR of *prgB* from pCF10 with the NcoI-prgB-F and
122 BamHI-stop-prgB-R primer pair and placement into the pMSP3545S-MCS vector via
123 NcoI/BamHI restriction, gel-purification of the correct DNA fragments and subsequent
124 ligation. The constructed pMSP3545S-*prgB* was subsequently used as a template for
125 mutagenesis creating pMSP3545S-*prgB* deletion or point mutation variants. This was carried
126 out with inverse PCR (iPCR) using partially overlapping primer pairs. The iPCR products were
127 gel-purified with the DNA clean-up kit and digested with DpnI to remove residual template
128 plasmid DNA. The processed iPCR products were then transformed to Top10 competent cells.
129 For overexpression, $prgB_{188\text{-}1233}$ was cloned into the p7XC3GH vector using the FX cloning
130 system[38]. Mutations were introduced to p7XC3GH- $prgB_{188\text{-}1233}$ with the same iPCR approach
131 to obtain the derivative plasmid with $prgB_{188\text{-}1233:\ S442A,\ N444A}$. All constructed plasmids were
132 screened by PCR and verified by sequencing.
133
134 *Protein purification and crystallization*

135 PrgB was produced as previously described[32]. Briefly, $PrgB_{188-1233}$ was expressed with an N-
136 terminal hexa-histidine tag from pET-*prgB*$_{188-1233}$ or a C-terminal deca-histidine and GFP tag
137 from p7XC3GH-*prgB* in *E. coli* BL21(DE3). The cells were grown at 37 °C in TB medium
138 until they reached an OD_{600} of 1.5. Then the temperature was lowered to 18 °C and protein
139 production was induced by adding 0.5 mM IPTG. Cells were grown for 16 hours before
140 harvesting by centrifugation. Cells were resuspended in 20 mM HEPES/NaOH (pH 7.0), 300
141 mM NaCl, 30 mM Imidazole and 0.02 mg/ml DNase I and broken with a Constant cell disruptor
142 at 4 °C and 25 kPsi (Constant Systems). The cell lysate was clarified by centrifugation for 30
143 minutes at 30,000 x g, 4°C and incubated at 4 °C with Ni-NTA (Macherey-Nagel). The Ni-
144 NTA column was washed with 10 column volumes of 20 mM HEPES/NaOH (pH 7.0), 300
145 mM NaCl, 30-50 mM imidazole and bound proteins were eluted from the column with the
146 same buffer supplemented with 500 mM Imidazole. The histidine affinity tags and the GFP
147 when present, were cleaved off from the purified protein fractions by incubation with TEV
148 protease (for pET-*prgB*) or Prescission protease (for p7XC3GH-*prgB*) in a 1:100 ratio for 20
149 h at 4 °C. The cleaved proteins were loaded on a Superdex-200 Increase 10/300 GL column
150 (Cytiva) equilibrated in 20 mM HEPES/NaOH (pH 7.0) and 150 mM NaCl. The elution profile
151 showed two peaks corresponding to a PrgB dimer and monomer[32]. These two peak fractions
152 were pooled separately and concentrated on an Amicon Ultra Centrifugal Filter with a 30 kDa
153 cut-off (Merck-Millipore). 10% glycerol was added to the concentrated protein fractions, which
154 were subsequently flash frozen in liquid nitrogen and stored at -80 °C.
155
156 *Structure determination via X-ray crystallography*
157 Purified $PrgB_{188-1233}$ from the monomeric fraction, with a protein concentration of 15 mg/mL,
158 were thawed and used in crystallization trials. Crystals of $PrgB_{188-1233}$ were grown in 8-12
159 weeks, at 20 °C by sitting drop vapor diffusion in a condition with 0.2 M $CaCl_2$ and 20% PEG

160 3350 and a protein to reservoir ratio of 1:1 in the drop. Crystals were flash cooled in liquid
161 nitrogen. X-ray diffraction data of $PrgB_{188-1233}$ was collected on ID23-1 at the ESRF, France.
162 The data was processed using XDS[39]. The $PrgB_{188-1234}$ crystals belong to the monoclinic space
163 group $P2_12_12_1$ and contain two molecules in the asymmetric unit. The crystallographic phase
164 problem was solved using molecular replacement using PHASER, using the SspB homology
165 model of the Ig-domains as search models (PDB: 2WOY)[40]. Further building of the model was
166 conducted in COOT[41]. The structure was refined to 1.85 Å with crystallographic R_{work} and R_{free}
167 values of 20.9/24.6 using Refmac5 and PHENIX refine[42,43]. The final $PrgB_{188-1233}$ model
168 consists of residues 584-1234, and was validated using MolProbity[44]. Atomic coordinates and
169 structure factors have been deposited in the Protein Data Bank (PDB Code: 8BEG).
170
171 *Sample preparation for electron microscopy*
172 $PrgB_{188-1233}$ dimer fractions were thawed on ice and loaded on a Superdex 200 10/300 GL gel
173 filtration column (GE Healthcare) equilibrated in 20 mM HEPES/NaOH pH 7.0 and 150 mM
174 NaCl. Protein peak fractions, corresponding to the dimer, were diluted to 0.1-0.3 mg/mL and
175 for the DNA-bound structures 120 bp ssDNA (Table S1) was added in 1: 1.2 molar ratio
176 (protein:DNA) and samples were incubated for 15 minutes on ice. For both apo and DNA-
177 bound samples, 4 µl of sample was applied to glow discharged Quantifoil 300 mesh 1.2/1.3
178 (Quantifoil) grids at 4 °C and 90-100% humidity, blotted for 1 s with blot force -5, and plunge-
179 frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific).
180
181 *Cryo-EM data collection*
182 Cryo-EM data were collected on an FEI Titan Krios transmission electron microscope (Thermo
183 Fisher Scientific), operated at 300 keV that was equipped with a K2 direct electron detector.
184 Data was collected by the AFIS method using the EPU software V2.8.0 (Thermo scientific) at

185 a nominal magnification of 165,000x (0.82 Å pixel size). Data collection parameters are listed
186 in Table S2. A total number of 2787 movie stacks were collected for apo PrgB and 1670 for
187 DNA-bound PrgB.
188
189 *Cryo-EM data processing*
190 Cryo-EM data of apo PrgB and DNA-bound PrgB were processed in the same way, but
191 separately using cryoSPARC (v3.2.0-3.3.1)[45]. Beam-induced motion was corrected using
192 standard settings, where start frame 1 was excluded, followed by per-micrograph contrast
193 transfer function (CTF) estimation. For apo PrgB a subset of 170,826 particles from 500
194 micrographs were extracted with a box size of 384 Å. They were then picked using the blob
195 picking tool with a 100-300 Å particle diameter. Picked particles were subjected to consecutive
196 rounds of 2D classifications. Template particles were selected from representative 2D classes,
197 and the subsequent full dataset was picked using the template picker tool and the input of the
198 subset processing. PrgB with DNA was directly picked using blob picker with a 100-300 Å
199 particle diameter and standard settings and extracted with 384 pix. PrgB without and with DNA
200 were then separately subjected to 2D classifications resulting in a final number of 283,630 and
201 163,566 particles respectively. Particles from selected classes were combined and used in ab-
202 initio reconstruction. The initial volume was then subjected to homogenous 3D refinement and
203 the resolution was calculated using the gold standard Fourier shell correlation (FSC threshold,
204 0.143) and found to be 8 Å and 11 Å for the apo and DNA-bound structure, respectively. The
205 volumes of apo and DNA-bound PrgB have been deposited in the Electron Microscopy Data
206 Bank (EMDB Codes: EMD-16001 and EMD-16002).
207
208 The adhesion domain (PDB Code: 6EVU)[32] and stalk domain (PDB Code: 8BEG) were
209 initially docked into the EM volume using Chimera[46] and subsequently run through

210 Namdinator[47] using 10 Å resolution and standard settings. The output was then fitted in the EM
211 volume in Chimera (v 1.15rc[46]), where figures also were generated.
212
213 *Western blotting*
214 Cell pellets of each bacterial strain were harvested after induction and incubation overnight,
215 and treated with lysozyme buffer (10 mM Tris, pH 8.0, 1 mM EDTA, 25% sucrose, 15 mg/ml
216 lysozyme) for 30 min at 37°C. The lysozyme-treated bacterial lysates were then mixed with
217 protein loading dye and boiled at 100°C for 12 min. The prepared samples were run on 8%
218 SDS-PAGE, transferred to Western blot and probed with the PrgB antibody produced in rabbit.
219
220 *Aggregation assay*
221 *E. faecalis* strains were inoculated in BHI medium with the indicated antibiotics and cultured
222 O/N at 37°C. Overnight cultures were diluted in a 1:100 ratio in BHI with the indicated
223 antibiotics, 10 ng/mL cCF10, and 50 ng/ mL nisin and dispensed into polystyrene cuvettes
224 (Sarstedt) in 0.9 mL triplicates. These were incubated for 24 h at 37°C without agitation.
225 Afterwards, the optical density of each sample was determined at 600 nm both before (OD_{sup})
226 and after (OD_{mix}) vigorously mixing of the bacterial culture by pipetting. The autoaggregation
227 percentage was then calculated as follows: $100 \times [1 - (OD_{sup}/OD_{mix})]$[17,29].
228
229 *Biofilm assay*
230 *E. faecalis* strains were inoculated in BHI with the indicated antibiotics and kept 16 h at 37°C.
231 The next morning, they were diluted in a 1:100 ratio in BHI with the indicated antibiotics, 10
232 ng/mL cCF10, and 50 ng/mL nisin. 200 µl fractions were dispensed into a 96-well micro-titer
233 plate (Costar) with 8 replicates per strain. 200 µL BHI fractions were used as blanks. The 96-
234 well plate was then incubated aerobically at 37°C without agitation in a humidified chamber

235 for 24 h. The suspension was transferred to another 96-well plate to determine the optical
236 density at 600 nm (OD_{600}). The plate containing the biofilm was washed with distilled water
237 three times and then left to air dry at room temp for 2.5 h. The biofilm was stained with 100 µl
238 0.1% (w/v) safranin (Sigma) at room temp for 20 min, then washed three times with distilled
239 water and left to air dry at room temperature. Afterwards the absorbance was determined using
240 a plate reader (BMG Labtech) at 450 nm. Biofilm production was calculated as an index of
241 safranin staining of the cell biomass divided by absorbance of its optical density
242 (OD_{450}/OD_{600})[48].
243
244 *Conjugation assay*
245 Donor (OG1RF) and recipient (OG1ES) strains were inoculated in BHI with the indicated
246 antibiotics and incubated overnight at 37 °C with agitation. Overnight cultured strains were
247 refreshed in BHI without antibiotics in a 1:10 ratio, and donor strains were induced with 50
248 ng/mL nisin (Sigma). All strains were then incubated at 37 °C for 1h without agitation.
249 Afterwards each of the donor strains was mixed with the recipient cells in ratio of 1:10 and
250 mated at 37 °C statically for 3.5 h. These mixtures were then serially diluted with BHI and
251 plated out in triplicates on BHI agar plates supplemented with tetracycline and spectinomycin
252 (to select for donor cells), or with tetracycline, erythromycin, and streptomycin (to select for
253 transconjugants). Plates were incubated at 37 °C for 48 h, counted and enumerated for colony
254 forming units (CFU). The plasmid transfer rate was determined as CFU of transconjugant over
255 CFU of donor (Tc's/Donors)[32].
256
257 Electrophoretic Mobility Shift Assay (EMSA)
258 EMSA was carried out in the same way as previously described[32]. 0.1 to 20 µM PrgB (wild
259 type and variants) were mixed with 50 nM 100 bp long double stranded DNA[32]. Samples were

260 incubated for 1 hour at 20 °C before loading them onto a 6% TBE-based native acrylamide gel
261 for electrophoresis for 90 min at 50 V and 6 °C. Gels were subsequently stained with 3× GelRed
262 (Biotium) in distilled water for 30 minutes and imaged with a Chemidoc system (BioRad).
263 Quantification of the DNA bands after imaging was done in ImageLab (BioRad).
264
265 *Statistical analysis*
266 All shown *in vivo* data is from three independent experiments and was plotted and analyzed
267 using GraphPad Prism (version 5.0) (GraphPad Software). The indicated error is the standard
268 deviation over 3 biologically independent replicates.
269
270 **Results**
271 *$PrgB_{584-1233}$ contains 4 immunoglobulin-like domains*
272 Previous bioinformatics and structural analysis of PrgB proposed that PrgB consisted of 3
273 domains; the previously crystallized polymer adhesin domain responsible for eDNA and LTA
274 binding, and two domains with RGD (Arg-Gly-Asp) motifs[32]. However, when we reanalyzed
275 the PrgB sequence with the new structure-prediction tools that are available, such as
276 AlphaFold[49], it became clear that this model was partially incorrect. PrgB seems to consist of
277 an N-terminal disordered region (residues 35-197), followed by a newly identified coiled-coil
278 (COI) domain (residues 198-257), the previously crystallized polymer adhesin domain
279 (residues 261-558), a linker region (residues 559-582), 4 immunoglobulin (Ig)-like domains
280 (residues 583-1232) and finally the C-terminal disordered region containing the LPXTG motif
281 (residues 1263-1305) that gets anchored to the cell wall (Fig. 1A). The Ig-like domains seem
282 to come in pairs of two slightly different structures, denoted as CSA1-CSC1 (first pair) and
283 CSA2-CSC2 (second pair), as named in InterPro (CSA from IPR026345; adhesin isopeptide-
284 forming adherence domain, and CSC from IPR032300; cell-surface antigen C-terminal). To

285 verify this updated domain organization of PrgB, we set out to experimentally determine its
286 structure using a combination of X-ray crystallography and cryo-EM methods.
287
288 As described in Schmitt *et al*[32], we were not able to produce full-length PrgB in *E. coli*, but
289 instead produced and purified $PrgB_{188-1233}$. This version of the protein only lacks the disordered
290 N-terminal region and the LPXTG anchor and elutes from size exclusion chromatography in
291 two peaks corresponding to dimeric and a monomeric PrgB[7]. The monomeric fraction was
292 successfully used for crystallization trials. Crystals belonging to space group $P2_12_12_1$ appeared
293 after 8-12 weeks, diffracted to 1.85 Å and contained 2 molecules in the asymmetric unit. The
294 crystallographic phase problem was solved using molecular replacement with SspB (PDB:
295 2WOY) as a search model. Surprisingly, the resulting electron density lacked the previously
296 crystallized polymer adhesin domain of PrgB, and instead only contained residues 584-1233.
297 Likely, the polymer adhesin domain had been cleaved off in the crystallization drop before the
298 crystals were formed. The modelled protein indeed consists of four Immunoglobulin (Ig)-like
299 domains: two CSA and two CSC domains (Fig. 2A). Previous bioinformatics analysis showed
300 that various homologous adhesin proteins contain different numbers of Ig-like domains[16], but
301 a DALI[50] analysis of $PrgB_{584-1233}$ showed that there is no previously solved structure in the
302 PDB that contains four of these Ig-domains coupled together. There are, however, homologous
303 structures available with either 2 or 3 Ig-domains. The closest structural homologs are the C-
304 terminal parts of Antigen I/II proteins from oral Streptococci, e.g. the surface protein AspA
305 from *Streptococcus pyogenes*[51] (PDB code: 4OFQ), which has 3 Ig-domains and an r.m.s.d. to
306 the Ig-like domains from PrgB of 3.2 Å over 337 residues or the BspA protein from
307 *Streptococcus agalactiae* [52](PDB code: 4ZLP) which has 2 Ig-domains and an r.m.s.d. of 3.3
308 Å over 334 residues (see Table S3 for an overview of the highest ranked DALI hits). The

309 r.m.s.d. decreases to ca. 1 Å if the individual Ig-like domains are superimposed upon each
310 other.
311
312 In each of the Ig-like domains of PrgB, isopeptide bonds are formed between a lysine and an
313 asparagine, a bond which is further stabilized by an aspartic acid (Fig. 2B). This feature is also
314 present in various homologous Ig-like domains from Antigen I/II proteins[40,51,53]. There is
315 density in the conserved metal binding site of the CSA2 domain that has been suggested to
316 bind Ca^{2+} in the Antigen I/II proteins. Refinement of our structure indicated that Mg^{2+} was the
317 best fit to the density (Fig. 2C). The homologous C2 domains from AspA[51], Pas[54], SpaP[53] and
318 SspB[40], each have an extra structural feature termed the BAR (SspB adherence region) domain,
319 which mediates adherence in these proteins. This BAR domain is absent in PrgB (Fig. S1).
320
321 *DNA binding leads to large displacement of the polymer adhesin domain of PrgB*
322 As we didn't manage to crystallize $PrgB_{188-1233}$ without the loss of the polymer adhesin domain,
323 we determined its structure via cryo-EM and single particle analysis. To determine the
324 conformational changes of PrgB upon DNA binding, we collected two datasets: one without
325 and one with ssDNA (120 bp). This yielded volumes with an average resolution of 8 and 11 Å,
326 respectively. Even though dimeric fractions of PrgB were used for the grid preparations, the
327 resulting maps of PrgB are monomeric, indicating that the dimer fell apart. Despite the
328 relatively low resolution of the EM volume, mainly due to a preferred orientation of the protein
329 on the cryo-EM grids, we could dock in the X-ray structures: the previously solved polymer
330 adhesin domain (PDB code: 6EVU)[32] and the newly solved structure of $PrgB_{584-1233}$. The
331 resulting model of apo-PrgB shows that the polymer adhesin domain is folded back on top of
332 the Ig-like domains (Fig. 3A). In the DNA-bound PrgB model the polymer adhesin domain has

333 undergone a large displacement away from the Ig-like domains. There is also a larger volume
334 for this domain, indicating that DNA has indeed been bound (Fig. 3B).
335
336 *In vivo assays*
337 Based on the new structural insights for PrgB, we decided to study the importance of the newly
338 defined coiled-coil domain and the Ig-like domains. This was done by complementing *E.*
339 *faecalis* OG1RF pCF10Δ*prgB* with different *prgB* mutants (from a nisin-inducible plasmid)
340 and characterizing their phenotypes in cellular aggregation, biofilm formation, and plasmid
341 transfer efficiency. In line with previous experiments[17], complementing pCF10Δ*prgB* with
342 exogenous PrgB from the pMSP3545S vector rescues all phenotypes. Aggregation and biofilm
343 formation are even slightly increased as compared to wild-type pCF10 (Fig. 4A-C, column 1
344 and 3), possibly due to a slightly increased production of PrgB (Fig. S2, lane 1 and 3).
345
346 We found that PrgB without the newly identified coiled-coil domain could not rescue the
347 aggregation phenotype of the deletion strain (Fig. 4A, column 4). Western blot analysis showed
348 reduced levels of $PrgB_{\Delta COI}$ as compared to wild type (Fig. S2) probably due to a decreased
349 stability, which could explain the observed decrease in aggregation. Deletion of either the
350 CSA1 or the CSC2 domain did not affect PrgB-mediated aggregation (Fig. 4A, column 8 and
351 9). However, complementation with PrgBΔCSA2-CSC2 could only partially rescue the
352 aggregation phenotype of the *E. faecalis* OG1RF pCF10Δ*prgB* strain (Fig. 4A, column 7) and
353 no rescue at all was seen in the PrgB variants with both the CSA1 and CSC1 domain deleted
354 or without any Ig-like domains (CSA1-CSC1-CSA2-CSC2) (Fig. 4A, column 5, and 6).
355
356 As expected[17], deletion of *prgB* also lead to a large decrease in biofilm formation (Fig. 4B). In
357 line with our observations from the aggregation assays, PrgB with a deletion of either the

358 coiled-coil domain or more than a single Ig-like domain could not rescue the *E. faecalis*
359 OG1RF pCF10Δ*prgB* biofilm formation phenotype. Only exogenous expression of $PrgB_{\Delta CSC2}$
360 can rescue the level of biofilm formation, but only to the level found in OG1RF pCF10, not to
361 the level of exogenously expressed wild-type *prgB* (Fig 4B, column 1, 3 and 9). Intriguingly,
362 the expression of exogenous $PrgB_{\Delta CSA1}$ in the OG1RF pCF10Δ*prgB* background did not restore
363 biofilm formation, while it did restore the aggregation phenotype (Fig. 4A and 4B, column 8).
364
365 In the conjugation assay, exogenous expression of the PrgB variants that failed to rescue the
366 aggregation phenotype of the OG1RF pCF10Δ*prgB* strain ($PrgB_{\Delta COI}$, $PrgB_{\Delta CSA1\text{-}CSC2}$,
367 $PrgB_{\Delta CSA1\text{-}CSC1}$ and $PrgB_{\Delta CSA2\text{-}CSC2}$) were also found to have a decreased plasmid transfer
368 efficiency (Fig. 4C, column 4-7). However, $PrgB_{\Delta CSA1}$ and $PrgB_{\Delta CSC2}$ could only partially
369 rescue the conjugation rate of the OG1RF pCF10Δ*prgB* strain (Fig. 4C, column 8-9), while
370 they did rescue the aggregation phenotype.
371
372 *The conserved binding cleft in the polymer adhesin domain is essential for conjugation and*
373 *biofilm formation, but not for aggregation*
374 To investigate the role of the binding cleft in the polymer adhesin domain, we introduced
375 single, double, and triple mutations to alter its conserved residues. The resulting *prgB* mutants
376 were exogenously expressed in the background of pCF10Δ*prgB* for functional
377 complementation; or in the background of wild-type pCF10 to test for any potential dominant
378 negative effects as previously observed for $PrgB_{\Delta 246\text{-}558}$ (deletion of the polymer adhesin
379 domain)[32]. Notably, all the resulting *prgB* mutants fully restored the aggregation phenotype of
380 pCF10Δ*prgB* (Fig. 5A), but did not rescue the defective biofilm formation, nor the reduced
381 conjugation efficiency (Fig. 5B and C, column 2-6). No dominant negative effects on
382 aggregation or biofilm formation could be observed when these variants were expressed in the

383 wild-type pCF10 background (Fig. 5A and B, column 7-10), although slightly reduced
384 conjugation rates were observed (Fig. 5C, column 7-10). We have previously shown that the
385 polymer adhesin domain binds eDNA and that this binding correlates with both biofilm
386 formation and conjugation efficiency. Therefore, we wanted to test whether eDNA binding was
387 affected in these PrgB variants. To do so, we purified both wild-type $PrgB_{188-1235}$ and the double
388 mutant $PrgB_{188-1233}$:S442A,N444A to compare their DNA binding affinities via electrophoretic
389 mobility shift assays (EMSA). The results indicate that the introduced changes in PrgB did not
390 affect its ability to bind eDNA, as the affinity for $PrgB_{188-1233}$ and $PrgB_{188-1233}$:S442A,N444A
391 were the same within experimental error (Fig. S3).
392
393 **Discussion**
394 The presented data provides important insights for a widespread virulence factor in Gram-
395 positive bacteria, since genes encoding PrgB homologs exists on a large number of conjugative
396 plasmids[16]. In this study, we expand our structural knowledge of PrgB beyond the polymer
397 adhesin domain, to now encompass almost the entire protein.
398
399 The crystal structure of $PrgB_{583-1233}$ shows that this part of PrgB consists of four tandemly
400 arranged immunoglobulin (Ig)-like domains. These Ig-like domains show a high degree of
401 structural homology to Streptococcal surface proteins, usually found in the oral cavity[16].
402 These homologous proteins have been indicated to bind various molecules, such as fimbria,
403 collagen and salivary agglutinin (SAG, also designated as glycoprotein 340), and these binding
404 interactions have been shown to be vital for their function[40,54–57]. However, we have not found
405 any evidence that the Ig-like domains of PrgB bind to a specific substrate in our own
406 experiments, nor have we found this in other reports. PrgB also does not contain a BAR
407 domain, which is crucial for stable interactions between e.g. SspB from *S. gordonni* and Mfa-

408 1 of *P. gingivalis* [58,59]. However, since Ig-like domains are known to bind a large variety of
409 ligands[60], we don't exclude the possibility that ligands for the Ig-like domains in PrgB will be
410 found in the future. However, the only ligands that have been verified to interact with PrgB so
411 far are eDNA and LTA, which have high affinity to the polymer adhesin domain[32].
412
413 The crystal structure of the Ig-like domains from PrgB, $PrgB_{583-1233}$, was complemented by
414 single particle analysis of $PrgB_{188-1233}$ via cryo-EM (Fig. 3). Despite the low resolution of the
415 EM volumes, we could dock in the high-resolution crystal structures of the Ig-like domains and
416 the polymer adhesin domain. eDNA binding is a key step in PrgB mediated cellular aggregation
417 and biofilm formation[32], and the models obtained from our cryo-EM volumes provided the first
418 mechanistical insight into the conformational changes that occur during this step. The polymer
419 adhesin domain is nestled to the side of the Ig-like domains (Fig. 3A), likely reflecting how
420 PrgB is present on the outside of the cell wall in the absence of a binding partner. Possibly, the
421 folding back of the polymer adhesin domain is protecting it from proteolytic cleavage (see
422 below). Our model of PrgB bound to DNA suggests that the polymer adhesin domain gets
423 displaced from the Ig-like domains upon eDNA binding (Fig. 3B). This would expose the
424 positively charged surface of the polymer adhesin domain and allow it to wrap the eDNA
425 around itself to condense it, as we observed in previous experiments[32,33]. We speculate that this
426 in turn could also expose the PrgA-binding site, predicted to be located in the flexible region
427 between the polymer adhesin domain and the CSA1 domain[15]. This would allow PrgA-
428 mediated cleavage and removal of the polymer adhesin domain[15]. How this cleavage is
429 regulated is not yet understood, but the folding back of the polymer adhesin domain onto the
430 Ig-like domains in the absence of substrate could be part of this regulation.
431

432 We have now obtained a structural basis to interpret almost all phenotypic data that is available
433 for PrgB. Unfortunately, most of the mutants that were previously described did not correlate
434 well with the newly determined domain boundaries. Therefore, we decided to create specific
435 deletion mutants that were based on the new PrgB structure to determine the role of the various
436 domains. At the N-terminus of PrgB, before the polymer adhesin domain, there is a predicted
437 coiled-coil region (Fig. S4A) that we wanted to investigate. To our surprise, the expression of
438 $PrgB_{\Delta COI}$ could not rescue any of the aggregation, biofilm formation or conjugation phenotypes
439 from a *prgB* deletion strain. However, in our experiments $PrgB_{\Delta COI}$ was only produced at low
440 levels (Fig. S2), indicating that the coiled-coil region might be important for protein production
441 and/or stability. Deletion of all Ig-like domains ($PrgB_{\Delta CSA1-\Delta CSC2}$) renders the protein incapable
442 to support aggregation, biofilm formation and conjugation. However, exogenous expression of
443 *prgB* with single Ig-like domain deletions, $prgB_{\Delta CSA1}$ and $prgB_{\Delta CSC2}$ in the pCF10Δ*prgB*
444 background, restore cellular aggregation, while they do not rescue biofilm formation and
445 conjugation (Fig. 4). This was unexpected, as the polymer adhesin domain is predicted to
446 mediate all the functions that were tested in these assays: aggregation, biofilm formation and
447 conjugation[17,32]. Our results, however, indicate that it is important to have all Ig-like domains
448 present and properly folded. The cellular aggregation assays seem to indicate that PrgB can
449 function when at least 3 Ig-like domains are present, as expression of both *prgB*ΔCSA1 and
450 *prgB*ΔCSC2 can complement pCF10Δ*prgB*, but unfortunately this assay is not suitable to
451 detect small changes. Based on the biofilm formation and conjugation efficiency assays, which
452 are more sensitive, we therefore conclude that even removing a single Ig-like domain strongly
453 decreases the function of PrgB.
454
455 The conserved Ser-Asn-Glu site in the negatively charged cleft of the polymer adhesin domain
456 intrigued us, as its function is unknown. Any changes that we made in these conserved residues

457 resulted in a PrgB variant that didn't facilitate biofilm formation or conjugation (Fig. 5).
458 Surprisingly these PrgB variants did fully support cellular aggregation. This is thought-
459 provoking, since various literature has shown that the PrgB functions in cellular aggregation,
460 biofilm formation and conjugation are strongly correlated. However, even a single point
461 mutation in this conserved site produced a PrgB variant that completely separates the cellular
462 clumping phenotype from biofilm formation and conjugation. *In vitro* experiments showed that
463 these point mutations did not impair PrgB binding to eDNA (Fig S3). A similar phenotypic
464 pattern was observed with $PrgB_{\Delta CSA1}$ and $PrgB_{\Delta CSC2}$ which both could fully rescue aggregation
465 but not biofilm formation. Our data therefore strongly indicates that PrgB has additional role(s),
466 besides mediating cellular aggregation, to further support conjugation and biofilm formation.
467 We therefore propose that PrgB performs at least one, currently unknown, additional function
468 besides binding to eDNA and/or LTA from the cell wall of the recipient cell. It is highly likely
469 that at least one of these additional functions is mediated by the conserved site in the polymer
470 adhesin domain.
471
472 As described in the introduction, there is a plethora of *prgB* mutants made (see Fig. S4) and
473 phenotypically analyzed, predominantly by the group of Prof. Gary Dunny. In supplementary
474 table S4 we provide a summary of all *prgB* mutants that we have identified in the literature and
475 our brief reinterpretation based on our new structural knowledge. Below, we will discuss a
476 selected number of these mutants in detail.
477
478 As expected, most mutants that have an insertion or a mutation in the polymer adhesin domain
479 (Fig. S4A) show a loss of protein function. Shortly, insertions at amino acids (a.a.) 358 and
480 359 are on the surface of the protein and likely leads to steric clashes that prevent the domain
481 from binding to eDNA and LTA, whereas insertions at a.a. 439, 473, 517, and 546 are all in

482 secondary structures that are central parts of the polymer adhesin domain and therefore are
483 likely to disrupt folding of this domain.
484
485 The RGD motifs in PrgB were previously proposed to be involved in integrin binding and to
486 promote adherence to human neutrophils, as well as internalization in cultured intestinal
487 epithelial cells[25,61]. The new domain classification showed that these two RGD motifs are in
488 CSA1 and CSA2, respectively (Fig. 1A). The structure that we determined shows that these
489 two sequence motifs are not surface exposed, but instead play an important role in the structural
490 integrity of the interfaces between CSA1 and CSC1, and between CSA2 and CSC2 (Fig. 6).
491 Mutations in these motifs would thus very likely destabilize the folding of the tandem Ig-
492 domains, which would explain the decreased PrgB biofilm formation observed in these
493 strains[18,23]. Thus, our new data strongly argues against the previously proposed direct binding
494 interaction between the RGD sequences and integrins of host origin.
495
496 Deletion of residues 993-1138 leads to a loss of about half of both the CSA2 and CSC2
497 domains. Therefore, we were surprised to find that this mutant was described to behave like
498 wild type in aggregation assays[31]. Possibly the remaining parts of CSA2 and CSC2 form a
499 (unfolded) linker region of similar length to the tandem CSA2-CSC2 structure, which allows
500 PrgB to retain its function of promoting aggregation. Similarly, $PrgB_{\Delta 668\text{-}1138}$, corresponding to
501 a complete removal of CSC1 and CSA2 and approximately half of both CSA1 and CSC2
502 domains, could still support PrgB mediated aggregation. Potentially the remaining residues (ca
503 180 amino acids) could also form an unfolded linker region allowing for the variant to retain
504 its aggregation phenotype. For $PrgB_{\Delta 993\text{-}1138}$ and $PrgB_{\Delta 668\text{-}1138}$, unfortunately no biofilm
505 formation or conjugation efficiency were reported, but based on our results it is likely that those
506 capabilities would have been severely compromised.

507

508 Taking past and present results into account combined with our new structural insights, we
509 propose a new mechanistic model for the function of PrgB. Our EM data indicates that the Ig-
510 domains provide structural support for the polymer adhesin domain in the absence of bound
511 ligand(s) and our *in vivo* data show that removal of one of the Ig-like domains does not largely
512 affect PrgB function. Previous data also indicates that parts of the Ig-like domains can be
513 deleted without affecting the function of PrgB. We therefore hypothesize that the Ig-like
514 domains provide two important features to the protein. First, a rigid stalk that is needed to
515 present the polymer adhesin domain at the correct distance from the cell wall. Second,
516 providing protection from PrgA-mediated cleavage when the polymer adhesin domain is folded
517 on top of the Ig-like domains. Binding of eDNA (and likely also LTA) releases the polymer
518 adhesin domain from this stalk domain, and with that likely also exposes the predicted PrgA
519 proteolysis site in the linker region between the polymer adhesin domain and the CSA1
520 domain[15,16]. The exact distance of the polymer adhesin domain from the cell wall does not seem
521 important for LTA-binding, since aggregation can still take place even with partially disrupted
522 Ig-like domains. However, the length of the stalk may be very important when it comes to
523 facilitating both biofilm formation and conjugation. This indicates that effective mating-pair
524 formation in conjugation may require a PrgB-mediated function distinct from pure aggregation,
525 something that is further shown by the findings of the mutations in the conserved site in the
526 polymer adhesin domain.

527

528 Besides providing a structural basis to explain about 30 years of work on PrgB, we here also
529 uncovered that the conserved Ser-Asn-Glu site in the polymer adhesin domain likely provides
530 additional functionality to PrgB that is needed for optimal biofilm formation and conjugation,
531 but that does not affect cellular aggregation. To fully investigate the function of this conserved

site in PrgB and other homologous virulence factors from Gram-positive bacteria, remains an exciting question to address in future research.

Acknowledgments

The authors would like to thank Prof. Gary Dunny for very fruitful discussions about the results and Dr. Karim Rafie and Annika Breidenstein for discussions about EM data processing. We acknowledge MAX IV Laboratory for time on Beamline BioMax under Proposal 20180236. Research conducted at MAX IV, a Swedish national user facility, is supported by the Swedish Research council under contract 2018-07152, the Swedish Governmental Agency for Innovation Systems under contract 2018-04969, and Formas under contract 2019-02496. We also acknowledge the synchrotrons Swiss Light Source (Paul Scherrer Institute, Switzerland) for time at beamline PX1 and the ESRF (France) for time at beamlines ID23 and ID30. The EM data was collected at the Umeå Core Facility for Electron Microscopy, a node of the Cryo-EM Swedish National Facility, funded by the Knut and Alice Wallenberg, Family Erling Persson and Kempe Foundations, SciLifeLab, Stockholm University and Umeå University. This work was supported by grants from the Swedish Research Council (2016-03599), Knut and Alice Wallenberg Foundation, Kempestiftelserna (SMK-1762 & SMK-1869) and Carl-Tryggers stiftelse (CTS 18:39) to R.P-A.B.

CRediT statement

Wei-Sheng Sun: Conceptualization, Investigation, Writing - Original Draft, Writing – Revision. Lena Lassinantti: Investigation, Writing - Original Draft, Writing – Revision. Michael Järvå: Conceptualization, Investigation, Andreas Schmitt: Investigation. Josy ter

Beek: Investigation, Writing - Original Draft, Writing – Revision. Ronnie P-A Berntsson: Conceptualization, Writing - Original Draft, Writing – Revision, Supervision, Funding acquisition.

561 **References**

562 1. Gilmore, M. S., Lebreton, F. & van Schaik, W. Genomic transition of enterococci from
563 gut commensals to leading causes of multidrug-resistant hospital infection in the
564 antibiotic era. *Current Opinion in Microbiology* vol. 16 10–16 (2013).

565 2. Hidron, A. I., Edwards, J. R., Patel, J., Horan, T. C., Sievert, D. M., Pollock, D. A., *et*
566 *al.* NHSN annual update: antimicrobial-resistant pathogens associated with healthcare-
567 associated infections: annual summary of data reported to the National Healthcare
568 Safety Network at the Centers for Disease Control and Prevention, 2006-2007. *Infect.*
569 *Control Hosp. Epidemiol.* **29**, 996–1011 (2008).

570 3. Palmer, K. L., Kos, V. N. & Gilmore, M. S. Horizontal gene transfer and the genomics
571 of enterococcal antibiotic resistance. *Current Opinion in Microbiology* vol. 13 632–
572 639 (2010).

573 4. Palmer, K. L. & Gilmore, M. S. Multidrug-resistant enterococci lack CRISPR-cas.
574 *MBio* **1**, e00227-10 (2010).

575 5. Trokter, M. & Waksman, G. Translocation through the conjugative Type 4 secretion
576 system requires unfolding of its protein substrate. *J. Appl. Microbiol.* **200**, e00615-17
577 (2018).

578 6. Cabezón, E., Ripoll-Rozada, J., Peña, A., de la Cruz, F. & Arechaga, I. Towards an
579 integrated model of bacterial conjugation. *FEMS Microbiol. Rev.* **39**, 81–95 (2015).

580 7. Koraimann, G. Spread and Persistence of Virulence and Antibiotic Resistance Genes:
581 A Ride on the F Plasmid Conjugation Module. *EcoSal Plus* **8**, (2018).

582 8. Grohmann, E., Christie, P. J., Waksman, G. & Backert, S. Type IV secretion in Gram-
583 negative and Gram-positive bacteria. *Molecular Microbiology* vol. 107 455–471
584 (2018).

585 9. Hirt, H., Manias, D. A., Bryan, E. M., Klein, J. R., Marklund, J. K., Staddon, J. H., *et*

586 *al.* Characterization of the pheromone response of the Enterococcus faecalis
587 conjugative plasmid pCF10: Complete sequence and comparative analysis of the
588 transcriptional and phenotypic responses of pCF10-containing cells to pheromone
589 induction. *J. Bacteriol.* **187**, 1044–1054 (2005).
590 10. Dunny, G. M. & Berntsson, R. P.-A. Enterococcal sex pheromones: Evolutionary
591 pathways to complex, two-signal systems. *J. Bacteriol.* **198**, 1556–1562 (2016).
592 11. Chandler, J. R., Hirt, H. & Dunny, G. M. A paracrine peptide sex pheromone also acts
593 as an autocrine signal to induce plasmid transfer and virulence factor expression in
594 vivo. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15617–15622 (2005).
595 12. Dunny, G. M. & Berntsson, R. P. A. Enterococcal sex pheromones: Evolutionary
596 pathways to complex, two-signal systems. *J. Bacteriol.* **198**, 1556–1562 (2016).
597 13. Dunny, G. M. Enterococcal Sex Pheromones: Signaling, Social Behavior, and
598 Evolution. *Annu. Rev. Genet.* **47**, 457–482 (2013).
599 14. Lassinantti, L., Camacho, M. I., Erickson, R. J. B., Willett, J. L. E., De Lay, N. R., ter
600 Beek, J., *et al.* Enterococcal PrgU Provides Additional Regulation of Pheromone-
601 Inducible Conjugative Plasmids. *mSphere* **6**, e00264--21 (2021).
602 15. Schmitt, A., Hirt, H., Järvå, M. A., Sun, W. S., ter Beek, J., Dunny, G. M., *et al.*
603 Enterococcal PrgA Extends Far Outside the Cell and Provides Surface Exclusion to
604 Protect against Unwanted Conjugation. *J. Mol. Biol.* **432**, 5681–5695 (2020).
605 16. Järvå, M. A., Hirt, H., Dunny, G. M. & Berntsson, R. P. Polymer Adhesin Domains in
606 Gram-Positive Cell Surface Proteins. *Front. Microbiol.* **11**, 599899 (2020).
607 17. Bhatty, M., Cruz, M. R., Frank, K. L., Laverde Gomez, J. A., Andrade, F., Garsin, D.
608 A., *et al.* Enterococcus faecalispCF10-encoded surface proteins PrgA, PrgB
609 (aggregation substance) and PrgC contribute to plasmid transfer, biofilm formation and
610 virulence. *Mol. Microbiol.* **95**, 660–667 (2015).

611 18. Chuang, O. N., Schlievert, P. M., Wells, C. L., Manias, D. A., Tripp, T. J. & Dunny,
612 G. M. Multiple functional domains of Enterococcus faecalis aggregation substance
613 Asc10 contribute to endocarditis virulence. *Infect. Immun.* **77**, 539–48 (2009).
614 19. Olmsted, S. B., Kao, S. M., van Putte, L. J., Gallo, J. C. & Dunny, G. M. Role of the
615 pheromone-inducible surface protein Asc10 in mating aggregate formation and
616 conjugal transfer of the Enterococcus faecalis plasmid pCF10. *J. Appl. Microbiol.* **173**,
617 7665–7672 (1991).
618 20. Olmsted, S. B., Erlandsen, S. L., Dunny, G. M. & Wells, C. L. High-resolution
619 visualization by field emission scanning electron microscopy of Enterococcus faecalis
620 surface proteins encoded by the pheromone-inducible conjugative plasmid pCF10. *J.*
621 *Appl. Microbiol.* **175**, 6229–6237 (1993).
622 21. Süssmuth, S. D., Muscholl-Silberhorn, A., Wirth, R., Susa, M., Marre, R. &
623 Rozdzinski, E. Aggregation substance promotes adherence, phagocytosis, and
624 intracellular survival of Enterococcus faecalis within human macrophages and
625 suppresses respiratory burst. *Infect. Immun.* **68**, 4900–4906 (2000).
626 22. Schlievert, P. M., Chuang-Smith, O. N., Peterson, M. L., Cook, L. C. C. & Dunny, G.
627 M. Enterococcus faecalis Endocarditis Severity in Rabbits Is Reduced by IgG Fabs
628 Interfering with Aggregation Substance. *PLoS One* **5**, e13194 (2010).
629 23. Chuang-Smith, O. N., Wells, C. L., Henry-Stanley, M. J. & Dunny, G. M.
630 Acceleration of Enterococcus faecalis biofilm formation by aggregation substance
631 expression in an Ex Vivo model of cardiac valve colonization. *PLoS One* **5**, e15798
632 (2010).
633 24. Rakita, R. M., Vanek, N. N., Jacques-Palaz, K., Mee, M., Mariscalco, M. M., Dunny,
634 G. M., *et al.* Enterococcus faecalis bearing aggregation substance is resistant to killing
635 by human neutrophils despite phagocytosis and neutrophil activation. *Infect. Immun.*

636 **67**, 6067–6075 (1999).

637 25. Vanek, N. N., Simon, S. I., Jacques-Palaz, K., Mariscalco, M. M., Dunny, G. M. &
638 Rakita, R. M. Enterococcus faecalis aggregation substance promotes opsonin-
639 independent binding to human neutrophils via a complement receptor type 3-mediated
640 mechanism. *FEMS Immunol. Med. Microbiol.* **26**, 49–60 (1999).

641 26. Hirt, H., Greenwood-Quaintance, K. E., Barnes, A. M. T., Karau, M. J., Till, L. M.,
642 Palzer, E., *et al.* Dynamics of plasmid-mediated niche invasion, immunity to invasion,
643 and pheromone-inducible conjugation in the murine gastrointestinal tract. *Nat.*
644 *Commun.* **13**, 1377 (2022).

645 27. MUSCHOLL, A., GALLI, D., WANNER, G. & WIRTH, R. Sex pheromone plasmid
646 pAD1-encoded aggregation substance of Enterococcus faecalis is positively regulated
647 in trans by traE1. *Eur. J. Biochem.* **214**, 333–8 (1993).

648 28. Galli, D., Friesenegger, A. & Wirth, R. Transcriptional control of sex-pheromone-
649 inducible genes on plasmid pAD1 of Enterococcus faecalis and sequence analysis of a
650 third structural gene for (pPD1 -encoded) aggregation substance. *Mol. Microbiol.* **6**,
651 1297–1308 (1992).

652 29. Waters, C. M. & Dunny, G. M. Analysis of functional domains of the Enterococcus
653 faecalis pheromone-induced surface protein aggregation substance. *J. Bacteriol.* **183**,
654 5659–67 (2001).

655 30. Waters, C. M., Wells, C. L. & Dunny, G. M. The Aggregation Domain of Aggregation
656 Substance, Not the RGD Motifs, Is Critical for Efficient Internalization by HT-29
657 Enterocytes. *Infect. Immun.* **71**, 5682–5689 (2003).

658 31. Waters, C. M., Hirt, H., McCormick, J. K., Schlievert, P. M., Wells, C. L. & Dunny,
659 G. M. An amino-terminal domain of Enterococcus faecalis aggregation substance is
660 required for aggregation, bacterial internalization by epithelial cells and binding to

661 lipoteichoic acid. *Mol. Microbiol.* **52**, 1159–71 (2004).
662 32. Schmitt, A., Jiang, K., Camacho, M. I., Jonna, V. R., Hofer, A., Westerlund, F., *et al.*
663 PrgB promotes aggregation, biofilm formation, and conjugation through DNA binding
664 and compaction. *Mol. Microbiol.* **109**, 291–305 (2018).
665 33. Kohler, V., Keller, W. & Grohmann, E. Enterococcus adhesin PrgB facilitates type IV
666 secretion by condensation of extracellular DNA. *Molecular Microbiology* vol. 109
667 263–267 (2018).
668 34. Mieher, J. L., Larson, M. R., Schormann, N., Purushotham, S., Wu, R., Rajashankar,
669 K. R., *et al.* Glucan binding protein C of *Streptococcus mutans* mediates both sucrose-
670 independent and sucrose-dependent adherence. *Infect. Immun.* **86**, 1–18 (2018).
671 35. Forsgren, N., Lamont, R. J. & Persson, K. Crystal structure of the variable domain of
672 the Streptococcus gordonii surface protein SspB. *Protein Sci.* **18**, 1896–1905 (2009).
673 36. Bae, T., Kozlowicz, B. & Dunny, G. M. Two targets in pCF10 DNA for PrgX binding:
674 Their role in production of Qa and prgX mRNA and in regulation of pheromone-
675 inducible conjugation. *J. Mol. Biol.* **315**, 995–1007 (2002).
676 37. Laverde Gomez, J. A., Bhatty, M. & Christie, P. J. PrgK, a Multidomain
677 Peptidoglycan Hydrolase, Is Essential for Conjugative Transfer of the Pheromone-
678 Responsive Plasmid pCF10. *J. Appl. Microbiol.* **196**, 527–539 (2014).
679 38. Geertsma, E. R. & Dutzler, R. A versatile and efficient high-throughput cloning tool
680 for structural biology. *Biochemistry* **50**, 3272–3278 (2011).
681 39. Kabsch, W. XDS. *Acta Crystallogr. Sect. D* **66**, 125–132 (2010).
682 40. Forsgren, N., Lamont, R. J. & Persson, K. Two intramolecular isopeptide bonds are
683 identified in the crystal structure of the Streptococcus gordonii SspB C-terminal
684 domain. *J. Mol. Biol.* **397**, 740–751 (2010).
685 41. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of

686 Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
687 42. Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model
688 anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D. Biol.*
689 *Crystallogr.* **57**, 122–33 (2001).
690 43. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J.,
691 Moriarty, N. W., *et al.* PHENIX : building new software for automated
692 crystallographic structure determination. *Acta Crystallogr. Sect. D Biol. Crystallogr.*
693 **58**, 1948–1954 (2002).
694 44. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. a, Immormino, R. M., Kapral, G.
695 J., *et al.* MolProbity: all-atom structure validation for macromolecular crystallography.
696 *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 12–21 (2010).
697 45. Punjani, A., Brubaker, M. A. & Fleet, D. J. Building Proteins in a Day: Efficient 3D
698 Molecular Structure Estimation with Electron Cryomicroscopy. *IEEE Trans. Pattern*
699 *Anal. Mach. Intell.* **39**, 706–718 (2017).
700 46. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng,
701 E. C., *et al.* UCSF Chimera - A visualization system for exploratory research and
702 analysis. *J. Comput. Chem.* **25**, 1605–12 (2004).
703 47. Kidmose, R. T., Juhl, J., Nissen, P., Boesen, T., Karlsen, J. L. & Pedersen, B. P.
704 Namdinator - Automatic molecular dynamics flexible fitting of structural models into
705 cryo-EM and crystallography experimental maps. *IUCrJ* **6 (Pt 4)**, 526–531 (2019).
706 48. Willett, J. L. E., Ji, M. M. & Dunny, G. M. Exploiting biofilm phenotypes for
707 functional characterization of hypothetical genes in Enterococcus faecalis. *npj Biofilms*
708 *Microbiomes* **5**, 23 (2019).
709 49. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., *et al.*
710 Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589

711 (2021).
712 50. Holm, L. DALI and the persistence of protein shape. *Protein Sci.* **29**, 128–140 (2020).
713 51. Hall, M., Nylander, A., Jenkinson, H. F. & Persson, K. Structure of the C-terminal
714 domain of AspA (antigen I/II-family) protein from Streptococcus pyogenes. *FEBS*
715 *Open Bio* **4**, 283–289 (2014).
716 52. Rego, S., Heal, T. J., Pidwill, G. R., Till, M., Robson, A., Lamont, R. J., *et al.*
717 Structural and functional analysis of cell wall-anchored polypeptide adhesin BspA in
718 Streptococcus agalactiae. *J. Biol. Chem.* **291**, 15985–6000 (2016).
719 53. Larson, M. R., Rajashankar, K. R., Crowley, P. J., Kelly, C., Mitchell, T. J., Brady, L.
720 J., *et al.* Crystal Structure of the C-terminal Region of Streptococcus mutans Antigen
721 I/II and Characterization of Salivary Agglutinin Adherence Domains. *J. Biol. Chem.*
722 **286**, 21657–21666 (2011).
723 54. Mieher, J. L., Schormann, N., Wu, R., Patel, M., Purushotham, S., Wu, H., *et al.*
724 Structure-function characterization of Streptococcus intermedius surface antigen Pas.
725 *J. Bacteriol.* **203**, e0017521 (2021).
726 55. Brady, L. J., Maddocks, S. E., Larson, M. R., Forsgren, N., Persson, K.,
727 Deivanayagam, C. C., *et al.* The changing faces of Streptococcus antigen I/II
728 polypeptide family adhesins. *Mol. Microbiol.* **77**, 276–286 (2010).
729 56. Maddocks, S. E., Wright, C. J., Nobbs, A. H., Brittan, J. L., Franklin, L., Strömberg,
730 N., *et al.* Streptococcus pyogenes antigen I/II-family polypeptide AspA shows
731 differential ligand-binding properties and mediates biofilm formation. *Mol. Microbiol.*
732 **81**, 1034–49 (2011).
733 57. Franklin, L., Nobbs, A. H., Bricio-Moreno, L., Wright, C. J., Maddocks, S. E., Sahota,
734 J. S., *et al.* The AgI/II Family Adhesin AspA Is Required for Respiratory Infection by
735 Streptococcus pyogenes. *PLoS One* **8**, e62433 (2013).

736 58. Purushotham, S. & Deivanayagam, C. The Calcium-induced Conformation and
737 Glycosylation of Scavenger-rich Cysteine Repeat (SRCR) Domains of Glycoprotein
738 340 Influence the High Affinity Interaction with Antigen I/II Homologs. *J. Biol. Chem.*
739 **289**, 21877–21887 (2014).
740 59. Daep, C. A., Novak, E. A., Lamont, R. J. & Demuth, D. R. Structural dissection and in
741 vivo effectiveness of a peptide inhibitor of porphyromonas gingivalis adherence to
742 streptococcus gordonii. *Infect. Immun.* **79**, 67–74 (2011).
743 60. Halaby, D. M. & Mornon, J. P. E. The immunoglobulin superfamily: An insight on its
744 tissular, species, and functional diversity. *J. Mol. Evol.* **46**, 389–400 (1998).
745 61. Olmsted, S. B., Dunny, G. M., Erlandsen, S. L. & Wells, C. L. A Plasmid-Encoded
746 Surface Protein On Enterococcus Faecalis Augments Its Internalization By Cultured
747 Intestinal Epithelial Cells. *J. Infect. Dis.* **170**, 1549–1556 (1994).
748
749

750 **Tables**

751

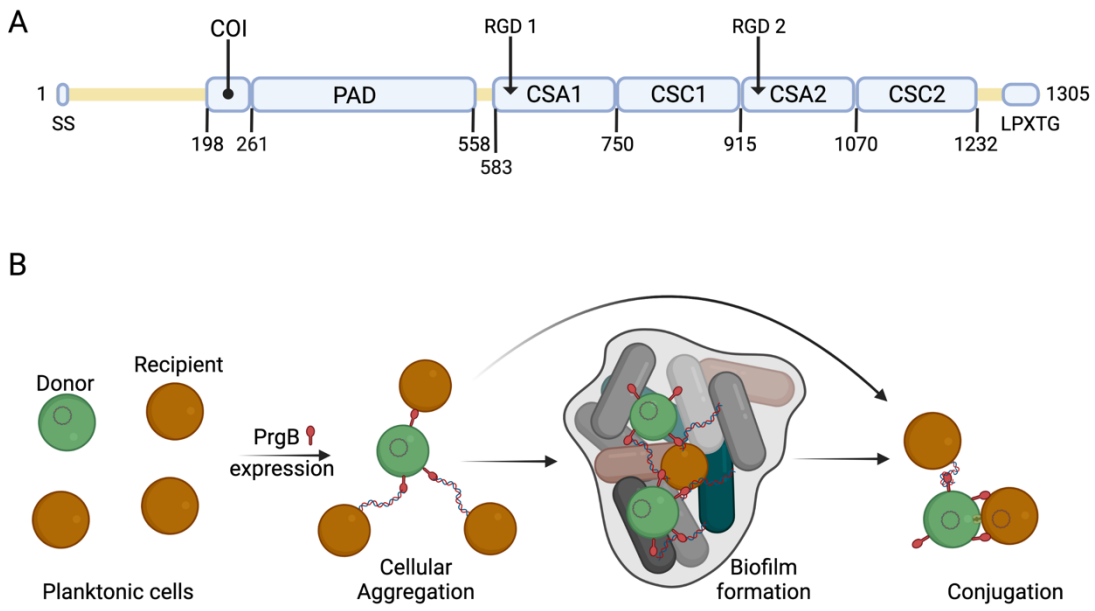
752 **Table 1**. X-ray data collection and refinement statistics. Values within parenthesis correspond
753 to the highest resolution shell.

Data collection summary	$PrgB_{584-1233}$
Space group	$P2_12_12_1$
Cell dimensions	
a, b, c (Å)	76.0, 87.5, 224.8
α, β, γ (°)	90, 90, 90
Resolution (Å)	47.3 – 1.84 (1.91 – 1.84)
Completeness (%)	99.6 (99.5)
R_{meas} (%)	0.03 (0.99)
I/σ (I)	12.1 (1.0)
CC(1/2)*	1 (0.67)
Redundancy	2.0 (2.0)
No. unique reflections	130301 (12877)
Refinement summary	
Resolution (Å)	47.3 – 1.84
R_{work} (%)	20.9
R_{free} (%)	24.5
Number of atoms	
protein	10080
water	816
other ligands	2
B-factors	
protein	44.9
water	47.7
other ligands	48.3
r.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.79
Ramachandran statistics	
outliers (%)	0
allowed (%)	1.5
favored (%)	98.5

754

755

756 **Figures**



757

758 **Figure 1.** Schematic overview of PrgB domain organization and function. A) Updated
759 schematic overview of the domain organization of PrgB. SS: Signal sequence, COI: Coiled-
760 coil, PAD: Polymer adhesin domain, CSA: adhesin isopeptide-forming adherence domain,
761 CSC: cell-surface antigen C-terminal domain, LPXTG: cell wall anchor sequence. B) In a
762 setting with a donor cell (green) and multiple recipient cells (brown), PrgB is produced and sits
763 on the cell wall. There it enhances cellular aggregation and/or biofilm formation, either by
764 directly binding LTA from the cell-wall of a recipient or first to eDNA to. PrgB compacts the
765 eDNA, and thereby likely pulls the recipient cells closer. Once close enough, PrgB binds to the
766 LTA of the recipient bacteria and facilitates mating-pair formation and conjugation.

767

768

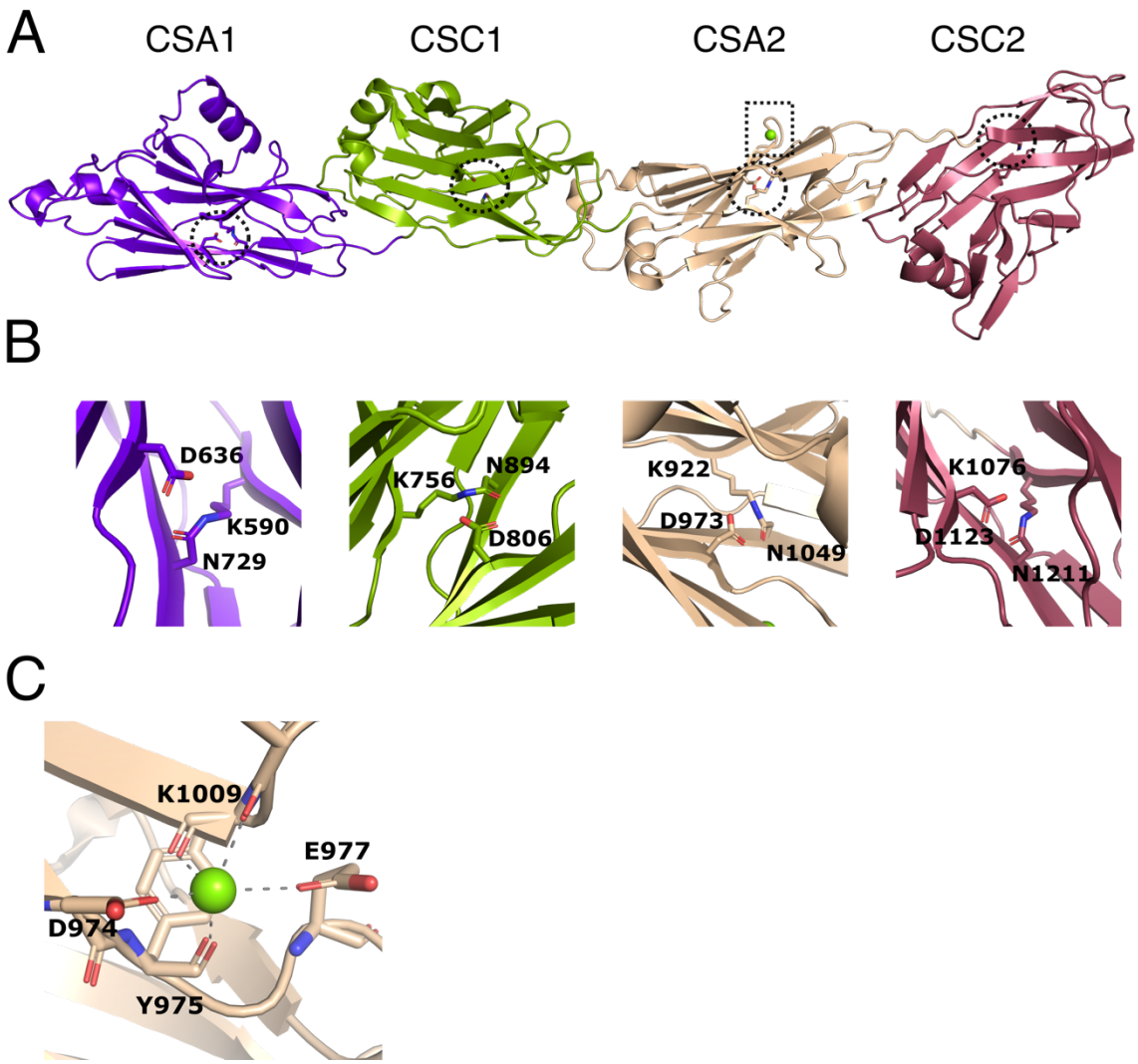


Figure 2. Structure of Ig-like domains of PrgB. A) The Ig-like domains are arranged as tandem pairs (CSA1-CSC1 and CSA2-CSC2). B) Each Ig-like domain has an internal isopeptide bond (indicated by the striped circle in panel A) strengthening the structural integrity of the domain. Each of the four isopeptide bonds is between a lysine and an asparagine and further stabilized by an aspartic acid residue (highlighted residues are shown in stick representation). C) A conserved metal binding site is situated in the CSA2 domain (highlighted by a striped box in panel A), here modelled with a Mg^{2+} (green sphere).

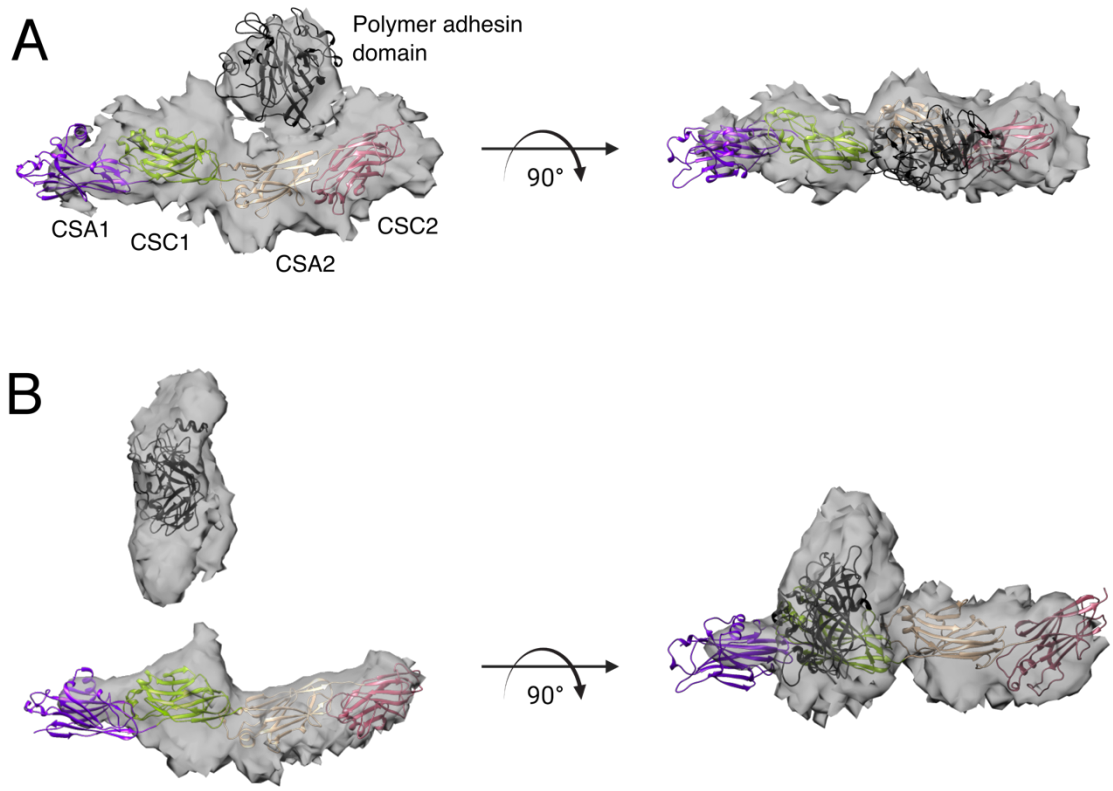


Figure 3. Crystal structures of the PrgB polymer adhesin domain and Ig-like domains in cartoon representation docked into the volumes acquired by single particle cryo-EM. A) The $PrgB_{188-1233}$ model shows the polymer adhesin domain folded back onto the Ig-like domains. B) In the DNA-bound $PrgB_{188-1233}$ model the polymer adhesin domain has undergone a large conformation change away from the Ig-like domain.

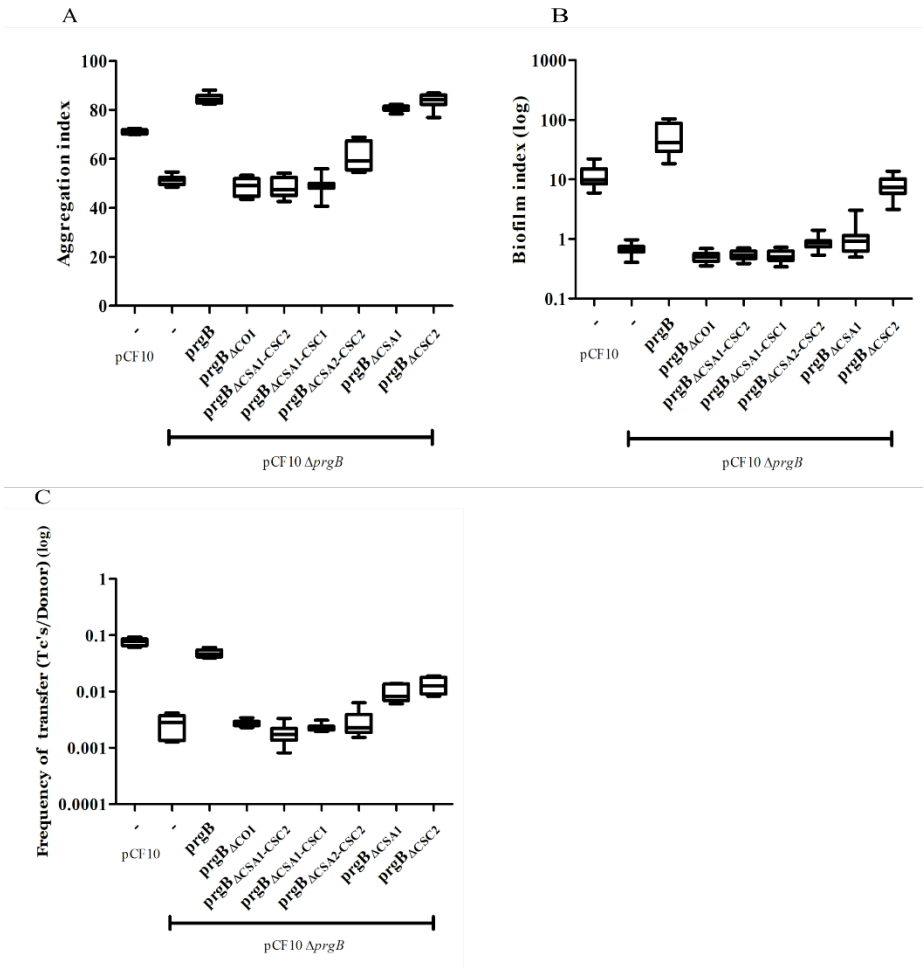


Figure 4. *In vivo* phenotypes of PrgB variants. PrgB variants are expressed from the pMSP3545S-MCS vector in the OG1RF pCF10Δ*prgB* background for phenotypic analysis with three assays: A) cellular aggregation B) biofilm formation and C) conjugation assays. For all assays OG1RF pCF10 carrying an empty vector serves as positive control and OG1RF pCF10Δ*prgB* with an empty vector as negative control. The value of each column represents the average of three independent experiments and the error bars represent the standard error of the mean (SEM).

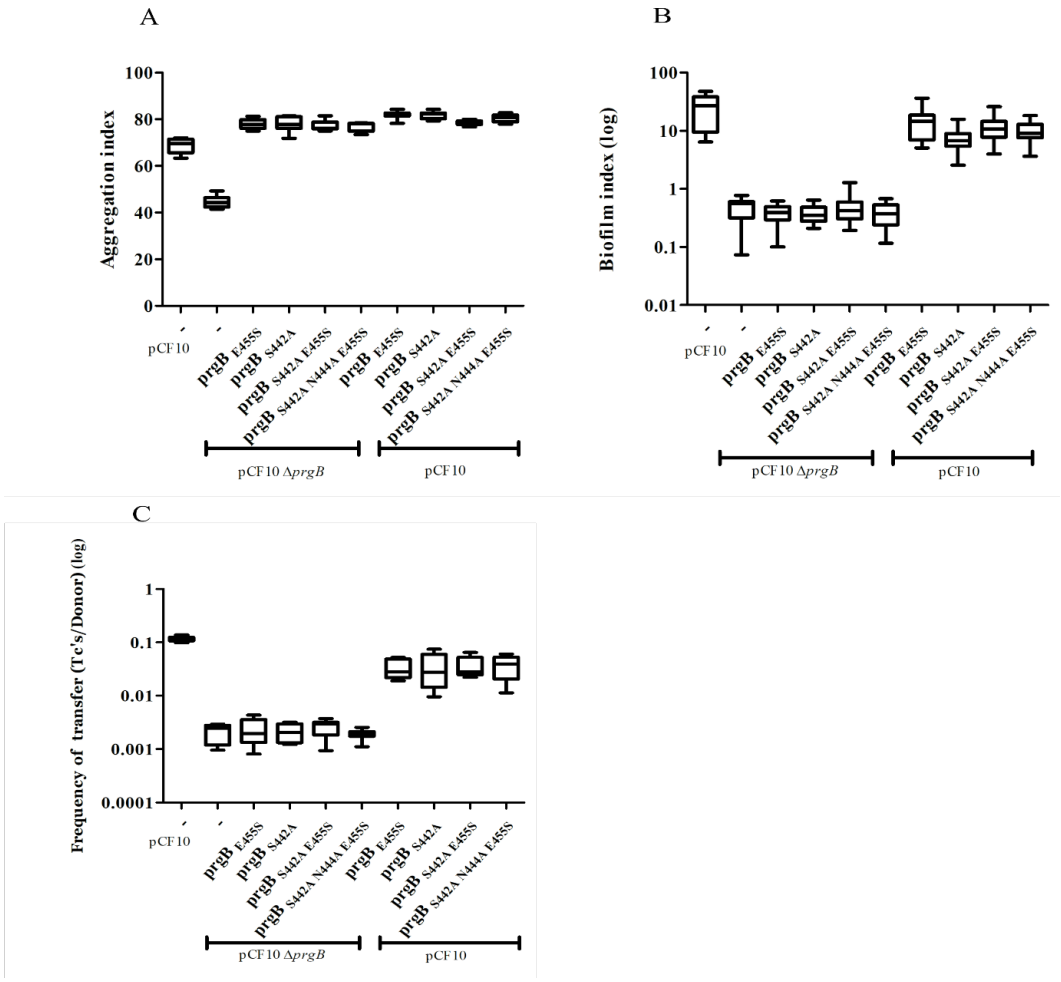


Figure 5. *In vivo* phenotypes of PrgB variants with point mutation(s) in the conserved site of the polymer adhesin domain. PrgB variants are expressed from the pMSP3545S-MCS vector in the OG1RF pCF10Δ*prgB* or OG1RF pCF10 strain and analyzed in: A) cellular aggregation, B) biofilm formation, and C) conjugation assays. OG1RF pCF10 carrying the empty vector serves as positive control and OG1RF pCF10Δ*prgB* with the empty vector as negative control. The height of each column represents the average of three independent experiments and the error bars indicate the standard error of the mean (SEM).

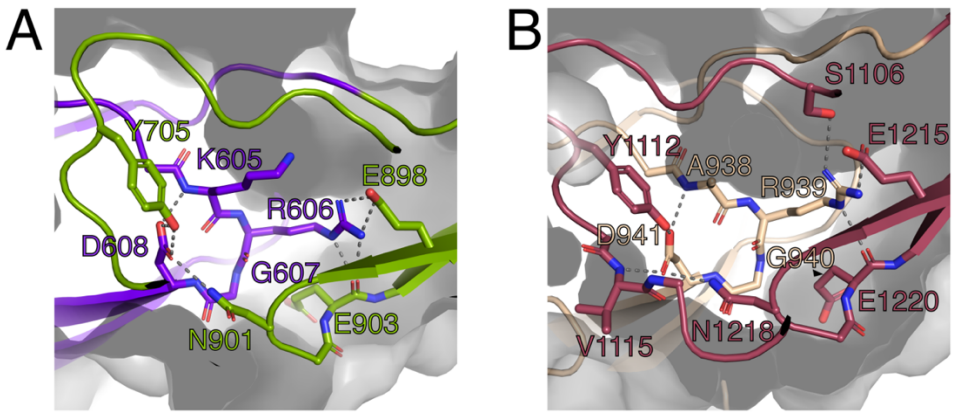


Figure 6. The RGD motifs play a role in the structural integrity of PrgB. A) Close up view of the RGD motif in the CSA1 domain (purple). Most of the important interactions, mainly hydrogen bonds, formed by this motif is to residues on the CSC1 domain (green). B) Close up view of the RGD motif in the CSA2 domain (sand colored). Most of the important interactions, mainly hydrogen bonds, formed by this motif is to residues on the CSC2 domain (dark red). Potential hydrogen bonds are marked by lines and the surface of the protein indicated by transparent gray.