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. After characterizing the polymer adhesin domain of PrgB previously, we here report the structure for almost the entire remainder of PrgB, which reveals that PrgB contains four immunoglobulin-like domains. Based on this new insight, we re-evaluate previously studied variants and present new *in vivo* data where specific domains or conserved residues have been removed. 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 presents the polymer adhesin domain at the right distance from the cell wall.

35 **Introduction**

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

45

46 One of the few well-characterized Gram-positive conjugative plasmids is pCF10 from *E.*
47 *faecalis*[6,9,10]. This conjugative plasmid contains a ~27 kbp operon that is tightly regulated by
48 the P_Q promoter[11–14] and that encodes all proteins needed for conjugation. This operon also
49 encodes three cell-wall anchored proteins: PrgA, PrgB, and PrgC. PrgA is a conjugation
50 regulator that provides surface exclusion to prevent unwanted conjugation. We have previously
51 shown that PrgA consists of a protease domain that is presented far away from the cell wall via
52 a long stalk and that it's likely mediating the proteolytic cleavage of PrgB[15,16]. PrgC is a
53 virulence factor, but its function and structure remain unknown[17]. PrgB is the main adhesin
54 produced by pCF10 and has been studied for well over 3 decades. This protein is around 140
55 kDa in size and possesses an N-terminal signal sequence and a C-terminal LPXTG cell wall
56 anchor motif[18]. PrgB, which is indicated to function as a dimer *in vivo*[19], distributes over the
57 entire surface of the cell-wall and increases cellular aggregation, biofilm formation and the
58 efficiency of plasmid transfer[20,21]. Several mammalian infection model systems have shown
59 that PrgB is a strong virulence factor[18,22–27]. One reason for this virulence is that PrgB mediates

60 biofilm formation in an extracellular DNA (eDNA) dependent manner[17]. Homologs of PrgB
61 have been identified in many other conjugative plasmids[16], suggesting that PrgB-like proteins
62 confer important roles in a large number of bacterial species[28–30].
63
64 PrgB was initially identified as one of the driving forces in cellular aggregation[20]. To
65 understand how it mediated this process, previous research has tried to identify the various
66 protein domains that are present in PrgB and evaluate their function(s). Two RGD (Arg-Gly-
67 Asp) motifs were identified (see figure 1A) and found to be important for vegetation and
68 biofilm formation in the host tissue environment[18,24]. The N-terminal half of PrgB was found
69 to be required for aggregation and to bind lipoteichoic acid (LTA)[30–32], which is a major
70 constituent of the cell-wall in Gram-positive bacteria. In 2018, we solved the structure of
71 $PrgB_{246-558}$ and showed that it has a lectin-like fold that was most similar to adhesins from
72 various oral Streptococci. As these adhesins are known to bind various polymers, we
73 subsequently referred to $PrgB_{246-558}$ as the polymer adhesin domain[16]. We have shown that this
74 domain can bind both LTA and eDNA in a competitive manner. Bound eDNA is thereby
75 strongly compacted as it is wrapped around the domain's positively charged surface[19]. We
76 therefore proposed that PrgB could use eDNA to promote cell-to-cell contacts, as an alternative
77 to direct binding to the LTA from a recipient cell (Fig. 1B). As all described polymer adhesin
78 domains, PrgB has a central ridge with a conserved cation binding site[16,33]. In the homologous
79 GbpC, from *Streptococcus mutans*, this site has been suggested to bind glucans[34]. However, no
80 interaction with glucans has been observed for PrgB or any other homologs[35]. Thus, the
81 importance of this conserved motif remains an open question.
82
83 The polymer adhesin domain plays an important role in the function of PrgB, but it only
84 accounts for around a quarter of the entire protein. Here we present the structure of almost the

85 entire remainder of PrgB. This allows us to put the large amount of available phenotypic data
86 into a structural context and explain a lot of previous observations. We also constructed several
87 new mutants of *prgB* that better fitted the found domain organization and analyzed their *in vivo*
88 effects on cellular aggregation, biofilm formation and conjugation efficiency. Based on our
89 findings, we conclude with an updated model of how PrgB mediates its different functions.
90

91 **Results**

92 *$PrgB_{584-1233}$ contains 4 immunoglobulin-like domains*

93 Previous bioinformatics and structural analysis of PrgB proposed that PrgB consisted of 3
94 domains; the previously crystallized polymer adhesin domain responsible for eDNA and LTA
95 binding, and two domains with RGD (Arg-Gly-Asp) motifs[19]. However, when we reanalyzed
96 the PrgB sequence with the new structure-prediction tools that are available, such as
97 AlphaFold[36], it became clear that this model was partially incorrect. PrgB seems to consist of
98 an N-terminal disordered region (residues 35-197), followed by a newly identified coiled-coil
99 (COI) domain (residues 198-257), the previously crystallized polymer adhesin domain
100 (residues 261-558), a linker region (residues 559-582), 4 immunoglobulin (Ig)-like domains
101 (residues 583-1232) and finally the C-terminal disordered region containing the LPXTG motif
102 (residues 1263-1305) that gets anchored to the cell wall (Fig. 1A). The Ig-like domains seem
103 to come in pairs of two slightly different structures, denoted as CSA1-CSC1 (first pair) and
104 CSA2-CSC2 (second pair), as named in InterPro (CSA from IPR026345; adhesin isopeptide-
105 forming adherence domain, and CSC from IPR032300; cell-surface antigen C-terminal). To
106 verify this updated domain organization of PrgB, we set out to experimentally determine its
107 structure using a combination of X-ray crystallography and cryo-EM methods.
108

109 As described in Schmitt *et al*[19], we were not able to produce full-length PrgB in *E. coli*, but
110 instead produced and purified $PrgB_{188-1233}$. This version of the protein only lacks the disordered
111 N-terminal region and the LPXTG anchor and elutes from size exclusion chromatography in
112 two peaks corresponding to dimeric and a monomeric PrgB. PrgB is in a monomer-dimer
113 equilibrium and the dimer has been described as the main biologically functional unit *in vivo*[7].
114 The monomeric fraction (Fig.1-figure supplement 1) was successfully used for crystallization
115 trials. Crystals belonging to space group $P2_12_12_1$ appeared after 8-12 weeks, diffracted to 1.85
116 Å and contained 2 molecules in the asymmetric unit. The crystallographic phase problem was
117 solved using molecular replacement with SspB (PDB: 2WOY) as a search model. Surprisingly,
118 the resulting electron density lacked the previously crystallized polymer adhesin domain of
119 PrgB, and instead only contained residues 584-1233. There is also no space in the crystal
120 packing to allow for a flexible polymer adhesin domain, so it was likely cleaved off in the
121 crystallization drop before the crystals were formed. The modelled protein indeed consists of
122 four Immunoglobulin (Ig)-like domains: two CSA and two CSC domains (Fig. 2A). Previous
123 bioinformatics analysis showed that various homologous adhesin proteins contain different
124 numbers of Ig-like domains[16], but a DALI[37] analysis of $PrgB_{584-1233}$ showed that there is no
125 previously solved structure in the PDB that contains four of these Ig-domains coupled together.
126 There are, however, homologous structures available with either 2 or 3 Ig-domains. The closest
127 structural homologs are the C-terminal parts of Antigen I/II proteins from oral Streptococci,
128 e.g. the surface protein AspA from *Streptococcus pyogenes*[38] (PDB code: 4OFQ), which has 3
129 Ig-domains and an r.m.s.d. to the Ig-like domains from PrgB of 3.2 Å over 337 residues, or the
130 BspA protein from *Streptococcus agalactiae* [39](PDB code: 4ZLP) which has 2 Ig-domains and
131 an r.m.s.d. of 3.3 Å over 334 residues (see Supplementary file 3 for an overview of the highest
132 ranked DALI hits). The r.m.s.d. decreases to ca. 1 Å if the individual Ig-like domains are
133 superimposed upon each other.

134

135 In each of the Ig-like domains of PrgB, isopeptide bonds are formed between a lysine and an
136 asparagine, a bond which is further stabilized by an aspartic acid (Fig. 2B). This feature is also
137 present in various homologous Ig-like domains from Antigen I/II proteins[38,40,41]. There is
138 density in the conserved metal binding site of the CSA2 domain that has been suggested to
139 bind Ca^{2+} in the Antigen I/II proteins. Refinement of our structure indicated that Mg^{2+} was the
140 best fit to the density (Fig. 2C). The homologous C2 domains from AspA[38], Pas[42], SpaP[40] and
141 SspB[41], each have an extra structural feature termed the BAR (SspB adherence region) domain,
142 which mediates adherence in these proteins. This BAR domain is absent in PrgB (Fig. 2 –
143 figure supplement 1).

144

145 As we didn't manage to crystallize $PrgB_{188-1233}$ without the loss of the polymer adhesin domain,
146 we tried to determine its structure via cryo-EM and single particle analysis. Two datasets of
147 PrgB, one with and one without ssDNA (120 bases), were collected. This yielded relatively
148 low-resolution volumes (8 and 11 Å, respectively). See Fig. 2 – figure supplement 2 for an
149 overview of the processing. Docking in the X-ray structures of the stalk domain ($PrgB_{584-1233}$)
150 and the previously solved polymer adhesin domain (PDB code: 6EVU)[19] into the volumes
151 weakly indicated that the polymer adhesin domain could be interacting with the stalk domain
152 in the absence of DNA (Fig. 2 – figure supplement 3). Therefore, we set out to test whether
153 the polymer adhesin domain binds to the Ig-like domains from the stalk domain ($PrgB_{584-1233}$)
154 *in vitro*. However, neither size exclusion chromatography nor native PAGE indicated any
155 binding of the polymer adhesin domain to the stalk domain *in vitro* (Fig. 2 – figure supplement
156 4).

157

158 *In vivo assays*

159 Based on the new structural insights for PrgB, we decided to study the importance of the newly
160 defined coiled-coil domain and the Ig-like domains. This was done by complementing *E.*
161 *faecalis* OG1RF pCF10Δ*prgB* with different *prgB* mutants (from a nisin-inducible plasmid)
162 and characterizing their phenotypes in cellular aggregation, biofilm formation, and plasmid
163 transfer efficiency. In line with previous experiments[17], complementing pCF10Δ*prgB* with
164 exogenous PrgB from the pMSP3545S vector rescues all phenotypes. Aggregation and biofilm
165 formation, both measured after overnight incubation, are even slightly increased as compared
166 to wild-type pCF10 (Fig. 3A-B, column 1 and 3), possibly due to a slightly increased
167 production of PrgB (Fig. 3 – figure supplement 2B, lane 1 and 3).
168
169 We found that expression of PrgB without the newly identified coiled-coil domain could not
170 rescue the aggregation phenotype of the deletion strain (Fig. 3A, column 4). Deletion of either
171 the CSA1 or the CSC2 domain did not affect PrgB-mediated aggregation (Fig. 3A, column 8
172 and 9). However, complementation with PrgBΔCSA2-CSC2 could only partially rescue the
173 aggregation phenotype of the *E. faecalis* OG1RF pCF10Δ*prgB* strain (Fig. 3A, column 7) and
174 no rescue at all was seen in the PrgB variants with both the CSA1 and CSC1 domain deleted
175 or without any Ig-like domains (CSA1-CSC1-CSA2-CSC2) (Fig. 3A, column 5, and 6).
176
177 As expected[17], deletion of *prgB* also leads to a large decrease in biofilm formation (Fig. 3B).
178 In line with our observations from the aggregation assays, PrgB with a deletion of either the
179 coiled-coil domain or more than a single Ig-like domain could not rescue the *E. faecalis*
180 OG1RF pCF10Δ*prgB* biofilm formation phenotype. Only exogenous expression of $PrgB_{\Delta CSC2}$
181 can rescue the level of biofilm formation, but only to the level found in OG1RF pCF10, not to
182 the level of exogenously expressed wild-type PrgB (Fig 3B, column 1, 3 and 9). Intriguingly,

183 the expression of exogenous $PrgB_{\Delta CSA1}$ in the OG1RF pCF10Δ*prgB* background did not restore
184 biofilm formation, while it did restore the aggregation phenotype (Fig. 3A and 3B, column 8).
185
186 The PrgB variants that failed to rescue the aggregation phenotype of the OG1RF pCF10Δ*prgB*
187 strain ($PrgB_{\Delta COI}$, $PrgB_{\Delta CSA1\text{-}CSC2}$, $PrgB_{\Delta CSA1\text{-}CSC1}$ and $PrgB_{\Delta CSA2\text{-}CSC2}$) were also found to have
188 a decreased plasmid transfer efficiency in the conjugation assay (Fig. 3C, column 4-7).
189 However, $PrgB_{\Delta CSA1}$ and $PrgB_{\Delta CSC2}$ could only partially rescue the conjugation rate of the
190 OG1RF pCF10Δ*prgB* strain (Fig. 3C, column 8-9), while they did rescue the aggregation
191 phenotype.
192
193 To determine whether all PrgB variants were properly expressed, translocated and linked to the
194 cell wall, we probed the protein levels in the cell wall fraction by Western Bot after 1 hr
195 induction (corresponding to a time point relevant for the conjugation assay) and after overnight
196 incubation (corresponding to a time point relevant to the aggregation and biofilm assays). All
197 PrgB variants were present in the cell wall extract after 1 hr induction (Fig. 3 – figure
198 supplement 1A), indicating that the production and folding of them are normal. However, the
199 protein levels of $PrgB_{\Delta COI}$, $PrgB_{\Delta CSA1\text{-}CSC2}$, $PrgB_{\Delta CSA1\text{-}CSC1,}$ and $PrgB_{\Delta CSA2\text{-}CSC2}$ were reduced
200 after overnight incubation, as compared to wild type (Fig. 3 – figure supplement 1B, lane 5-8
201 compared to lane 3). This indicates that the protein stability is decreased when these domains
202 are missing, which could explain the observed aggregation deficiency after overnight
203 incubation. Intriguingly, even though $PrgB_{\Delta CSA1}$ and $PrgB_{\Delta CSC2}$ are relatively stable (Fig. 3 –
204 figure supplement 1B, lane 9-10) and complemented the aggregation phenotype, they still
205 could not fully rescue the biofilm phenotype. This suggests a functional loss of in these two
206 variants.
207

208 *The conserved binding cleft in the polymer adhesin domain is essential for conjugation and*
209 *biofilm formation, but not for aggregation*
210 To investigate the role of the binding cleft in the polymer adhesin domain, we introduced
211 single, double, and triple mutations to alter its conserved residues. The resulting *prgB* mutants
212 were exogenously expressed in the background of pCF10Δ*prgB* for functional
213 complementation; or in the background of wild-type pCF10 to test for any potential dominant
214 negative effects as previously observed for $PrgB_{\Delta 246\text{-}558}$ (deletion of the polymer adhesin
215 domain)[19]. Notably, all the resulting PrgB variants fully restored the aggregation phenotype of
216 pCF10Δ*prgB* (Fig. 4A), but did not rescue the defective biofilm formation, nor the reduced
217 conjugation efficiency (Fig. 4B and C, column 2-6). No dominant negative effects on
218 aggregation or biofilm formation could be observed when these variants were expressed in the
219 wild-type pCF10 background (Fig. 4A and B, column 7-10), although slightly reduced
220 conjugation rates were observed (Fig. 4C, column 7-10). We have previously shown that the
221 polymer adhesin domain binds eDNA and that this binding correlates with both biofilm
222 formation and conjugation efficiency. Therefore, we wanted to test whether eDNA binding was
223 affected in these PrgB variants. To do so, we purified both wild-type $PrgB_{188\text{-}1235}$ and the double
224 binding cleft variant $PrgB_{188\text{-}1233}$:S442A,N444A to compare their DNA binding affinities via
225 electrophoretic mobility shift assays (EMSA). The results indicate that the introduced changes
226 in PrgB did not affect its ability to bind eDNA, as the DNA binding affinities of $PrgB_{188\text{-}1233}$
227 and $PrgB_{188\text{-}1233}$:S442A,N444A were the same within experimental error (Fig. 4 – figure
228 supplement 1).
229
230 **Discussion**
231 The presented data provides important insights for a widespread virulence factor in Gram-
232 positive bacteria, since genes encoding PrgB homologs exist on a large number of conjugative

233 plasmids[16]. In this study, we expand our structural knowledge of PrgB beyond the polymer
234 adhesin domain, to now encompass almost the entire protein.
235
236 The crystal structure of $PrgB_{583-1233}$ shows that this part of PrgB consists of four tandemly
237 arranged immunoglobulin (Ig)-like domains. These Ig-like domains show a high degree of
238 structural homology to Streptococcal surface proteins, usually found in the oral cavity[16].
239 These homologous proteins have been indicated to bind various molecules, such as fimbria,
240 collagen and salivary agglutinin (SAG, also designated as glycoprotein 340), and these binding
241 interactions have been shown to be vital for their function[41–45]. However, we have not found
242 any evidence that the Ig-like domains of PrgB bind to a specific substrate in our own
243 experiments, nor have we found this in other reports. PrgB also does not contain a BAR
244 domain, which is crucial for stable interactions between *e.g.* SspB from *S. gordonni* and Mfa-
245 1 of *P. gingivalis*[46,47]. However, since Ig-like domains are known to bind a large variety of
246 ligands[48], we don't exclude the possibility that ligands for the Ig-like domains in PrgB will be
247 found in the future. The only ligands that have been verified to interact with PrgB so far are
248 eDNA and LTA, which have high affinity to the polymer adhesin domain[19].
249
250 The crystal structure of the Ig-like domains from PrgB, $PrgB_{583-1233}$, was complemented by
251 single particle analysis of $PrgB_{188-1233}$ via cryo-EM (Fig. 2 - figure supplement 3). Despite the
252 low resolution of the EM volumes, we tried to dock in the high-resolution crystal structures of
253 the Ig-like domains and the polymer adhesin domain. The model of PrgB from cryo-EM
254 indicated that the polymer adhesin domain might interact with the Ig-like domains in the
255 absence of substrate, which could have implications for the function or regulation of PrgB. To
256 test this hypothesis, we assayed the interaction between the polymer adhesin domain and the
257 separately purified Ig-like stalk domain *in vitro.* The results did not show any interaction

258 between these domains. However, in the full-length protein these two domains are attached via
259 a linker region, which make their local concentration very high. Thus, we cannot exclude that
260 these two domains can interact, but in that case the dissociation constant must be high (at least
261 high μM range) and the interaction is thus unlikely to be physiologically relevant.
262
263 We have now obtained a structural basis to interpret almost all phenotypic data that is available
264 for PrgB. Unfortunately, most of the mutants that were previously described did not correlate
265 well with the newly determined domain boundaries. Therefore, we decided to create specific
266 deletion mutants that were based on the new PrgB structure to determine the role of the various
267 domains. At the N-terminus of PrgB, before the polymer adhesin domain, there is a predicted
268 coiled-coil region (Fig. 1A) that we wanted to investigate. To our surprise, the expression of
269 $PrgB_{\Delta COI}$ could not rescue any of the aggregation, biofilm formation or conjugation phenotypes
270 from a *prgB* deletion strain. However, in our experiments $PrgB_{\Delta COI}$ was unstable with
271 substantially decreased amounts present in the cell-wall extract collected after overnight
272 induction as compared to 1 hour induction (Fig. 3 – figure supplement 1), indicating that the
273 coiled-coil region is important for protein production and/or stability. In line with this
274 hypothesis, Alphafold predicts this coiled-coil domain to interact with the linker between the
275 polymer adhesin domain and the Ig-like domain (Fig. 2 – figure supplement 5). Since the linker
276 region contains the sequence that PrgA recognizes for cleavage of PrgB[15], the lack of the
277 coiled-coil domain and its potential shielding effect could explain the decreased stability of
278 $PrgB_{\Delta COI}$. Deletion of all Ig-like domains ($PrgB_{\Delta CSA1-\Delta CSC2}$) renders the protein incapable to
279 support aggregation, biofilm formation and conjugation. However, exogenous expression of
280 *prgB* with single Ig-like domain deletions, $prgB_{\Delta CSA1}$ and $prgB_{\Delta CSC2}$ in the pCF10Δ*prgB*
281 background, restore cellular aggregation, while they do not rescue biofilm formation and
282 conjugation (Fig. 3). This was unexpected, as the polymer adhesin domain is predicted to

283 mediate all the functions that were tested in these assays: aggregation, biofilm formation and
284 conjugation[17,19]. Our results, however, indicate that it is important to have all Ig-like domains
285 present and properly folded. The cellular aggregation assays seem to indicate that PrgB can
286 function when at least 3 Ig-like domains are present, as expression of both *prgB*ΔCSA1 and
287 *prgB*ΔCSC2 can complement pCF10Δ*prgB*, but unfortunately this assay is not suitable to
288 detect small changes. Based on the biofilm formation and conjugation efficiency assays, which
289 are more sensitive, we therefore conclude that even removing a single Ig-like domain strongly
290 decreases the function of PrgB. We hypothesize that these Ig-like domains are required to
291 present the polymer adhesin domain at the right distance from the cell.
292
293 The conserved Ser-Asn-Glu site in the negatively charged cleft of the polymer adhesin domain
294 intrigued us, as its function is unknown. Any changes that we made in these conserved residues
295 resulted in PrgB variants that didn't facilitate biofilm formation or conjugation (Fig. 4).
296 Surprisingly the same PrgB variants did fully support cellular aggregation. This is thought-
297 provoking, since various literature has shown that the PrgB functions in cellular aggregation,
298 biofilm formation and conjugation are strongly correlated. However, even a single point
299 mutation in this conserved site produced a PrgB variant that completely separates the cellular
300 clumping phenotype from biofilm formation and conjugation. *In vitro* experiments showed that
301 these point mutations did not impair PrgB binding to eDNA (Figure 4 – figure supplement 1).
302 A similar phenotypic pattern was observed with $PrgB_{\Delta CSA1}$ and $PrgB_{\Delta CSC2}$, which both could
303 fully rescue aggregation but not biofilm formation. Our data therefore strongly indicates that
304 PrgB has additional role(s), besides mediating cellular aggregation, to further support
305 conjugation and biofilm formation. We therefore propose that PrgB performs at least one,
306 currently unknown, additional function besides binding to eDNA and/or LTA from the cell

307 wall of the recipient cell. It is highly likely that at least one of these additional functions is
308 mediated by the conserved site in the polymer adhesin domain.
309
310 As described in the introduction, there is a plethora of *prgB* mutants made (see Fig. 5 – figure
311 supplement 1) and phenotypically analyzed, predominantly by the group of Prof. Gary Dunny.
312 In supplementary file 4 we provide a summary of all *prgB* mutants that we have identified in
313 the literature and our brief reinterpretation based on our new structural knowledge. Below, we
314 will discuss a selected number of these mutants in detail.
315
316 As expected, most mutants that have an insertion or a mutation in the polymer adhesin domain
317 (Fig. 5 – figure supplement 1) show a loss of protein function. Shortly, insertions at amino
318 acids (a.a.) 358 and 359 are on the surface of the protein and likely leads to steric clashes that
319 prevent the domain from binding to eDNA and LTA, whereas insertions at a.a. 439, 473, 517,
320 and 546 are all in secondary structures that are central parts of the polymer adhesin domain and
321 therefore are likely to disrupt folding of this domain.
322
323 The RGD motifs in PrgB were previously proposed to be involved in integrin binding and to
324 promote adherence to human neutrophils, as well as internalization in cultured intestinal
325 epithelial cells[26,49]. The new domain classification showed that these two RGD motifs are in
326 CSA1 and CSA2, respectively (Fig. 1A). The structure that we determined shows that these
327 two sequence motifs are not surface exposed, but instead play an important role in the structural
328 integrity of the interfaces between CSA1 and CSC1, and between CSA2 and CSC2 (Fig. 5).
329 Mutations in these motifs would thus very likely destabilize the folding of the tandem Ig-
330 domains, which would explain the decreased PrgB biofilm formation observed in these

331 strains[18,24]. Thus, our new data strongly argues against the previously proposed direct binding
332 interaction between the RGD sequences and integrins of host origin.
333
334 Deletion of residues 993-1138 leads to a loss of about half of both the CSA2 and CSC2
335 domains. Therefore, we were surprised to find that this mutant was described to behave like
336 wild type in aggregation assays[32]. Possibly the remaining parts of CSA2 and CSC2 form a
337 (unfolded) linker region of similar length to the tandem CSA2-CSC2 structure, which allows
338 PrgB to retain its function of promoting aggregation. Similarly, $PrgB_{\Delta 668\text{-}1138}$, corresponding to
339 a complete removal of CSC1 and CSA2 and approximately half of both CSA1 and CSC2
340 domains, could still support PrgB mediated aggregation. Potentially the remaining residues (ca
341 180 amino acids) could also form an unfolded linker region allowing for the variant to retain
342 its aggregation phenotype. For $PrgB_{\Delta 993\text{-}1138}$ and $PrgB_{\Delta 668\text{-}1138}$, unfortunately no biofilm
343 formation or conjugation efficiency were reported, but based on our results it is likely that those
344 capabilities would have been severely compromised.
345
346 Taking past and present results into account combined with our new structural insights, we
347 propose a new mechanistic model for the function of PrgB. Our *in vivo* data show that removal
348 of one of the Ig-like domains does not largely affect PrgB function. Previous data also indicates
349 that parts of the Ig-like domains can be deleted without affecting the function of PrgB. We
350 therefore hypothesize that the Ig-like domains provide two important features to the protein.
351 First, a rigid stalk that is needed to present the polymer adhesin domain at the correct distance
352 from the cell wall. Second, providing the correct structural positioning for PrgA-mediated
353 cleavage of the polymer adhesin domain, as the potential protease domain in PrgA is also
354 presented away from the cell on a ~40 nm long stalk [15,16]. The exact distance of the polymer
355 adhesin domain from the cell wall does not seem important for LTA-binding, since aggregation

356 can still take place even with partially disrupted Ig-like domains. However, the length of the
357 stalk may be very important when it comes to facilitating both biofilm formation and
358 conjugation. This indicates that effective mating-pair formation in conjugation may require a
359 PrgB-mediated function distinct from pure aggregation, something that is further shown by the
360 findings of the mutations in the conserved site in the polymer adhesin domain.
361
362 Besides providing a structural basis to explain about 30 years of work on PrgB, we here also
363 uncovered that the conserved Ser-Asn-Glu site in the polymer adhesin domain likely provides
364 additional functionality to PrgB that is needed for optimal biofilm formation and conjugation,
365 but that does not affect cellular aggregation. To fully investigate the function of this conserved
366 site in PrgB and other homologous virulence factors from Gram-positive bacteria, remains an
367 exciting question to address in future research.
368
369 **Materials and Methods**
370 *Bacterial strains and growth conditions*
371 See Supplementary File 1 for a full list of all strains, plasmids and oligonucleotides used.
372 *Escherichia coli* Top10 was used in molecular cloning and grown in Lysogeny broth (LB). *E.*
373 *coli* BL21 (DE3) was used for recombinant protein expression and grown in Terrific Broth
374 (TB). The *E. faecalis* strains were cultured in Brain-Heart infusion broth (BHI) or Tryptic Soy
375 broth without dextrose (TSB-D) as indicated in each assay. Concentrations of antibiotics for *E.*
376 *coli* selection were as follows: ampicillin (100 µg/ml), kanamycin (50 µg/ml), spectinomycin
377 (50 µg/ml), and erythromycin (150 µg/ml). In *E. faecalis* cultures, antibiotics were used as the
378 following concentrations: tetracycline (10 µg/ml), fusidic acid (25 µg/ml), erythromycin (20
379 µg/ml for chromosome-encoded resistance; 100 µg/ml for plasmid-encoded resistance),
380 spectinomycin (250 µg/ml for chromosome-encoded resistance; 1000 µg/ml for plasmid-

381 encoded resistance), streptomycin (1000 µg/ml). Plasmids were transformed to *E. coli* by heat-
382 shock transformation, whereas *E. faecalis* strains were transformed by electroporation[50].
383
384 *Cloning and mutagenesis*
385 To insert a multiple cloning site in pMSP3545S, DNA oligos of MCS_fwd and MCS_rev were
386 resuspended in miliQ to 100 µM, mixed 1:1, denatured at 95°C for 15 min and slowly cooled
387 to room temperature. The annealed MCS oligo was ligated into the gel-purified backbone
388 fragment from pMSP3545S-*prgK* vector[51] digested with NcoI and XbaI (removing the *prgK*
389 insert). This was done in a 30 µL ligation mixture with 90 ng of the digested vector and a 9
390 times molar access of insert. 5 µL of this ligation mixture was transformed into Top10
391 competent cells and plated on LB agar plates with spectinomycin (50 µg/ml), and erythromycin
392 (150 µg/ml). The constructed plasmid was analyzed by restriction digestion and sequenced to
393 confirm the correct insertion of the multiple cloning site and is further called pMSP3545S-
394 MCS.
395
396 pMSP3545S-*prgB* was constructed by PCR of *prgB* from pCF10 with the NcoI-prgB-F and
397 BamHI-stop-prgB-R primer pair and placement into the pMSP3545S-MCS vector via
398 NcoI/BamHI restriction, gel-purification of the correct DNA fragments and subsequent
399 ligation. The constructed pMSP3545S-*prgB* was subsequently used as a template for
400 mutagenesis creating pMSP3545S-*prgB* deletion or point mutation variants. This was carried
401 out with inverse PCR (iPCR) using partially overlapping primer pairs. The iPCR products were
402 gel-purified with the DNA clean-up kit and digested with DpnI to remove residual template
403 plasmid DNA. The processed iPCR products were then transformed to Top10 competent cells.
404 For over-expression and protein purification, pET-$prgB_{246\text{-}558}$ and pET-$prgB_{188\text{-}1233}$ was
405 transformed into *E. coli* BL21(DE3) as previously described[15]. $prgB_{188\text{-}1233}$ and $prgB_{580\text{-}1233}$

406 DNA fragments were amplified with primers mentioned in Supplementry File 1, and cloned
407 into pINIT vector and then sub-cloned into the p7XC3GH vector via the FX cloning system[52].
408 Mutations were introduced to p7XC3GH-*prgB*$_{188\text{-}1233}$ with the same iPCR approach to obtain
409 the derivative plasmid of *prgB*$_{188\text{-}1233:\ S442A,\ N444A}$. All constructed plasmids were screened by
410 PCR and verified by sequencing.
411
412 *Protein purification and crystallization*
413 PrgB was produced as previously described[19]. Briefly, PrgB$_{188\text{-}1233}$ was expressed with an N-
414 terminal hexa-histidine tag from pET-*prgB*$_{188\text{-}1233}$ or a C-terminal deca-histidine and GFP tag
415 from p7XC3GH-*prgB* in *E. coli* BL21(DE3). The cells were grown at 37 °C in TB medium
416 until they reached an OD_{600} of 1.5. Then the temperature was lowered to 18 °C and protein
417 production was induced by adding 0.5 mM IPTG. Cells were grown for 16 hours before
418 harvesting by centrifugation. Cells were resuspended in 20 mM HEPES/NaOH (pH 7.0), 300
419 mM NaCl, 30 mM Imidazole and 0.02 mg/ml DNase I and broken with a Constant cell disruptor
420 at 4 °C and 25 kPsi (Constant Systems). The cell lysate was clarified by centrifugation for 30
421 minutes at 30,000 x g, 4°C and incubated at 4 °C with Ni-NTA (Macherey-Nagel). The Ni-
422 NTA column was washed with 10 column volumes of 20 mM HEPES/NaOH (pH 7.0), 300
423 mM NaCl, 30-50 mM imidazole and bound proteins were eluted from the column with the
424 same buffer supplemented with 500 mM Imidazole. The histidine affinity tags and the GFP
425 when present, were cleaved off from the purified protein fractions by incubation with TEV
426 protease (for pET-*prgB*) or Prescission protease (for p7XC3GH-*prgB*) in a 1:100 ratio for 20
427 h at 4 °C. The cleaved proteins were loaded on a Superdex-200 Increase 10/300 GL column
428 (Cytiva) equilibrated in 20 mM HEPES/NaOH (pH 7.0) and 150 mM NaCl. The elution profile
429 showed two peaks corresponding to a PrgB dimer and monomer, as previously reported[19].
430 These two peak fractions were pooled separately and concentrated on an Amicon Ultra

431 Centrifugal Filter with a 30 kDa cut-off (Merck-Millipore). 10% glycerol was added to the
432 concentrated protein fractions, which were subsequently flash frozen in liquid nitrogen and
433 stored at -80 °C.
434
435 *Native gel electrophoresis*
436 Elution fractions from size-exclusion chromatography were mixed with native gel sample
437 buffer (Invitrogen) and loaded on a Novex 4-20% Tris-Glycine gel (Invitrogen). Staining was
438 carried out with InstantBlue Protein Stain (VWR).
439
440 *Structure determination via X-ray crystallography*
441 Purified $PrgB_{188-1233}$ from the monomeric fraction, with a protein concentration of 15 mg/mL,
442 were thawed and used in crystallization trials. Crystals of $PrgB_{188-1233}$ were grown in 8-12
443 weeks, at 20 °C by sitting drop vapor diffusion in a condition with 0.2 M $CaCl_2$ and 20% PEG
444 3350 and a protein to reservoir ratio of 1:1 in the drop. Crystals were flash cooled in liquid
445 nitrogen. X-ray diffraction data of $PrgB_{188-1233}$ was collected on ID23-1 at the ESRF, France.
446 The data was processed using XDS[53]. The $PrgB_{188-1234}$ crystals belong to the monoclinic space
447 group $P2_12_12_1$ and contain two molecules in the asymmetric unit. The crystallographic phase
448 problem was solved using molecular replacement using PHASER, using the SspB homology
449 model of the Ig-domains as search models (PDB: 2WOY)[41]. Further building of the model was
450 conducted in COOT[54]. The structure was refined to 1.85 Å with crystallographic R_{work} and R_{free}
451 values of 20.9/24.6 using Refmac5 and PHENIX refine[55,56]. The final $PrgB_{188-1233}$ model
452 consists of residues 584-1234, and was validated using MolProbity[57]. Atomic coordinates and
453 structure factors have been deposited in the Protein Data Bank (PDB Code: 8BEG).
454
455 *Sample preparation for electron microscopy*

456 $PrgB_{188-1233}$ dimer fractions were thawed on ice and loaded on a Superdex 200 10/300 GL gel
457 filtration column (GE Healthcare) equilibrated in 20 mM HEPES/NaOH pH 7.0 and 150 mM
458 NaCl. Protein peak fractions, corresponding to the dimer, were diluted to 0.1-0.3 mg/mL and
459 for the DNA-bound structures 120 bp ssDNA (Table S1) was added in 1: 1.2 molar ratio
460 (protein:DNA) and samples were incubated for 15 minutes on ice. For both apo and DNA-
461 bound samples, 4 µl of sample was applied to glow discharged Quantifoil 300 mesh 1.2/1.3
462 (Quantifoil) grids at 4 °C and 90-100% humidity, blotted for 1 s with blot force -5, and plunge-
463 frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific).
464
465 *Cryo-EM data collection*
466 Cryo-EM data were collected on an FEI Titan Krios transmission electron microscope (Thermo
467 Fisher Scientific), operated at 300 keV that was equipped with a K2 direct electron detector.
468 Data was collected by the AFIS method using the EPU software V2.8.0 (Thermo scientific) at
469 a nominal magnification of 165,000x (0.82 Å pixel size). Data collection parameters are listed
470 in Supplementary file 2, and the general workflow showing representative micrographs, 2D
471 and 3D classes are shown in Fig. 2-figure supplement 2. A total number of 2787 movie stacks
472 were collected for apo PrgB and 1670 for DNA-bound PrgB.
473
474 *Cryo-EM data processing*
475 Cryo-EM data of apo PrgB and DNA-bound PrgB were processed in the same way, but
476 separately using cryoSPARC (v3.2.0-3.3.1)[58]. Beam-induced motion was corrected using
477 standard settings, where start frame 1 was excluded, followed by per-micrograph contrast
478 transfer function (CTF) estimation. For apo PrgB a subset of 500 micrographs were picked
479 using the blob picking tool with a 100-200 Å particle diameter, followed by extraction of
480 170,826 particles with a box size of 384 pix. Picked particles were subjected to consecutive

481 rounds of 2D classifications. Subsequently, representative 2D classes were selected as input
482 for picking of the full dataset, using the template picker tool. PrgB with DNA was directly
483 picked using blob picker with a 100-300 Å particle diameter and standard settings and extracted
484 with 384 pix. PrgB without and with DNA were then separately subjected to 2D classifications
485 resulting in a final number of 283,630 and 163,566 particles respectively. Particles from
486 selected classes were combined and used in ab-initio reconstruction. The initial volume was
487 then subjected to homogenous 3D refinement and the resolution was calculated using the gold
488 standard Fourier shell correlation (FSC threshold, 0.143) and found to be 8 Å and 11 Å for the
489 apo and DNA-bound structure, respectively. The volumes of apo and DNA-bound PrgB have
490 been deposited in the Electron Microscopy Data Bank (EMDB Codes: EMD-16001 and EMD-
491 16002).
492
493 The adhesion domain (PDB Code: 6EVU)[19] and stalk domain (PDB Code: 8BEG) were
494 initially docked into the EM volume using Chimera[59] and subsequently run through
495 Namdinator[60] using 10 Å resolution and standard settings. The output was then fitted in the EM
496 volume in Chimera (v 1.15rc[59]), where figures also were generated.
497
498 *Aggregation assay*
499 *E. faecalis* strains were inoculated in BHI medium with the indicated antibiotics and cultured
500 overnight at 37°C. Overnight cultures were diluted in a 1:100 ratio in TSB-D with the indicated
501 antibiotics, 10 ng/mL cCF10, and 50 ng/ mL nisin and dispensed into polystyrene cuvettes
502 (Sarstedt) in 0.9 mL triplicates. These were incubated for 24 h at 37°C without agitation.
503 Afterwards, the optical density of each sample was determined at 600 nm both before (OD_{sup})
504 and after (OD_{mix}) vigorously mixing of the bacterial culture by pipetting. The autoaggregation
505 percentage was then calculated as follows: $100 \times [1 - (OD_{sup}/OD_{mix})]$[17,30].

506

507 *Biofilm assay*

508 *E. faecalis* strains were inoculated in BHI with the indicated antibiotics and kept 16 h at 37°C.
509 The next morning, they were diluted in a 1:100 ratio in TSB-D with the indicated antibiotics,
510 10 ng/mL cCF10, and 50 ng/mL nisin. 200 µl fractions were dispensed into a 96-well micro-
511 titer plate (Costar) with 8 replicates per strain. 200 µL TSB-D fractions were used as blanks.
512 The 96-well plate was then incubated aerobically at 37°C without agitation in a humidified
513 chamber for 24 h. The suspension was transferred to another 96-well plate to determine the
514 optical density at 600 nm (OD_{600}). The plate containing the biofilm was washed with distilled
515 water three times and then left to air dry at room temp for 2.5 h. The biofilm was stained with
516 100 µl 0.1% (w/v) safranin (Sigma) at room temp for 20 min, then washed three times with
517 distilled water and left to air dry at room temperature. Afterwards the absorbance was
518 determined using a plate reader (BMG Labtech) at 450 nm. Biofilm production was calculated
519 as an index of safranin staining of the cell biomass divided by absorbance of its optical density
520 (OD_{450}/OD_{600})[61].

521

522 *Analysis of the protein levels of PrgB variants in cell wall extracts*

523 For the 1-hour time-point, samples of each strain were diluted in a 1:25 ratio in TSB-D with
524 the indicated antibiotics, cultured at 37°C for 2h, until optical density 0.6 was reached, and then
525 induced with 10 ng/mL cCF10, and 50 ng/ mL nisin for 1 hour. Overnight samples were
526 harvested after induction and incubation overnight as described in the aggregation assay. The
527 cell pellets were treated with lysozyme buffer (10 mM Tris, pH 8.0, 1 mM EDTA, 25% sucrose,
528 15 mg/ml lysozyme) for 30 min at 37°C. The lysozyme-treated bacterial cells were spun down
529 at 13,000 *xg*, 4 °C for 5 minutes. The supernatant, containing the cell wall extract, was mixed
530 with protein loading dye and boiled at 100°C for 12 min. Subsequently, the samples were run

531 on a 8% SDS-PAGE, transferred to Western blot and probed with the PrgB antibody produced
532 in rabbit[62–64].
533
534 *Conjugation assay*
535 Donor (OG1RF pCF10 pMSP3545S derivative strains) and recipient (OG1ES) strains were
536 inoculated in BHI with the indicated antibiotics and incubated overnight at 37 °C with agitation.
537 Overnight cultured strains were refreshed in BHI without antibiotics in a 1:10 ratio, and donor
538 strains were induced with 50 ng/mL nisin (Sigma). All strains were then incubated at 37 °C for
539 1h without agitation. Afterwards each of the donor strains was mixed with the recipient cells
540 in ratio of 1:10 and mated at 37 °C statically for 3.5 h. These mixtures were then serially diluted
541 with BHI and plated out in triplicates on BHI agar plates supplemented with tetracycline and
542 spectinomycin (to select for donor cells), or with tetracycline, erythromycin, and streptomycin
543 (to select for transconjugants). Plates were incubated at 37 °C for 48 h, counted and enumerated
544 for colony forming units (CFU). The plasmid transfer rate was determined as CFU of
545 transconjugant over CFU of donor (Tc's/Donors)[19].
546
547 Electrophoretic Mobility Shift Assay (EMSA)
548 EMSA was carried out in the same way as previously described[19]. 0.1 to 20 µM PrgB (wild
549 type and variants) were mixed with 50 nM 100 bp long double stranded DNA[19]. Samples were
550 incubated for 1 hour at 20 °C before loading them onto a 6% TBE-based native acrylamide gel
551 for electrophoresis for 90 min at 50 V and 6 °C. Gels were subsequently stained with 3× GelRed
552 (Biotium) in distilled water for 30 minutes and imaged with a Chemidoc system (BioRad).
553 Quantification of the DNA bands after imaging was done in ImageLab (BioRad).
554
555 *Statistical analysis*

556 All *in vivo* data is from three independent experiments and was plotted and analyzed using
557 GraphPad Prism (version 5.0) (GraphPad Software). The indicated error is the standard
558 deviation over 3 biologically independent replicates. Statistical significance between the PrgB
559 variants were analyzed with One-way Anova, with * indicating $p < 0.05$, * * indicating $p <$
560 0.01, and * * * indicating $p < 0.001$.
561
562 *Material availability*
563 All data generated or analyzed during this study are included in this published article and its
564 supplemental information. All structural data has been deposited in the Protein Data Bank
565 (PDB) and the Electron Microscopy Data Base (EMDB) and is publicly available. DOIs are
566 listed in the key resources table. Any additional information required to reanalyze the data
567 reported in this paper, *e.g.* bacterial strains, is available from the corresponding author upon
568 request.
569
570 **Acknowledgments**
571 The authors would like to thank Prof. Gary Dunny for very fruitful discussions about the results
572 and Dr. Karim Rafie and Annika Breidenstein for discussions about EM data processing. We
573 acknowledge MAX IV Laboratory for time on Beamline BioMax under Proposal 20180236.
574 Research conducted at MAX IV, a Swedish national user facility, is supported by the Swedish
575 Research council under contract 2018-07152, the Swedish Governmental Agency for
576 Innovation Systems under contract 2018-04969, and Formas under contract 2019-02496. We
577 also acknowledge the synchrotrons Swiss Light Source (Paul Scherrer Institute, Switzerland)
578 for time at beamline PX1 and the ESRF (France) for time at beamlines ID23 and ID30. The
579 EM data was collected at the Umeå Core Facility for Electron Microscopy, a node of the Cryo-
580 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.

596 **References**

597 1. Gilmore, M. S., Lebreton, F. & van Schaik, W. Genomic transition of enterococci from
598 gut commensals to leading causes of multidrug-resistant hospital infection in the
599 antibiotic era. *Curr. Opin. Microbiol.* **16**, 10–16 (2013).
600 2. Hidron, A. I., Edwards, J. R., Patel, J., Horan, T. C., Sievert, D. M., Pollock, D. A., *et*
601 *al.* NHSN annual update: antimicrobial-resistant pathogens associated with healthcare-
602 associated infections: annual summary of data reported to the National Healthcare
603 Safety Network at the Centers for Disease Control and Prevention, 2006-2007. *Infect.*
604 *Control Hosp. Epidemiol.* **29**, 996–1011 (2008).
605 3. Palmer, K. L., Kos, V. N. & Gilmore, M. S. Horizontal gene transfer and the genomics
606 of enterococcal antibiotic resistance. *Current Opinion in Microbiology* vol. 13 (2010).
607 4. Palmer, K. L. & Gilmore, M. S. Multidrug-resistant enterococci lack CRISPR-cas.
608 *MBio* **1**, (2010).
609 5. Trokter, M. & Waksman, G. Translocation through the conjugative Type 4 secretion
610 system requires unfolding of its protein substrate. *J. Appl. Microbiol.* **200**, e00615-17
611 (2018).
612 6. Cabezón, E., Ripoll-Rozada, J., Peña, A., de la Cruz, F. & Arechaga, I. Towards an
613 integrated model of bacterial conjugation. *FEMS Microbiol. Rev.* (2014).
614 7. Koraimann, G. Spread and Persistence of Virulence and Antibiotic Resistance Genes:
615 A Ride on the F Plasmid Conjugation Module. *EcoSal Plus* **8**, (2018).
616 8. Grohmann, E., Christie, P. J., Waksman, G. & Backert, S. Type IV secretion in Gram-
617 negative and Gram-positive bacteria. *Mol. Microbiol.* **107**, 455–471 (2018).
618 9. Hirt, H., Manias, D. A., Bryan, E. M., Klein, J. R., Marklund, J. K., Staddon, J. H., *et*
619 *al.* Characterization of the pheromone response of the Enterococcus faecalis
620 conjugative plasmid pCF10: complete sequence and comparative analysis of the

621 transcriptional and phenotypic responses of pCF10-containing cells to pheromone
622 induction. *J. Appl. Microbiol.* **187**, 1044–1054 (2005).
623 10. Dunny, G. M. & Berntsson, R. P.-A. Enterococcal sex pheromones: Evolutionary
624 pathways to complex, two-signal systems. *J. Bacteriol.* **198**, 1556–1562 (2016).
625 11. Chandler, J. R., Hirt, H. & Dunny, G. M. A paracrine peptide sex pheromone also acts
626 as an autocrine signal to induce plasmid transfer and virulence factor expression in
627 vivo. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15617–15622 (2005).
628 12. Dunny, G. M. & Berntsson, R. P. A. Enterococcal sex pheromones: Evolutionary
629 pathways to complex, two-signal systems. *J. Bacteriol.* **198**, 1556–1562 (2016).
630 13. Dunny, G. M. Enterococcal Sex Pheromones: Signaling, Social Behavior, and
631 Evolution. *Annu. Rev. Genet.* **47**, 457–482 (2013).
632 14. Lassinantti, L., Camacho, M. I., Erickson, R. J. B., Willett, J. L. E., De Lay, N. R., ter
633 Beek, J., *et al.* Enterococcal PrgU Provides Additional Regulation of Pheromone-
634 Inducible Conjugative Plasmids. *mSphere* **6**, e00264-21 (2021).
635 15. Schmitt, A., Hirt, H., Järvå, M. A., Sun, W. S., ter Beek, J., Dunny, G. M., *et al.*
636 Enterococcal PrgA Extends Far Outside the Cell and Provides Surface Exclusion to
637 Protect against Unwanted Conjugation. *J. Mol. Biol.* **432**, 5681–5695 (2020).
638 16. Järvå, M. A., Hirt, H., Dunny, G. M. & Berntsson, R. P. Polymer Adhesin Domains in
639 Gram-Positive Cell Surface Proteins. *Front. Microbiol.* **11**, (2020).
640 17. Bhatty, M., Cruz, M. R., Frank, K. L., Gomez, J. A. L., Andrade, F., Garsin, D. A., *et*
641 *al.* Enterococcus faecalis pCF10-encoded surface proteins PrgA, PrgB (aggregation
642 substance) and PrgC contribute to plasmid transfer, biofilm formation and virulence.
643 *Mol. Microbiol.* **95**, 660–677 (2015).
644 18. Chuang, O. N., Schlievert, P. M., Wells, C. L., Manias, D. A., Tripp, T. J. & Dunny,
645 G. M. Multiple functional domains of Enterococcus faecalis aggregation substance

646 Asc10 contribute to endocarditis virulence. *Infect. Immun.* **77**, 539–548 (2009).
647 19. Schmitt, A., Jiang, K., Camacho, M. I., Jonna, V. R., Hofer, A., Westerlund, F., *et al.*
648 PrgB promotes aggregation, biofilm formation, and conjugation through DNA binding
649 and compaction. *Mol. Microbiol.* **109**, 291–305 (2018).
650 20. Olmsted, S. B., Kao, S. M., van Putte, L. J., Gallo, J. C. & Dunny, G. M. Role of the
651 pheromone-inducible surface protein Asc10 in mating aggregate formation and
652 conjugal transfer of the Enterococcus faecalis plasmid pCF10. *J. Appl. Microbiol.* **173**,
653 7665–7672 (1991).
654 21. Olmsted, S. B., Erlandsen, S. L., Dunny, G. M. & Wells, C. L. High-resolution
655 visualization by field emission scanning electron microscopy of Enterococcus faecalis
656 surface proteins encoded by the pheromone-inducible conjugative plasmid pCF10. *J.*
657 *Appl. Microbiol.* **175**, 6229–6237 (1993).
658 22. Süssmuth, S. D., Muscholl-Silberhorn, A., Wirth, R., Susa, M., Marre, R. &
659 Rozdzinski, E. Aggregation substance promotes adherence, phagocytosis, and
660 intracellular survival of Enterococcus faecalis within human macrophages and
661 suppresses respiratory burst. *Infect. Immun.* **68**, 4900–4906 (2000).
662 23. Schlievert, P. M., Chuang-Smith, O. N., Peterson, M. L., Cook, L. C. C. & Dunny, G.
663 M. Enterococcus faecalis Endocarditis Severity in Rabbits Is Reduced by IgG Fabs
664 Interfering with Aggregation Substance. *PLoS One* **5**, (2010).
665 24. Chuang-Smith, O. N., Wells, C. L., Henry-Stanley, M. J. & Dunny, G. M.
666 Acceleration of Enterococcus faecalis biofilm formation by aggregation substance
667 expression in an ex vivo model of cardiac valve colonization. *PLoS One* **5**, e15798
668 (2010).
669 25. Rakita, R. M., Vanek, N. N., Jacques-Palaz, K., Mee, M., Mariscalco, M. M., Dunny,
670 G. M., *et al.* Enterococcus faecalis bearing aggregation substance is resistant to killing

671 by human neutrophils despite phagocytosis and neutrophil activation. *Infect. Immun.*
672 **67**, 6067–6075 (1999).
673 26. Vanek, N. N., Simon, S. I., Jacques-Palaz, K., Mariscalco, M. M., Dunny, G. M. &
674 Rakita, R. M. Enterococcus faecalis aggregation substance promotes opsonin-
675 independent binding to human neutrophils via a complement receptor type 3-mediated
676 mechanism. *FEMS Immunol. Med. Microbiol.* **26**, 49–60 (1999).
677 27. Hirt, H., Greenwood-Quaintance, K. E., Barnes, A. M. T., Karau, M. J., Till, L. M.,
678 Palzer, E., *et al.* Dynamics of plasmid-mediated niche invasion, immunity to invasion,
679 and pheromone-inducible conjugation in the murine gastrointestinal tract. *Nat.*
680 *Commun.* **13**, (2022).
681 28. MUSCHOLL, A., GALLI, D., WANNER, G. & WIRTH, R. Sex pheromone plasmid
682 pAD1-encoded aggregation substance of Enterococcus faecalis is positively regulated
683 in trans by traE1. *Eur. J. Biochem.* **214**, (1993).
684 29. Galli, D., Friesenegger, A. & Wirth, R. Transcriptional control of sex-pheromone-
685 inducible genes on plasmid pAD1 of Enterococcus faecalis and sequence analysis of a
686 third structural gene for (pPD1 -encoded) aggregation substance. *Mol. Microbiol.* **6**,
687 1297–1308 (1992).
688 30. Waters, C. M. & Dunny, G. M. Analysis of functional domains of the Enterococcus
689 faecalis pheromone-induced surface protein aggregation substance. *J. Appl. Microbiol.*
690 **183**, 5659–5667 (2001).
691 31. Waters, C. M., Wells, C. L. & Dunny, G. M. The Aggregation Domain of Aggregation
692 Substance, Not the RGD Motifs, Is Critical for Efficient Internalization by HT-29
693 Enterocytes. *Infect. Immun.* **71**, 5682–5689 (2003).
694 32. Waters, C. M., Hirt, H., McCormick, J. K., Schlievert, P. M., Wells, C. L. & Dunny,
695 G. M. An amino-terminal domain of Enterococcus faecalis aggregation substance is

696 required for aggregation, bacterial internalization by epithelial cells and binding to
697 lipoteichoic acid. *Mol. Microbiol.* **52**, 1159–1171 (2004).
698 33. Kohler, V., Keller, W. & Grohmann, E. Enterococcus adhesin PrgB facilitates type IV
699 secretion by condensation of extracellular DNA. *Mol. Microbiol.* **109**, 263–267 (2018).
700 34. Mieher, J. L., Larson, M. R., Schormann, N., Purushotham, S., Wu, R., Rajashankar,
701 K. R., *et al.* Glucan binding protein C of *Streptococcus mutans* mediates both sucrose-
702 independent and sucrose-dependent adherence. *Infect. Immun.* **86**, 1–18 (2018).
703 35. Forsgren, N., Lamont, R. J. & Persson, K. Crystal structure of the variable domain of
704 the Streptococcus gordonii surface protein SspB. *Protein Sci.* **18**, 1896–1905 (2009).
705 36. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., *et al.*
706 Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, (2021).
707 37. Holm, L. DALI and the persistence of protein shape. *Protein Sci.* **29**, 128–140 (2020).
708 38. Hall, M., Nylander, A., Jenkinson, H. F. & Persson, K. Structure of the C-terminal
709 domain of AspA (antigen I/II-family) protein from Streptococcus pyogenes. *FEBS*
710 *Open Bio* **4**, 283–289 (2014).
711 39. Rego, S., Heal, T. J., Pidwill, G. R., Till, M., Robson, A., Lamont, R. J., *et al.*
712 Structural and functional analysis of cell wall-anchored polypeptide adhesin BspA in
713 Streptococcus agalactiae. *J. Biol. Chem.* **291**, (2016).
714 40. Larson, M. R., Rajashankar, K. R., Crowley, P. J., Kelly, C., Mitchell, T. J., Brady, L.
715 J., *et al.* Crystal Structure of the C-terminal Region of Streptococcus mutans Antigen
716 I/II and Characterization of Salivary Agglutinin Adherence Domains. *J. Biol. Chem.*
717 **286**, 21657–21666 (2011).
718 41. Forsgren, N., Lamont, R. J. & Persson, K. Two intramolecular isopeptide bonds are
719 identified in the crystal structure of the Streptococcus gordonii SspB C-terminal
720 domain. *J. Mol. Biol.* **397**, 740–751 (2010).

721 42. Mieher, J. L., Schormann, N., Wu, R., Patel, M., Purushotham, S., Wu, H., *et al.*
722 Structure-function characterization of Streptococcus intermedius surface antigen pas.
723 *J. Bacteriol.* **203**, (2021).
724 43. Brady, L. J., Maddocks, S. E., Larson, M. R., Forsgren, N., Persson, K.,
725 Deivanayagam, C. C., *et al.* The changing faces of Streptococcus antigen I/II
726 polypeptide family adhesins. *Mol. Microbiol.* **77**, 276–286 (2010).
727 44. Maddocks, S. E., Wright, C. J., Nobbs, A. H., Brittan, J. L., Franklin, L., Strömberg,
728 N., *et al.* Streptococcus pyogenes antigen I/II-family polypeptide AspA shows
729 differential ligand-binding properties and mediates biofilm formation. *Mol. Microbiol.*
730 **81**, (2011).
731 45. Franklin, L., Nobbs, A. H., Bricio-Moreno, L., Wright, C. J., Maddocks, S. E., Sahota,
732 J. S., *et al.* The AgI/II Family Adhesin AspA Is Required for Respiratory Infection by
733 Streptococcus pyogenes. *PLoS One* **8**, (2013).
734 46. Purushotham, S. & Deivanayagam, C. The Calcium-induced Conformation and
735 Glycosylation of Scavenger-rich Cysteine Repeat (SRCR) Domains of Glycoprotein
736 340 Influence the High Affinity Interaction with Antigen I/II Homologs. *J. Biol. Chem.*
737 **289**, 21877–21887 (2014).
738 47. Daep, C. A., Novak, E. A., Lamont, R. J. & Demuth, D. R. Structural dissection and in
739 vivo effectiveness of a peptide inhibitor of porphyromonas gingivalis adherence to
740 streptococcus gordonii. *Infect. Immun.* **79**, (2011).
741 48. Halaby, D. M. & Mornon, J. P. E. The immunoglobulin superfamily: An insight on its
742 tissular, species, and functional diversity. *J. Mol. Evol.* **46**, (1998).
743 49. Olmsted, S. B., Dunny, G. M., Erlandsen, S. L. & Wells, C. L. A Plasmid-Encoded
744 Surface Protein On Enterococcus Faecalis Augments Its Internalization By Cultured
745 Intestinal Epithelial Cells. *J. Infect. Dis.* **170**, 1549–1556 (1994).

746 50. Bae, T., Kozlowicz, B. & Dunny, G. M. Two targets in pCF10 DNA for PrgX binding:
747 their role in production of Qa and prgX mRNA and in regulation of pheromone-
748 inducible conjugation. *J. Mol. Biol.* **315**, 995–1007 (2002).
749 51. Laverde Gomez, J. A., Bhatty, M. & Christie, P. J. PrgK, a Multidomain
750 Peptidoglycan Hydrolase, Is Essential for Conjugative Transfer of the Pheromone-
751 Responsive Plasmid pCF10. *J. Appl. Microbiol.* **196**, 527–539 (2014).
752 52. Geertsma, E. R. & Dutzler, R. A versatile and efficient high-throughput cloning tool
753 for structural biology. *Biochemistry* **50**, 3272–3278 (2011).
754 53. Kabsch, W. XDS. *Acta Crystallogr. Sect. D* **66**, 125–132 (2010).
755 54. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of
756 Coot. **66**, 486–501 (2010).
757 55. Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model
758 anisotropic displacements in macromolecular refinement. **57**, 122–133 (2001).
759 56. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J.,
760 Moriarty, N. W., *et al.* PHENIX: building new software for automated crystallographic
761 structure determination. **58**, 1948–1954 (2002).
762 57. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral,
763 G. J., *et al.* MolProbity: all-atom structure validation for macromolecular
764 crystallography. **66**, 12–21 (2010).
765 58. Punjani, A., Brubaker, M. A. & Fleet, D. J. Building Proteins in a Day: Efficient 3D
766 Molecular Structure Estimation with Electron Cryomicroscopy. *IEEE Trans. Pattern*
767 *Anal. Mach. Intell.* **39**, (2017).
768 59. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng,
769 E. C., *et al.* UCSF Chimera - A visualization system for exploratory research and
770 analysis. *J. Comput. Chem.* **25**, (2004).

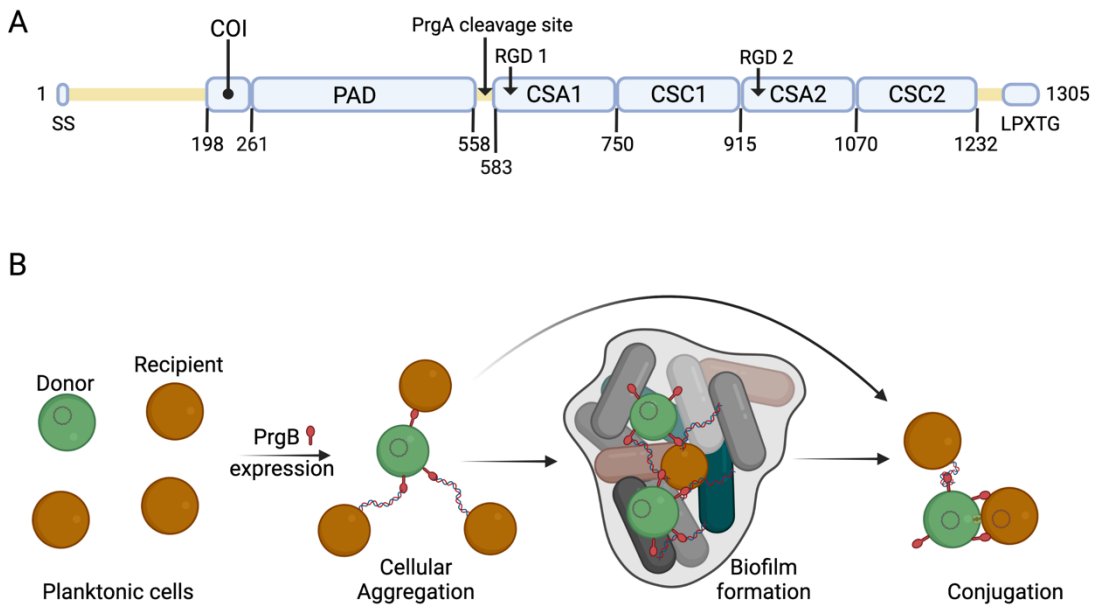
771 60. Kidmose, R. T., Juhl, J., Nissen, P., Boesen, T., Karlsen, J. L. & Pedersen, B. P.
772 Namdinator - Automatic molecular dynamics flexible fitting of structural models into
773 cryo-EM and crystallography experimental maps. *IUCrJ* **6**, (2019).
774 61. Willett, J. L. E., Ji, M. M. & Dunny, G. M. Exploiting biofilm phenotypes for
775 functional characterization of hypothetical genes in Enterococcus faecalis. *npj Biofilms*
776 *Microbiomes* **5**, (2019).
777 62. Christie, P. J., Kao, S. M., Adsit, J. C. & Dunny, G. M. Cloning and expression of
778 genes encoding pheromone-inducible antigens of Enterococcus (Streptococcus)
779 faecalis. *J. Appl. Microbiol.* **170**, 5161–5168 (1988).
780 63. Chen, Y., Staddon, J. H. & Dunny, G. M. Specificity determinants of conjugative
781 DNA processing in the Enterococcus faecalis plasmid pCF10 and the Lactococcus
782 lactis plasmid pRS01. *Mol. Microbiol.* **63**, 1549–1564 (2007).
783 64. Chen, Y., Zhang, X., Manias, D., Yeo, H.-J., Dunny, G. M. & Christie, P. J.
784 Enterococcus faecalis PcfC, a spatially localized substrate receptor for type IV
785 secretion of the pCF10 transfer intermediate. *J. Appl. Microbiol.* **190**, 3632–3645
786 (2008).
787
788

Tables

Table 1. X-ray data collection and refinement statistics. Values within parenthesis correspond 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

795 **Figures**



796

797 **Figure 1.** Schematic overview of PrgB domain organization and function. A) Updated
798 schematic overview of the domain organization of PrgB. SS: Signal sequence, COI: Coiled-
799 coil, PAD: Polymer adhesin domain, CSA: adhesin isopeptide-forming adherence domain,
800 CSC: cell-surface antigen C-terminal domain, LPXTG: cell wall anchor sequence. PrgA
801 cleavage site is located between the polymer adhesin domain and the first Ig-like domain, and
802 has the sequence IFNYGNPKEP. B) In a setting with a donor cell (green) and multiple
803 recipient cells (brown), PrgB is produced and sits on the cell wall. There it enhances cellular
804 aggregation and/or biofilm formation, either by directly binding LTA from the cell-wall of a
805 recipient or by binding first to eDNA. PrgB compacts the eDNA, and thereby likely pulls the
806 recipient cells closer. Once close enough, PrgB binds to the LTA of the recipient bacteria and
807 facilitates mating-pair formation and conjugation.

808

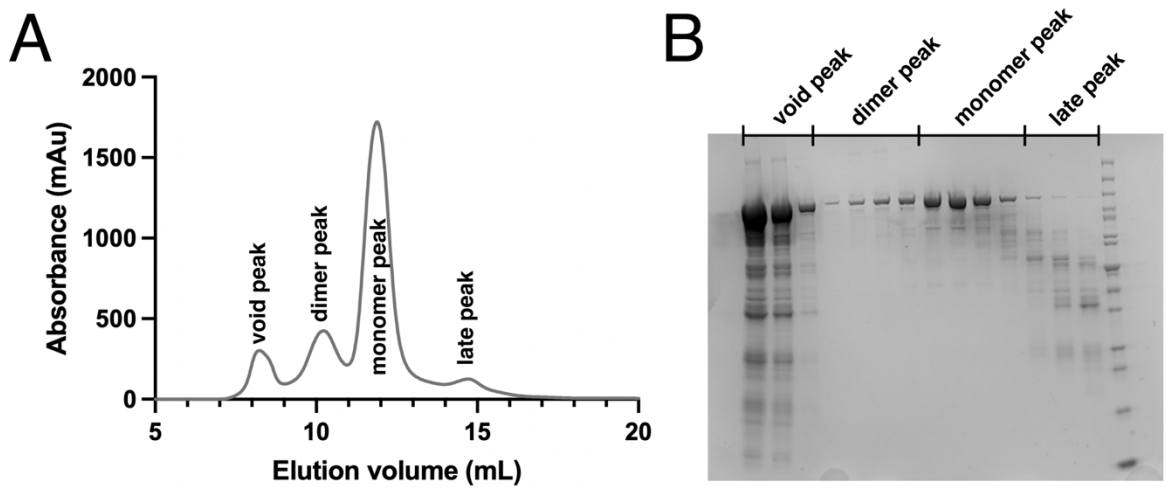


Figure 1 – figure supplement 1. Purification of PrgB. A) Representative chromatogram from size exclusion chromatography on a Superdex 200 Increase 10/300 GL column. 4 peaks are observed, corresponding to the void, PrgB dimer, PrgB monomer and a late peak. B) SDS-PAGE of the SEC elution fractions indicated in panel A.

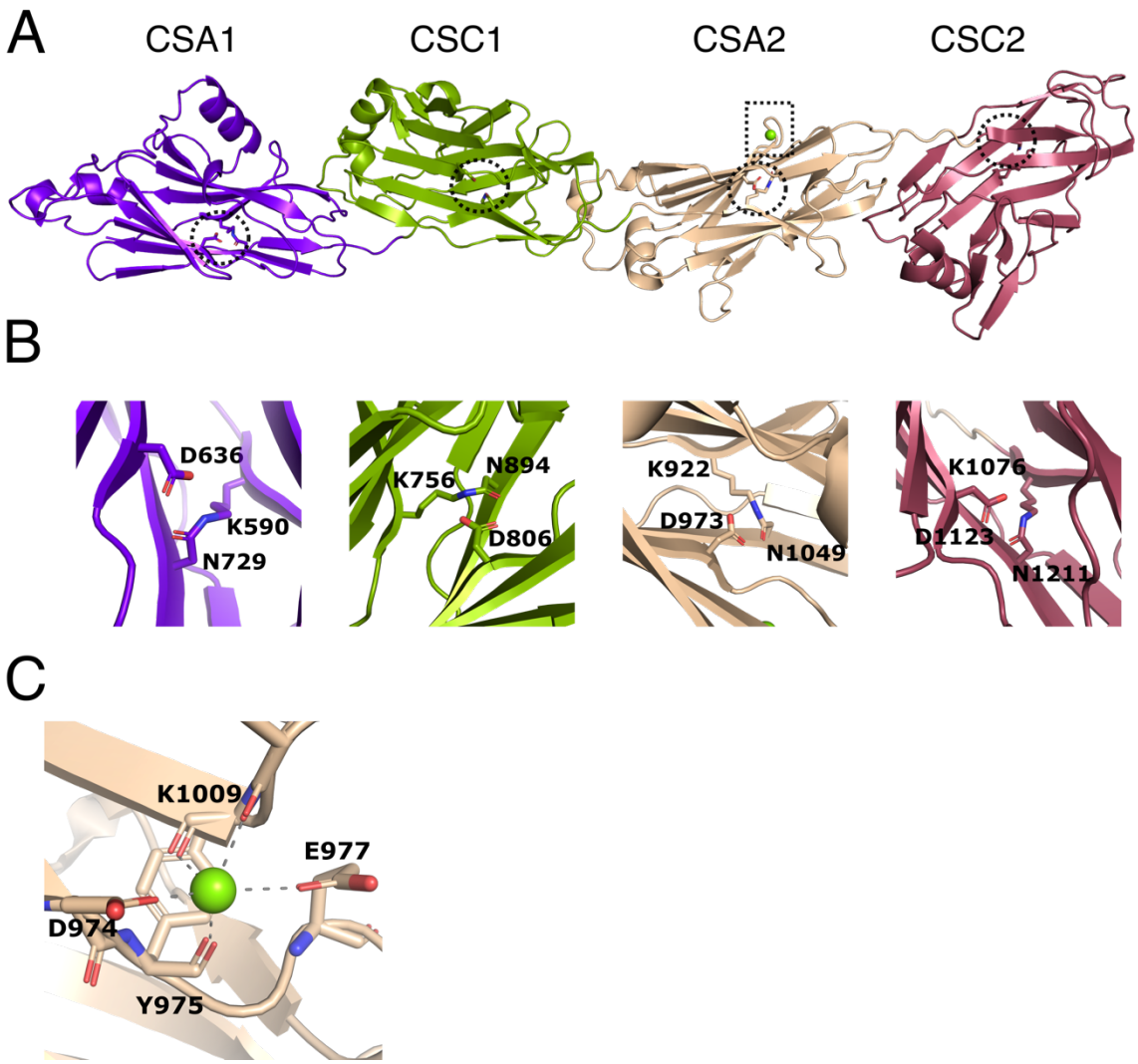
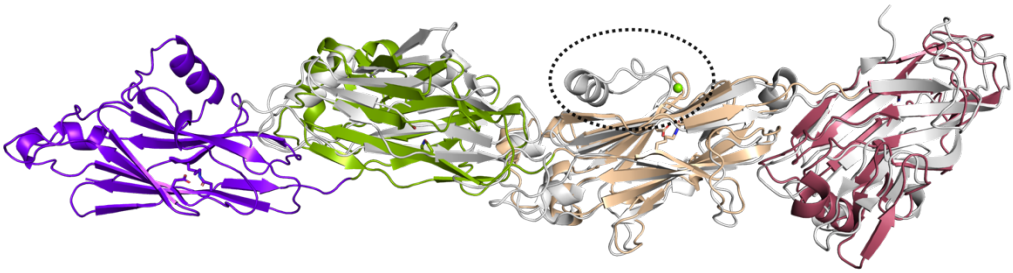


Figure 2. Structure of Ig-like domains of PrgB. A) The Ig-like domains are arranged as tandem pairs (CSA1-CSC1 and CSA2-CSC2). B) Each Ig-like domain has an internal isopeptide bond (indicated by the striped circle in panel A) strengthening the structural integrity of the domain. Each of the four isopeptide 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).



822

823 **Figure 2 – figure supplement 1.** Structure of PrgB compared to the C-terminal domain of Pas

824 (gray) (PDB Code: 6E3F). The BAR motif from Pas, missing in PrgB, is highlighted by the

825 striped circle.

826

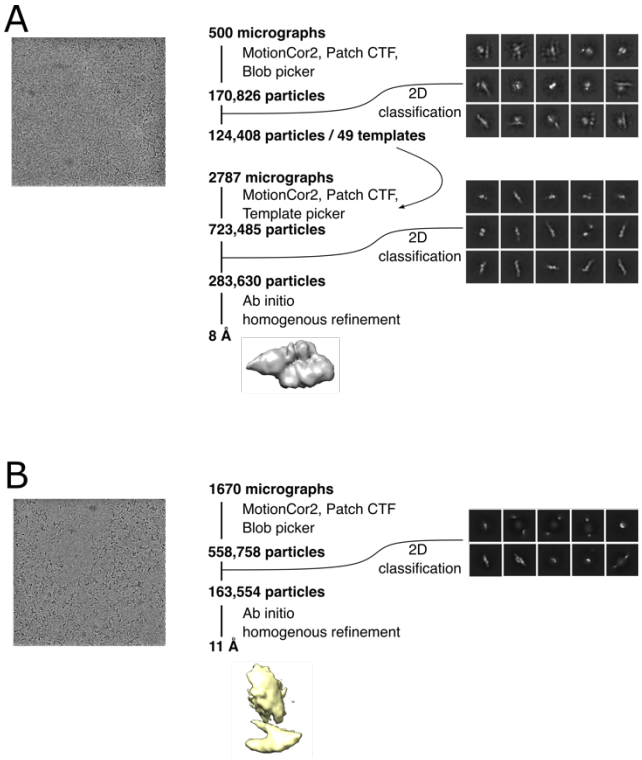


Figure 2 – figure supplement 2. Cryo-EM data processing scheme. Data collection and processing for the apo PrgB dataset (A) and PrgB bound to DNA (B). For both, representative micrographs and 2D classes used in the processing are shown.

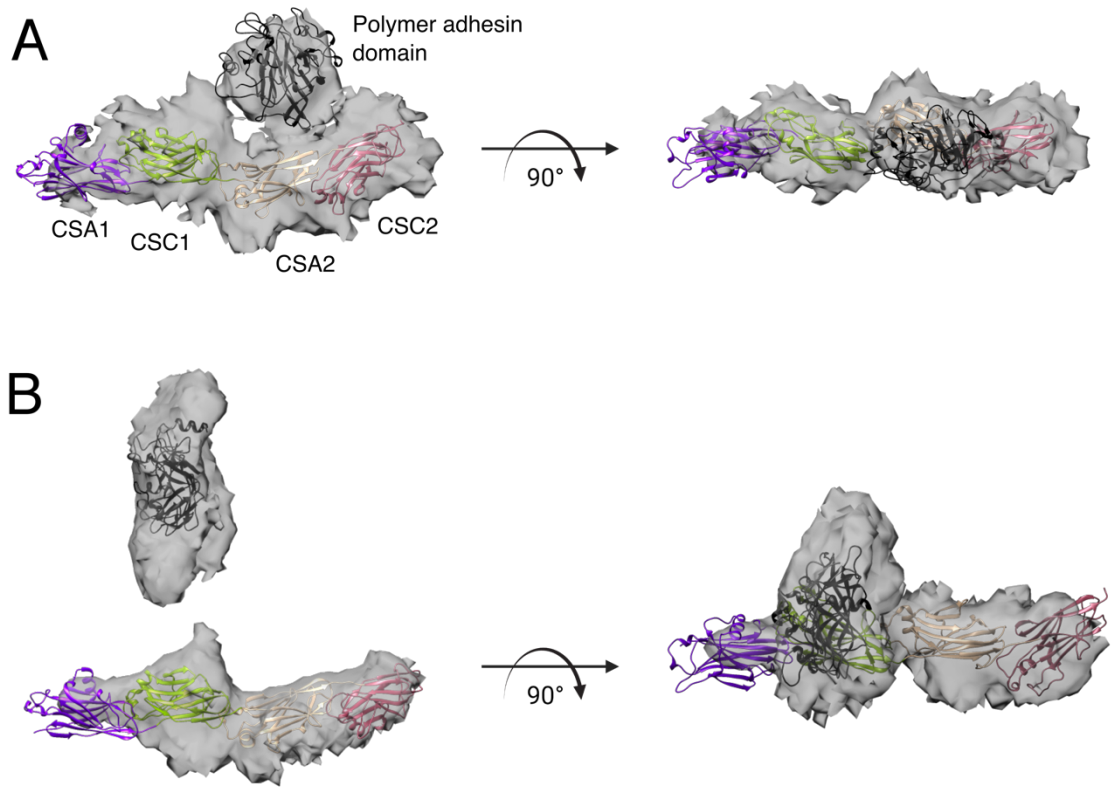


Figure 2 – figure supplement 3. Crystal structures of the PrgB polymer adhesin domain and Ig-like domains in cartoon representation docked into the volumes acquired by single particle cryo-EM. A) The PrgB188-1233 model shows the polymer adhesin domain folded back onto the Ig-like domains. B) In the DNA-bound PrgB188-1233 model the polymer adhesin domain has undergone a large conformation change away from the Ig-like domain.

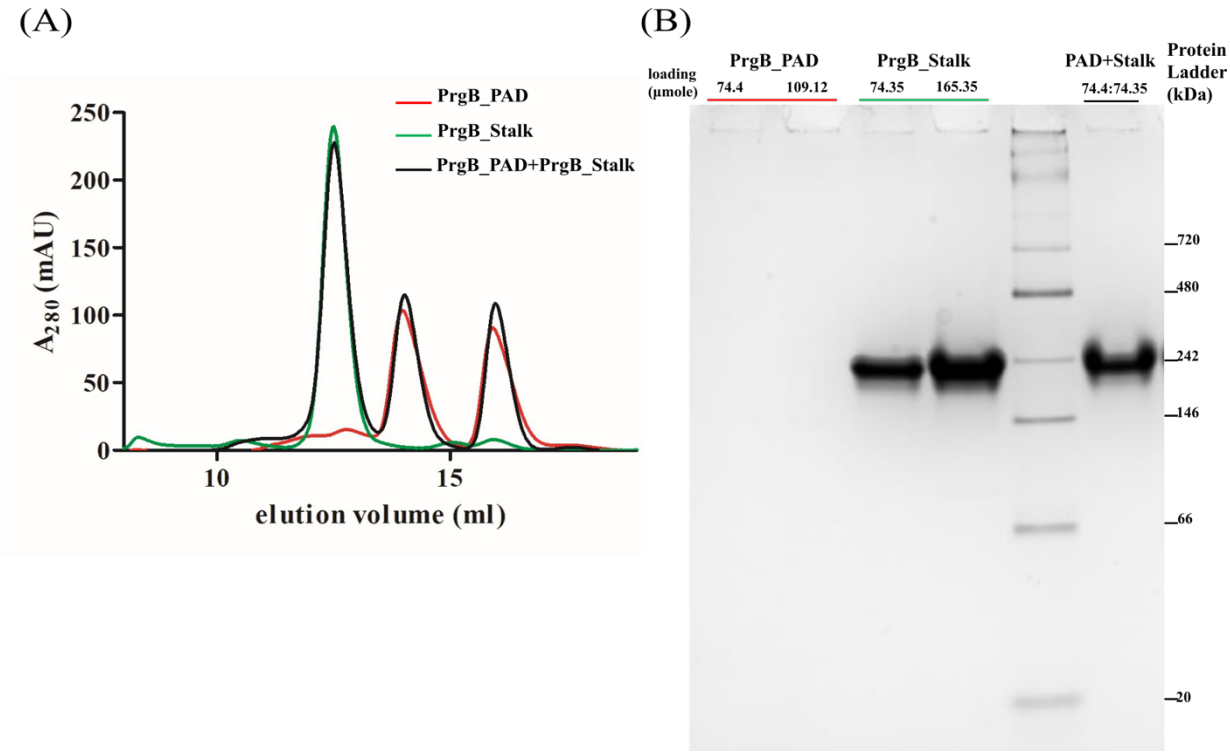


Figure 2 – figure supplement 4. Analysis of potential interaction of the Polymer adhesin domain and stalk-like domains of PrgB. (A) Size-exclusion chromatography of PrgB_PAD (red), PrgB_Stalk (green), and mixture (black). No shifted peaks are observed in the mixture. (B) Native PAGE electrophoresis of PrgB_PAD, PrgB_Stalk, and mixture in 4-20% native gradient gel. PrgB_PAD has a strongly positively charged surface, and does not enter the gel on its own. No shift is observed when mixed with the Stalk domain.

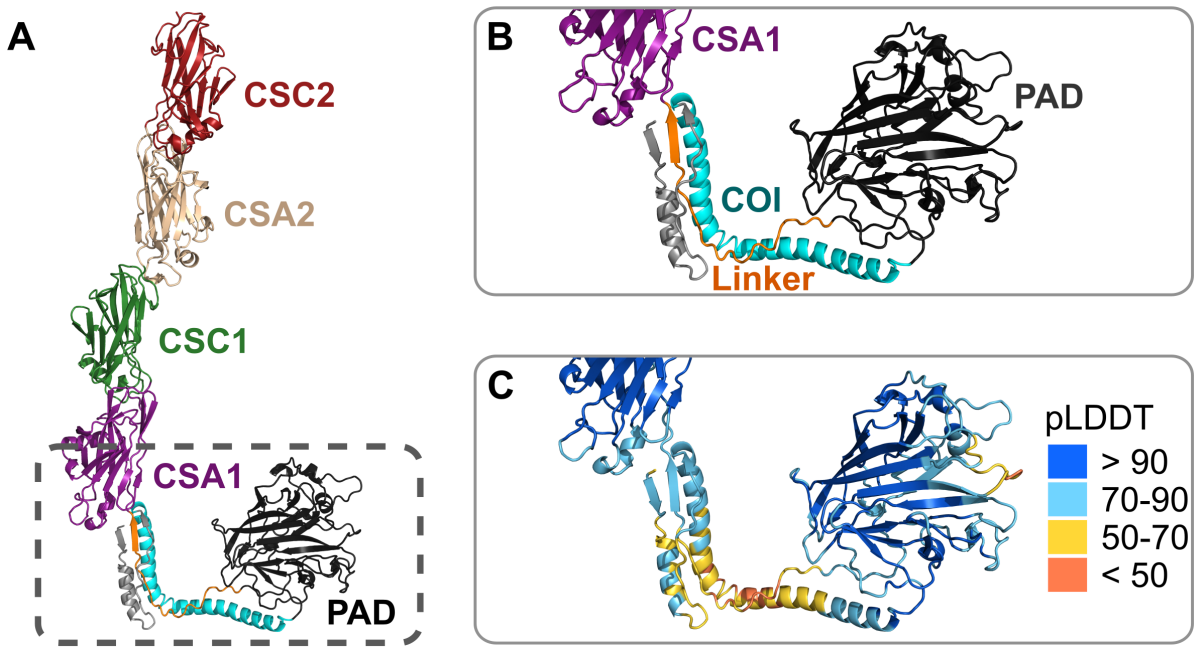


Figure 2 – figure supplement 5. AlphaFold model of PrgB. A) Model of PrgB with N-proximal residues 157-194 in grey, the coiled-coil (COI, residue 195-260) domain in cyan, the polymer adhesin domain (PAD, residue 261-558) in black, the following linker region (residue 559-583) in orange, CSA1 (residue 584-750) in purple, CSC1 in green, CSA2 in beige and CSC2 in red. Both N- and C-termini (residues 1-156 and 1233-1305) are removed for clarity as they were predicted to be largely unfolded. B) The linker region between the polymer adhesin domain and the CSA1 domain, as indicated in panel A. C) The same region as shown in panel B with the residues colored by their confidence measure (pLDDT score). Regions with a pLDDT score above 90 are expected to be modelled with high accuracy, regions with a score between 70-90 are expected to have a good prediction for the backbone, while regions with a pLDDT score between 50-70 are off low confidence. A pLDDT score < 50 is a reasonable strong predictor of disorder.

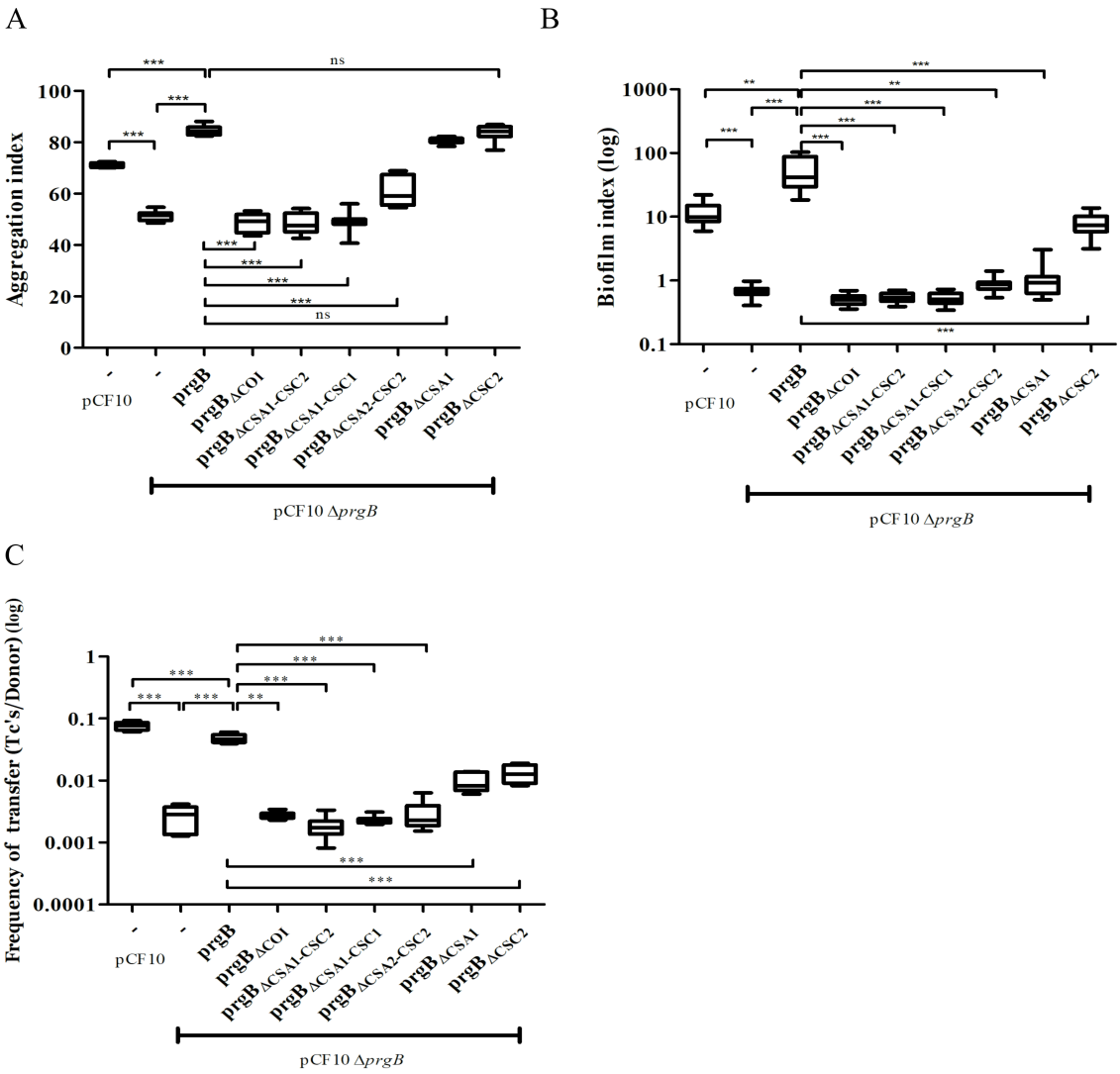


Figure 3. *In vivo* phenotypes of PrgB variants. PrgB variants are expressed from the pMSP3545S-MCS vector in the OG1RF pCF10Δ*prgB* background for phenotypic analysis with three assays: A) cellular aggregation B) biofilm formation and C) conjugation assays. For all assays OG1RF pCF10 carrying an empty 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).

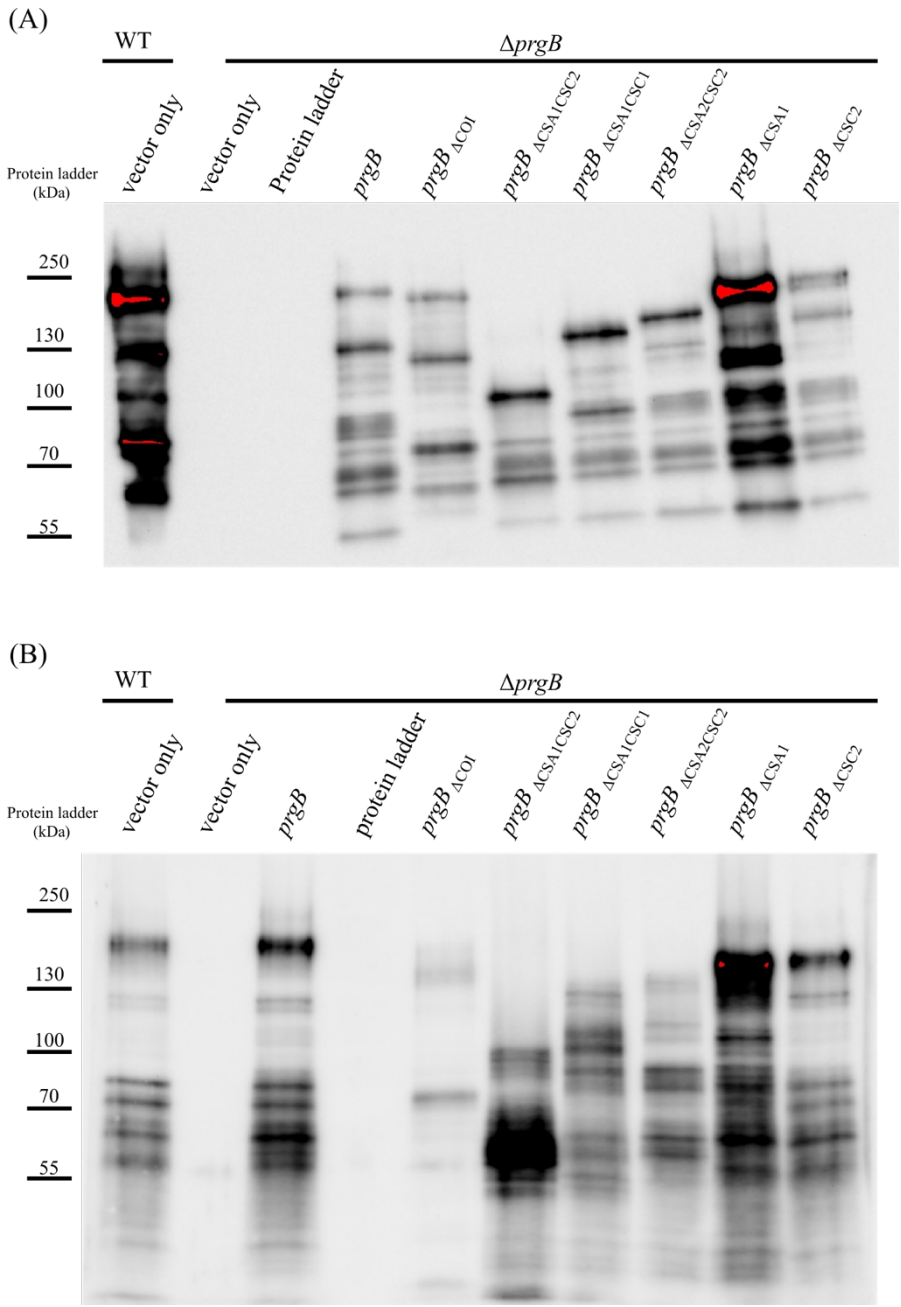


Figure 3 – figure supplement 1. Western-blot showing the expression levels of the PrgB variants in *E. faecalis* cell wall extracts. (A) Samples collected after 1 hour induction with cCF10 and nisin. (B) Samples collected after overnight induction with cCF10 and nisin. Estimated molecular weight of PrgB variants: WT (140 kDa); ΔCOI (137 kDa); ΔCSA1CSC2 (72 kDa); ΔCSA1CSC1 (103 kDa); ΔCSA2CSC2 (105 kDa); ΔCSA1 (121 kDa); ΔCSA2 (122 kDa).

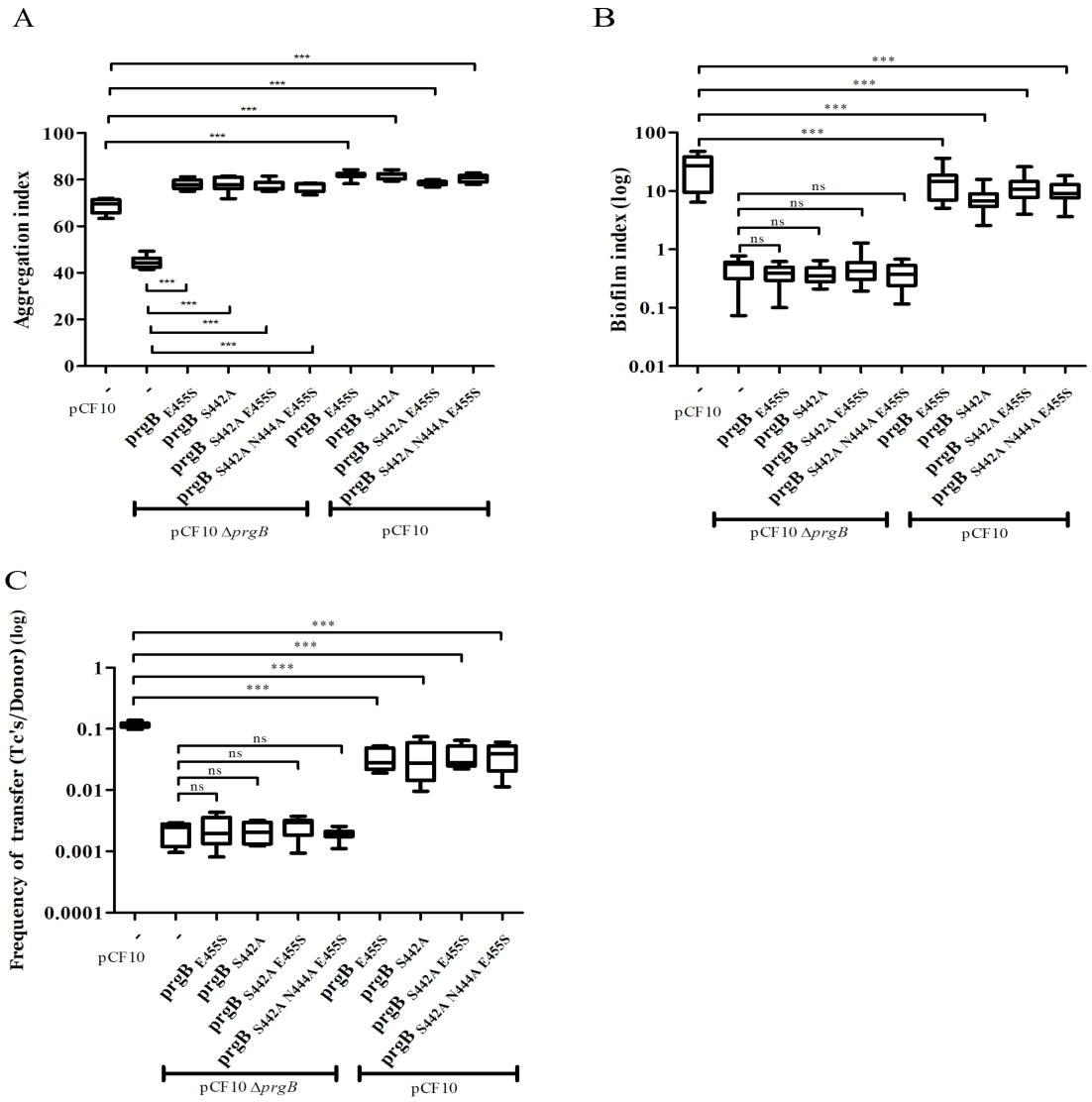
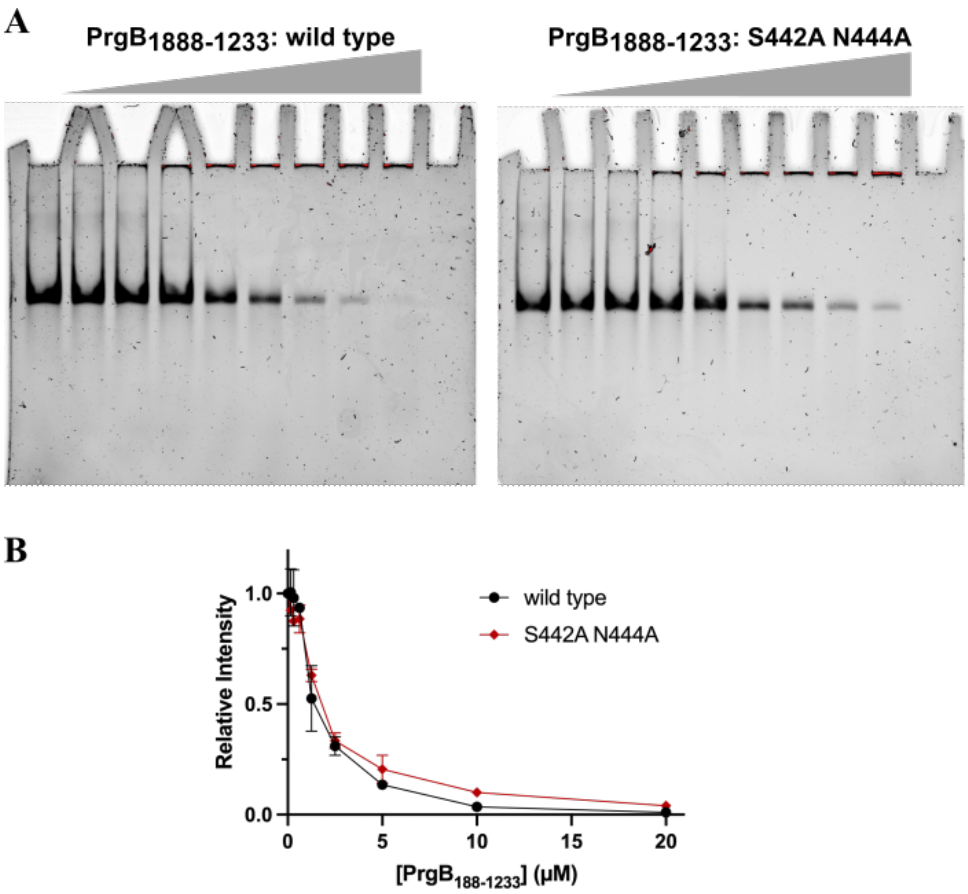


Figure 4. *In vivo* phenotypes of PrgB variants with point mutation(s) in the conserved site of the polymer adhesin domain. PrgB variants are expressed from the pMSP3545S-MCS vector in the OG1RF pCF10Δ*prgB* or OG1RF pCF10 strain and analyzed in: A) cellular aggregation, B) biofilm formation, and C) 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).



894

895 **Figure 4 – figure supplement 1.** Comparison of the DNA-binding affinity of the wild-type
896 polymer adhesion domain from PrgB ($PrgB_{188\text{-}1234\ WT}$) and its S442A & N444A variant
897 ($PrgB_{188\text{-}1234\ S442A\ N444A}$). A) For this mobility shift assay, protein was mixed and incubated with
898 50 nM DNA_{100} and applied to native gels. From left to right, DNA mixtures with 0, 0.16, 0.31,
899 0.63, 1.25, 2.5, 5, 10 or 20 mM protein were loaded and the final lane is a negative control with
900 only 20 mM protein (and no DNA). B) Plot of the relative intensities of the unbound (not
901 upshifted) DNA bands normalized against the protein-free condition (Lane 1). The error is the
902 standard deviation (N = 2).

903

904

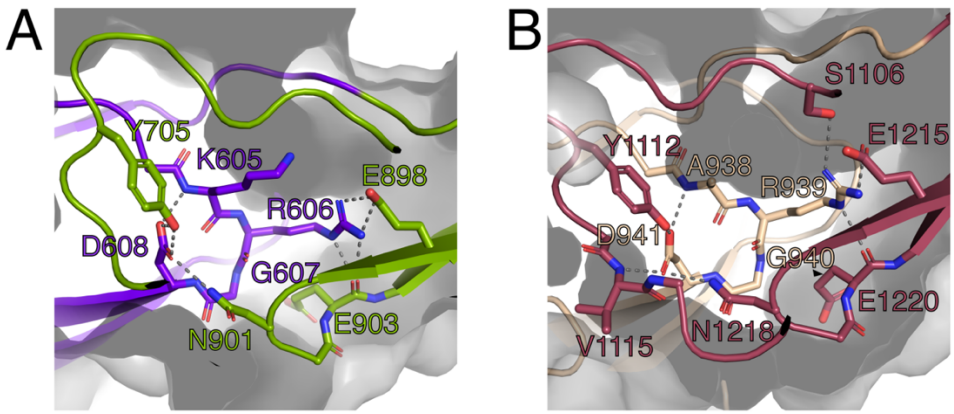
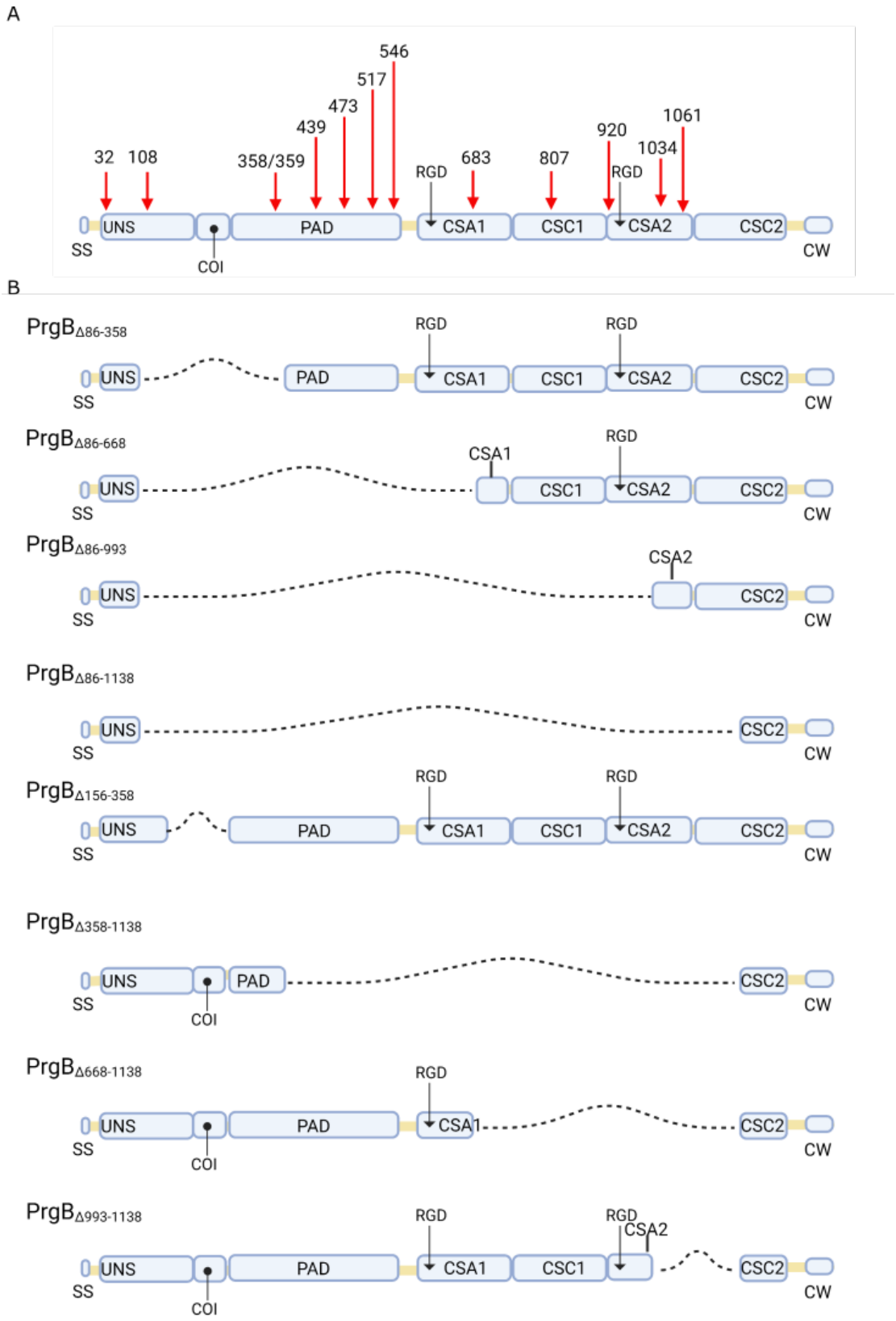


Figure 5. The RGD motifs play a role in the structural integrity of PrgB. A) Close up view of the RGD motif in the CSA1 domain (purple). Most of the important interactions, mainly hydrogen bonds, formed by this motif are to residues on the CSC1 domain (green). B) Close up view of the RGD motif in the CSA2 domain (sand colored). Most of the important interactions, mainly hydrogen bonds, formed by this motif are 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.



913

914 **Figure 5 – figure supplement 1.** Schematic overview of previously described *prgB* mutants.

915 A) Red arrows indicate the sites of transposon insertions that cause defective phenotypes in

916 aggregation, biotic tissue biofilm formation, or conjugation. B) Diagrams illustrating the
917 specific deletion mutants discussed in the main text. Dashed lines indicate deleted regions. SS,
918 signal sequence; UNS, unstructured region; PAD, polymer adhesin domain; COI, coiled-coil
919 domain; CSA, cell surface antigen; CSC, cell surface antigen C-terminus; CW, LPXTG-
920 containing cell wall anchor sequence.
921

922

923 **Supplementary files**

924 Supplementary File 1**.** Strains, plasmids and oligonucleotides

925 Supplementary File 2. Cryo-EM data collection and refinement

926 Supplementary File 3. Top hits of Dali search based on PDB of PrgB Ig-like domains

927 Supplementary File 4. Structural interpretation of previously observed PrgB phenotypes

928