

1 **The *Streptococcus agalactiae* R3 surface protein is encoded by *sar5***

2

3 Marte Singsås Dragset^{1,2,*}, Adelle Basson², Camilla Olaisen³, Linn-Karina Selvik^{1,2}, Randi

4 Valsø Lyng³, Hilde Lysvand², Christina Gabrielsen Aas^{2,3}, Jan Egil Afset^{2,3}

5

6 ¹ Centre for Molecular Inflammation Research (CEMIR), Norwegian University of Science and

7 Technology (NTNU), Trondheim, Norway

8 ² Department of Clinical and Molecular Medicine, Norwegian University of Science and

9 Technology (NTNU), Trondheim, Norway

10 ³ Department of Medical Microbiology, St. Olavs University Hospital, Trondheim, Norway

11

12 *Corresponding author

13 E-mail: marte.dragset@ntnu.no

14

15

16

17

18

19

20

21

22

23

24

25

26 **ABSTRACT**

27

28 *Streptococcus agalactiae* (a group B streptococcus; GBS) is an important human pathogen
29 causing pneumonia, sepsis and meningitis in neonates, as well as infections in pregnant women,
30 immunocompromised individuals, and the elderly. For the future control of GBS-inflicted
31 disease, GBS surface exposed proteins are particularly relevant as they may act as antigens for
32 vaccine development and/or as serosubtype markers in epidemiological settings. Even so, the
33 genes encoding some of the surface proteins established as serosubtype markers by antibody-
34 based methods are still unknown. Here, we identify *sar5* as the gene encoding the R3 surface
35 protein, a serosubtype marker of hitherto unknown genetic origin.

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51 **INTRODUCTION**

52

53 *Streptococcus agalactiae* (a group B streptococcus; GBS) is an important human pathogen,
54 most notably in neonates, but also in pregnant women as well as immunocompromised and
55 elderly individuals. Worldwide, 18 % of pregnant women are colonized with GBS in their
56 rectovaginal tract (1). Colonization of GBS during pregnancy is a risk factor for preterm birth,
57 stillbirth, and neonatal infection (2). To reduce the risk of vertical transmission of GBS to the
58 neonate during birth, routine screening for GBS colonization followed by intrapartum
59 antibiotic prophylaxis (IAP) to pregnant women with GBS is recommended (3). However,
60 administration of IAP poses a risk of allergic and anaphylactic reactions and the widespread
61 use of antibiotics may result in the emergence of antibiotic resistance. Another option to
62 prevent GBS infection is vaccine development. Currently, conserved GBS surface proteins are
63 considered as promising targets for vaccine development (4), as they may elicit a strong
64 immune response against the majority of GBS strains (5).

65

66 GBS surface proteins also play an important role as serosubtype markers, relevant for GBS
67 classification in epidemiological settings. While GBS strains can be distinguished into ten
68 serotypes due to differences in their capsular polysaccharide (CPS) (Ia, Ib, and II – IX),
69 surface-expressed protein antigens enable further division of these serotypes. Some of the
70 surface proteins are conserved and present in nearly all GBS strains, while others are associated
71 with specific serotypes, and thus used to define serosubtypes (6). Historically, detection of
72 serosubtypes by means of antibody-based methods has played a major role. In more recent
73 years, serosubtyping of GBS has benefitted greatly from the introduction of molecular
74 methods, such as PCR and whole genome sequencing (WGS) (7, 8).

75

76 GBS surface proteins have been classified according to two different and overlapping
77 classification systems (Table 1). However, there is still some discrepancy and confusion
78 surrounding the traditional nomenclature, and some surface proteins that have not yet been
79 definitely linked to a specific gene. One classification scheme of GBS surface proteins includes
80 C β and the C α -like proteins (Alps) C α , Alp1-4 and Rib. Nearly all GBS strains carry one of
81 the six alp genes (Alp GBS) although, occasionally, an Alp-encoding gene may be absent (non-
82 Alp GBS) (9). Another, and overlapping, classification system of GBS surface proteins is the
83 Streptococcal R proteins first described in 1952 (10), which are resistant to trypsin digestion
84 (thereby designated “R”). R proteins are categorized into five types, R1-5 (11-13). R1 is
85 probably non-existent as a distinct protein; the antiserum raised against R1 was later shown to
86 recognize the identical N-termini of Alp2 and Alp3, the gene products of *alp2* and *alp3*,
87 respectively (14). The R2 protein is expressed by group A and C streptococci and does not
88 seem to occur in GBS (13). The R4 protein has been shown to be identical to Rib and is encoded
89 by the *rib* gene (15), while R5 has been renamed group B protective surface protein (BPS) and
90 was shown to be the gene product of *sar5* (13, 16). The R3 protein has been characterized to
91 some extent (12, 17-19), and has proved useful as a serosubtype GBS marker (20, 21).
92 However, the gene encoding the R3 protein is still unknown (Table 1). BPS was initially
93 thought to be distinct from R3 (13), however, a later study pinpointed a correlation between
94 the presence of the BPS-encoding *sar5* gene and R3 expression (6). Here, we follow up on this
95 correlation, hypothesizing that *sar5* encodes R3. Unraveling the R3-encoding gene, and the
96 putative discrepancy in the nomenclature and nature of the *sar5* gene product, is important for
97 the *sar5* gene product as a prospective target in vaccine development and molecular based GBS
98 serosubtyping, as well as for functional studies on its mechanistic role in pathogenicity.

99

100 **Table 1.** Surface-proteins of GBS. Alps (in blue) and R proteins (in green).

Name	Gene	GenBank Number
Cα	<i>bca</i> ⁽²²⁾	M97256
Alp1 (epsilon)	<i>alp1</i> ⁽²³⁾	AH013348.2
Alp2/R1	<i>alp2</i> ^(14, 24)	AF208158
Alp3/R1	<i>alp3</i> ^(14, 24)	AF245663
Alp4	<i>alp4</i> ⁽²⁵⁾	AJ488912
R3	<i>unknown</i> *	-
Rib/R4	<i>rib</i> ⁽¹⁵⁾	U583333
R5/BPS	<i>sar5</i> ⁽¹³⁾	AJ133114

101 * in this study determined to be encoded by *sar5*.

102

103 RESULTS

104

105 Presence of the *sar5* gene correlates with R3 protein expression across GBS strains.

106 In a previous study, 121 GBS strains collected from pregnant women in Zimbabwe were tested
107 for (among other markers) the presence of the *sar5* gene and R3 protein expression (6). The
108 study found that 31 out of 35 (91.5%) *sar5* positive strains expressed R3. The remaining 86
109 strains were negative for both *sar5* and R3. Based on these findings we speculated that *sar5*
110 could encode R3, and that, consequently, the previously reported *sar5*-encoded BPS and the
111 R3 protein are the same protein. To further investigate this observed association between *sar5*
112 and R3 expression, we analyzed 140 clinical GBS strains from neonatal and adult GBS
113 infections from the Norwegian GBS reference laboratory (18). These strains were previously
114 characterized for R3 expression (and other serotype markers) by fluorescent antibody testing
115 using a monoclonal R3 antibody (18). This R3 antibody has been used and evaluated in several
116 previous studies (6, 20, 21, 26-28). We typed the strains for presence/absence of *sar5* using a
117 previously established PCR approach (6), with the R3 reference strain CCUG 29784 (also
118 known as Prague 10/84) as a positive control. Of the 140 GBS strains, the majority were *sar5*
119 negative (131), while nine strains were *sar5* positive. Seven of these strains were R3 positive

120 (Table 2, S1 Figure, and S1 Table). Hence, there was a strong, albeit not perfect, correlation
121 between the presence of the *sar5* gene and R3 expression across the 140 investigated GBS
122 strains.

123

124 **Table 2.** Distribution of the *sar5* gene and R3 expression among 140 GBS clinical strains.

	R3 positive	R3 negative	Sum
<i>sar5</i> positive	7	2	9
<i>sar5</i> negative	0	131	131
Sum	7	133	140

125

126

127 **The *sar5* positive R3 negative GBS strains express R3 but encode a *sar5* deletion variant.**

128 Two strains, 93-33 and 94-3, contained the *sar5* gene but were negative for R3 expression.

129 Thus, these strains were in conflict with our hypothesis that *sar5* encodes R3. The initial

130 detection of R3 in the 140 GBS strains included in this study was performed by fluorescent

131 monoclonal antibody testing on whole bacterial cells (18, 29). We speculated whether R3 from

132 93-33 and 94-3 could be detected by western blot analysis of denatured proteins from cell

133 lysates. Indeed, using the same R3 monoclonal antibody as in the initial whole cell R3

134 fluorescence testing, we could clearly see that both 93-33 and 94-3 expressed R3, although in

135 a seemingly truncated form compared to the R3 reference strain CCUG 29784 (Figure 1). For

136 all three strains the blot displayed a ladder-like pattern characteristic for the R3 protein (18).

137 The largest fragment was around the expected size of 109 kDa for the control strain (in

138 accordance with the size of the *sar5* gene at 2940 bp) and around 90 kDa for the 93-33 and

139 94-3 strains. This finding prompted us to subject GBS strains 93-33 and 94-3 to nanopore

140 WGS, to investigate whether the *sar5* genes of 93-33 and 94-3 could be truncated. Indeed, both

141 GBS strains possessed a *sar5* gene with an identical 531 bp in-frame deletion towards the 3'

142 end of the gene, when compared to the *sar5* gene of NCTC 9828 (the *sar5* reference strain
143 (13)) (Figure 2). The deletion occurred between two 102 bp long direct repeat regions, making
144 it feasible that the 531 bp region has been deletion by homologous recombination. The 531 bp
145 deletion corresponds perfectly to the 20 kDa difference in size between the R3 control strain
146 and the 93-33 and 94-3 strains observed by western blot analysis (Figure 1). Taken together,
147 our results show that the two *sar5* positive but initially R3 negative strains indeed express R3,
148 although in a truncated form compared to the control strain, and that they both possess a
149 deletion variant of the *sar5* gene.

150

151 **Figure 1:** Western blot analysis of GBS whole cell denatured lysates from strains 93-33 and
152 94-3. The blots were probed with α -R3 antibody (upper panel) and α -GAPDH antibody (lower
153 panel). The CCUG 29779 strain, known to not bind to the R3 antibody, serves as a negative
154 control. The CCUG 29784 strain, known to bind to the R3 antibody, serves as a positive
155 control. The protein standards are shown to the left of the blots.

156

157 **Figure 2.** (A) Alignment of the gene sequences of strains 93-33 and 94-3 to the *sar5* gene of
158 *sar5* reference strain NCTC 9828. Identity between all sequences is indicated by the top panel
159 in green and gene annotation is shown in yellow. Black vertical lines indicate mismatches at
160 the nucleotide level, while grey boxes indicate matching nucleotides to the reference strain.
161 93-33 and 94-3 have a 531 bp long deletion (marked with horizontal line) within the *sar5* gene.
162 The binding sites of the primers used to detect the *sar5* gene are shown as red triangles. Repeat
163 regions are shown in blue. (B) Close up of the 531 bp long *sar5* deletion in 93-33 and 94-3.
164 The figure is annotated as described for (A) with the additional visualization of *sar5* coding
165 strand's bases.

166

167 **The *sar5*-encoded protein is recognized by the R3 antibody.**

168 Based on the above results, we had strong indications that the *sar5* gene encodes the R3 protein.
169 We aimed to prove this experimentally by inducing *sar5* protein expression in a *sar5* negative
170 bacterial species, followed by R3 protein detection. First, we constructed a *sar5* inducible
171 expression vector by replacing the luciferase reporter gene of pKT1 (30) with *sar5*, creating
172 pKT1-*sar5*-F. In pKT1, luciferase expression is controlled by the XylS/*Pm* regulator/promoter
173 system, which is induced by the benzoic acid *m*-toluate. We added a FLAG tag to the C-
174 terminal end of the *sar5*-encoded protein, to allow for successful detection of the *sar5*-encoded
175 protein also if the protein was not recognized by the R3 specific antibody (Figure 3). Even so,
176 upon *m*-toluate induction of pKT1-*sar5*-F in *Escherichia coli* BL21 (DE3), we could clearly
177 detect the *sar5*-encoded 109 kDa protein with the monoclonal R3 antibody (Figure 4). We
178 could furthermore detect the FLAG-tag expressed from pKT1-*sar5*-F around the same
179 expected size of 109 kDa, confirming that it is indeed the *sar5* gene product that is detected.
180 When we induced pKT1 (expressing luciferase as opposed to *sar5*), we did not detect
181 expression of any protein around 109 kDa, neither with the R3 nor the FLAG-specific
182 antibodies.

183

184 Since the two *sar5* positive but initially R3 negative strains 93-33 and 94-3 actually expressed
185 R3, and were shown to possess a deletion variant of *sar5* (*sar5D*), we wanted to investigate
186 whether this deletion variant also encoded a protein which is recognized by the R3 antibody.
187 Hence, we constructed an inducible vector expressing FLAG-tagged *sar5D* (pKT1-*sar5D*-F,
188 Figure 3), and subjected it to induction and western blot analysis. As for the full-length *sar5*,
189 both the R3 and the FLAG antibody bound to the induced *sar5D* gene product (Figure 4).
190 Compared to the full-length *sar5*, the *sar5D* gene encoded a seemingly truncated R3 protein,

191 corresponding in size to the R3 protein expressed by the 93-33 and 94-3 strains. Taken together,
192 our results demonstrate that *sar5* encodes a protein recognized by the R3-specific antibody.

193

194 **Figure 3:** *To induce expression of sar5, the luciferase reporter gene of pKT1 (30) was replaced*
195 *by FLAG-tagged sar5 and FLAG-tagged sar5D, creating pKT1-sar5-F and pKT1-sar5D-F,*
196 *respectively. xylS, gene encoding the transcription activator XylS. Pm, promoter at which XylS*
197 *binds and activates transcription in response to the inducer m-toluate. kanR, gene encoding*
198 *resistance to kanamycin. oriV, origin of replication for RK2-derived plasmids. trfA, gene*
199 *encoding plasmid replication initiator protein TrfA, activating replication by binding to oriV.*
200 *rrnBT1T2, transcriptional terminator. oriT, origin of conjugal transfer. The plasmid maps*
201 *were generated using SnapGene software (from Insightful Science).*

202

203 **Figure 4:** *Western blot analysis of whole cell lysates from E. coli BL21 (DE3) carrying pKT1,*
204 *pKT1-sar5-F or pKT1-sar5D-F plasmids, induced with 2 mM m-toluate (+) or mock-induced*
205 *with the equivalent amount of the solvent ethanol (-). The blots were probed with α -R3*
206 *antibody (upper panel), α -FLAG antibody (middle panel), or α -GAPDH antibody (lower*
207 *panel). The protein standards are shown to the left of the blots.*

208

209 MATERIALS AND METHODS

210

211 **Bacterial strains.** Included in this study were 140 clinical GBS strains collected in the years
212 between 1993 and 1995 from neonatal or adult GBS disease from the Norwegian national
213 reference laboratory for GBS, Department of Medical Microbiology, St Olavs Hospital,
214 Trondheim, Norway. These strains were previously characterized for R3 expression (18), as
215 shown in S1 Table. *S. agalactiae* CCUG 29784 was used as an R3 reference strain, while *S.*

216 *agalactiae* CCUG 29779 (both from Culture Collection University of Gothenburg, Sweden)
217 was used as an R3 negative control in western blot analysis. *S. agalactiae* NCTC 9828 (also
218 called ComptonR, (13)) was used as a reference strain for the *sar5* gene in analysis of *sar5*
219 gene sequences. The GBS strains were cultured overnight on blood agar medium or in Todd-
220 Hewitt broth at 35° C.

221
222 **Detection of *sar5* by PCR.** Bacterial cells from a single colony were suspended in 100 µl TE-
223 buffer and 100 µl lysis buffer (1% Triton X-100, 0.5% Tween 20, 10 mM Tris-HCl with pH 8
224 and 1 mM EDTA) (31). The mixture was incubated at 95 °C for 15 minutes and centrifuged at
225 14 500 rpm for 2 minutes before 100 µl of the supernatant was transferred to a new tube. This
226 material was used as template in the PCR reaction, with AmpliTaq Gold DNA Polymerase with
227 Buffer I (5U/µl; Applied Biosystems). The *sar5*-specific primers used were identical to those
228 of Mavenyengwa et al (6).

229
230 **DNA isolation, WGS and assembly.** Bacterial cells were suspended in TE buffer and treated
231 with proteinase K (1.5 mg/mL), lysozyme (0.5 mg/mL) and mutanolysin (250 U/mL) for 15
232 minutes with shaking at 37 °C, before heating at 65 °C for 15 minutes. RNase A (2 mg/mL)
233 was then added to the lysate. Genomic DNA was subsequently isolated using the EZ1 DNA
234 tissue kit with an EZ1 Advanced XL instrument (Qiagen). Illumina sequencing libraries were
235 prepared using the Nextera XT sample prep kit and sequenced on the Illumina MiSeq platform
236 (Illumina) with 300-bp paired-end read configuration (MiSeq Reagent Kit v3). Nanopore
237 sequencing libraries were prepared using the Rapid Sequencing Kit (SQK-RAD004) and
238 sequenced on a minION instrument with Flongle adapter and flowcells (FLO-FLG001) (Oxford
239 Nanopore Technologies). Raw nanopore data was basecalled using Guppy v5.0.13 and
240 assembled using Flye v2.7. Assemblies were polished with nanopore data using Racon v1.4.20

241 and with Illumina.data using Pilon v1.23. Geneious vR9 was used for alignments and
242 visualization.

243

244 **Cloning of *sar5* into an inducible expression vector.** The *sar5* gene was cloned into the *m*-
245 toluate-inducible expression vector pKT1 (30). Briefly, the *sar5* ORF of GBS strain 13/87
246 (identical to the BPS-encoding gene of strain NCTC 9828 (13)) was amplified. To incorporate
247 a C-terminal FLAG-tag, the sequence encoding the FLAG epitope (DYKDDDDK), was
248 incorporated into the *sar5*-amplification reverse primers. In addition, the primer sets
249 amplifying *sar5* and the pKT1 vector backbone were extended with overlaps to enable Gibson
250 Assembly with the Gibson Assembly® Master Mix (NEB). KOD Xtreme™ Hot Start DNA
251 Polymerase (Sigma-Aldrich) was used for PCR amplifications. Illustrations generated using
252 SnapGene software (Insightful Science; available at snapgene.com) of the cloning strategy and
253 the primers used are found in S1 Materials and Methods.

254

255 **Induced expression of *sar5*.** For expression of *sar5* in *E. coli*, pKT1 (negative control), pKT1-
256 *sar5*_F and pKT1-*sar5D*_F were transformed into *E. coli* strain BL21 (DE3) and grown in LB
257 medium supplemented with 50 µg/ml kanamycin to stationary phase, then diluted
258 approximately 1:500 and grown to OD₆₀₀ 0.05–0.1 at 37 °C. At this point, the samples were
259 adjusted to the same OD₆₀₀, induced with 2 mM of *m*-toluate (Sigma, 1 M stock solution
260 solved in laboratory grade ethanol) and incubated for 5 hours with shaking at 30°C. For
261 uninduced samples the equivalent amount of ethanol was added as a mock treatment.

262

263 **Preparation of protein extracts and detection of protein expression by western blot**
264 **analysis.** Overnight cultures of GBS strains were pelleted by centrifugation and washed in

265 PBS. The pellets were resuspended in 1X LDS Sample Buffer (NuPage®, Invitrogen) with 50
266 mM dithiothreitol (DTT) and heated for 10 minutes at 95 °C. Samples were cleared for cellular
267 debris by centrifugation. To prepare protein extracts of *E. coli*, induced (or mock induced)
268 cultures were adjusted to OD₆₀₀ ~0.8, and pelleted by centrifugation. The pellet was
269 resuspended in 50 µl 1x LDS Loading Buffer with 50 mM DTT, heated for 10 minutes at 70°C
270 and sonicated 3 times for 1 minute each. Protein extracts were separated on 4-12% Bis-Tris
271 mini protein gels (NuPage®, Invitrogen) and blotted on polyvinylidene fluoride membranes
272 (Bio-Rad). Membranes were blocked with 1X blocking buffer (Roche) in PBS. The primary
273 antibodies against R3 (mouse monoclonal, from (18)), FLAG (monoclonal mouse anti-FLAG
274 M2 antibody, Sigma), and GAPDH antibody GA1R (Covalab) as well as the HRP conjugated
275 secondary antibody goat anti-mouse (Dako) were diluted in 0.5X blocking buffer/PBS. The
276 bound HRP-conjugated antibodies were visualized using SuperSignal™ West Femto
277 Maximum Sensitivity Substrate (Thermo Scientific) and Odyssey Fc imaging system (Licor).

278

279 **DISCUSSION**

280 GBS *sar5* was previously shown to encode BPS, a protein initially described to be different
281 from the R3 surface protein (13). Correlation between R3 expression and the presence of the
282 *sar5* gene was observed within a previously examined GBS strain collection, where 31 out of
283 35 (91.5%) *sar5* positive strains expressed R3 (6). Similarly, frequent co-expression of the Alp
284 protein C α and the non-Alp C β protein has been observed, where 81% of the C β positive strains
285 also contained C α (32). Even so, C α and C β are encoded by two different genes; *bca* and *bac*,
286 respectively (22, 33). To elucidate whether this was also the case for R3 and BPS, we further
287 investigated the correlation between *sar5* and R3 expression across 140 GBS strains from the
288 Norwegian GBS reference laboratory. We observed a perfect correlation between the presence
289 of the *sar5* gene and R3 expression, as well as between the absence of *sar5* and lack of R3

290 expression. Furthermore, when we induced *sar5* expression in a non-R3 bacterial strain, we
291 found that the monoclonal R3 antibody (18) recognized the *sar5*-encoded protein. During the
292 initial screening of our strain collection two strains were *sar5* positive but R3 negative (93-33
293 and 94-3). However, while these strains were deemed negative in R3 expression by fluorescent
294 antibody testing on whole bacterial cells (18), they were positive upon western blot analysis of
295 denatured cell lysate (Figure 1). We also found that GBS strains 93-33 and 94-33 possessed a
296 copy of *sar5* with a deletion (*sar5D*). We speculate that the *sar5D*-encoded protein is not
297 recognized by the monoclonal R3 antibody due to conformational changes masking the R3
298 epitope of the protein in its native form, or that the protein is simply not expressed on the
299 surface of the bacterial cells and thus only detected by immunoblotting of denatured whole cell
300 lysates. Even so, we have demonstrated that the *sar5* gene encodes R3 and that, consequently,
301 R3 and BPS must be one and the same protein.

302

303 When BPS was first identified in 2002 by Erdogan *et al*, it was considered a new protein and
304 different from R3 (13). BPS was described in the reference strain NCTC 9828 (called Compton
305 R by Erdogan *et al* (13)), which at that time was considered a Rib and R3 reference strain.
306 However, later that same year, Kong *et al*. reported that the gene thought to encode Rib in
307 strain NCTC 9828 (termed Prague 25/60 by Kong *et al*) had extensive similarities to the *rib*
308 gene but also possessed stretches which differed from *rib* (7). They named this new protein
309 Alp4, which has been the designation used since then (14). Later, it was reported that the C-
310 terminal antigenic determinant of Alp4 and Rib cross-reacted immunologically, while the N-
311 terminal antigenic determinants of Rib and Alp4 differed in immunological specificity (14).
312 The knowledge that NCTC 9828 carries *alp4* (and not *rib*) has consequences for the production
313 of specific antisera targeting R3 and BPS in the study identifying BPS as a distinct protein from
314 R3 (13). Production of specific antibodies was performed by immunizing a rabbit with strain

315 NCTC 9828. Harvested antiserum was adsorbed using a GBS strain expressing *rib*. This would
316 remove antibodies targeting Rib, and is a common procedure for generating specific polyclonal
317 antiserum. However, since NCTC 9828 does not express Rib, but the similar Alp4, antibodies
318 targeting epitopes common to both Rib and Alp4 would be removed by the adsorption whereas
319 antibodies specific only to Alp4 would remain in the antiserum. Immunoprecipitation-bands
320 that were considered evidence of R3 by Erdogan *et al.* in 2002 (13), may in fact have been
321 bands representing Alp4. Similarly, a study reporting 155 (of 4425 total) colonizing and
322 invasive GBS strains expressing BPS found no overlap between R3- and BPS-expression (34).
323 The presumed R3-specific antibody used in that study was also prepared by adsorbing antisera
324 made by immunizing a rabbit with GBS strain NCTC 9828. Regarding BPS and R3 as two
325 distinct proteins would result in an R3-designated antibody without antibodies targeting the
326 *sar5* gene product. Using NCTC 9828 as a reference strain for R3 would thus result in R3
327 antiserum that may actually detect Alp4.

328

329 As a consequence of our findings, already reported data on BPS and R3 are equally relevant
330 for the *sar5* gene product. Using BPS as the future designation of the *sar5* gene product makes
331 the historical R3 protein nonexistent, and vice versa. The nomenclature of GBS surface proteins
332 is already confusing (Table 1). BPS, R5, and now R3 are all names for the same protein. It is
333 important that communication and reports use unambiguous terminology for genes and gene
334 products. We therefore suggest using the designation R3/BPS for the *sar5*-encoded protein
335 henceforth.

336

337 Existing data suggest that the prevalence of *sar5* in GBS strains differs between geographical
338 regions. In Norway and the United States, the prevalence has been reported to lie in the range
339 2.3-8.1 % in invasive GBS strains (18, 27, 34). In a study from the United States, 3.6 % of

340 more than 4000 colonizing GBS strains carried R3/BPS (34), while in Zimbabwe it amounted
341 to near 30 % in healthy pregnant carriers (6). Thus, as a strain variable marker, the R3/BPS
342 protein has proved its potential in serotyping, as a serosubtype marker. Moreover, a
343 recombinant version of this protein has been reported as immunogenic and, on immunization,
344 induced formation of antibodies protective in an animal model, suggesting potential for this
345 protein as a vaccine component (13). R3/BPS may thus be suitable as one of the constituents
346 in a vaccine targeting GBS, particularly in vaccines aimed at populations in areas of Southern
347 Africa (6).

348

349 **Acknowledgements**

350

351 We thank the late Johan A. Mæland for the idea and encouragement for this project.

352

353 **References**

- 354 1. Russell NJ, Seale AC, O'Driscoll M, O'Sullivan C, Bianchi-Jassir F, Gonzalez-Guarin J, et al.
355 Maternal Colonization With Group B Streptococcus and Serotype Distribution Worldwide: Systematic
356 Review and Meta-analyses. *Clin Infect Dis*. 2017;65(suppl_2):S100-S111.
- 357 2. Brokaw A, Furuta A, Dacanay M, Rajagopal L, Adams Waldorf KM. Bacterial and Host
358 Determinants of Group B Streptococcal Vaginal Colonization and Ascending Infection in Pregnancy.
359 *Frontiers in cellular and infection microbiology*. 2021;11:720789.
- 360 3. Di Renzo GC, Melin P, Berardi A, Blennow M, Carbonell-Estrany X, Donzelli GP, et al.
361 Intrapartum GBS screening and antibiotic prophylaxis: a European consensus conference. *J Matern
362 Fetal Neonatal Med*. 2015;28(7):766-82.
- 363 4. Patras KA, Nizet V. Group B Streptococcal Maternal Colonization and Neonatal Disease:
364 Molecular Mechanisms and Preventative Approaches. *Front Pediatr*. 2018;6:27
- 365 5. Lin SM, Zhi Y, Ahn KB, Lim S, Seo HS. Status of group B streptococcal vaccine development.
366 *Clin Exp Vaccine Res*. 2018;7(1):76-81.
- 367 6. Mavenyengwa RT, Maeland JA, Moyo SR. Serotype markers in a *Streptococcus agalactiae*
368 strain collection from Zimbabwe. *Indian J Med Microbiol*. 2010;28(4):313-9.
- 369 7. Kong FR, Gowan S, Martin D, James G, Gilbert GL. Molecular profiles of group B streptococcal
370 surface protein antigen genes: Relationship to molecular serotypes. *J Clin Microbiol*. 2002;40(2):620-
371 6.
- 372 8. Kapatai G, Patel D, Efstratiou A, Chalker VJ. Comparison of molecular serotyping approaches
373 of *Streptococcus agalactiae* from genomic sequences. *BMC Genomics*. 2017;18(1)

- 374 9. Gabrielsen C, Maeland JA, Lyng RV, Radtke A, Afset JE. Molecular characteristics of
375 *Streptococcus agalactiae* strains deficient in alpha-like protein encoding genes. *J Med Microbiol.*
376 2017;66(1):26-33.
- 377 10. Lancefield RC, Perlmann GE. Preparation and Properties of a Protein (R-Antigen) Occurring in
378 *Streptococci* of Group-a, Type-28 and in Certain *Streptococci* of Other Serological Groups. *J Exp Med.*
379 1952;96(1):83-97.
- 380 11. Flores AE, Ferrieri P. Molecular-Species of R-Protein Antigens Produced by Clinical Isolates of
381 Group-B *Streptococci*. *J Clin Microbiol.* 1989;27(5):1050-4.
- 382 12. Wilkinson HW. Comparison of Streptococcal R Antigens. *Appl Microbiol.* 1972;24(4):669-+.
- 383 13. Erdogan S, Fagan PK, Talay SR, Rohde M, Ferrieri P, Flores AE, et al. Molecular analysis of group
384 B protective surface protein, a new cell surface protective antigen of group B streptococci. *Infect*
385 *Immun.* 2002;70(2):803-11.
- 386 14. Maeland JA, Afset JE, Lyng RV, Radtke A. Survey of immunological features of the alpha-like
387 proteins of *Streptococcus agalactiae*. *Clin Vaccine Immunol.* 2015;22(2):153-9.
- 388 15. Smith BL, Flores A, Dechaine J, Krepela J, Bergdall A, Ferrieri P. Gene encoding the group B
389 streptococcal protein R4, its presence in clinical reference laboratory isolates & R4 protein pepsin
390 sensitivity. *Indian J Med Res.* 2004;119:213-20.
- 391 16. Ferrieri P, Baker CJ, Hillier SL, Flores AE. Diversity of surface protein expression in group B
392 streptococcal colonizing & invasive isolates. *Indian J Med Res.* 2004;119 Suppl:191-6.
- 393 17. Maeland JA, Radtke A, Lyng RV, Mavenyengwa RT. Novel aspects of the Z and R3 antigens of
394 *Streptococcus agalactiae* revealed by immunological testing. *Clin Vaccine Immunol.* 2013;20(4):607-
395 12.
- 396 18. Kvam AI, Bevanger L, Maeland JA. Properties and distribution of the putative R3 protein of
397 *Streptococcus agalactiae*. *APMIS.* 1999;107(9):869-74.
- 398 19. Kvam AI, Bevanger L, Loseth K. An apparently new strain-variable *Streptococcus agalactiae*
399 protein. *Streptococci and the Host.* 1997;418:355-7.
- 400 20. Mavenyengwa RT, Maeland JA, Moyo SR. Distinctive features of surface-anchored proteins of
401 *Streptococcus agalactiae* strains from Zimbabwe revealed by PCR and dot blotting. *Clin Vaccine*
402 *Immunol.* 2008;15(9):1420-4.
- 403 21. Mavenyengwa RT, Maeland JA, Moyo SR. Putative Novel Surface-Exposed *Streptococcus*
404 *agalactiae* Protein Frequently Expressed by the Group B *Streptococcus* from Zimbabwe. *Clin Vaccine*
405 *Immunol.* 2009;16(9):1302-8.
- 406 22. Michel JL, Madoff LC, Olson K, Kling DE, Kasper DL, Ausubel FM. Large, Identical, Tandem
407 Repeating Units in the C-Protein Alpha-Antigen Gene, *bca*, of Group-B *Streptococci*. *Proc Natl Acad Sci*
408 *U S A.* 1992;89(21):10060-4.
- 409 23. National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National
410 Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2022 Jan 08].
411 Available from: <https://www.ncbi.nlm.nih.gov/>
- 412 24. Lachenauer CS, Creti R, Michel JL, Madoff LC. Mosaicism in the alpha-like protein genes of
413 group B streptococci. *Proc Natl Acad Sci U S A.* 2000;97(17):9630-5.
- 414 25. Creti R, Fabretti F, Orefici G, von Hunolstein C. Multiplex PCR assay for direct identification of
415 group B streptococcal alpha-protein-like protein genes. *J Clin Microbiol.* 2004;42(3):1326-9.
- 416 26. Radtke A, Kong FR, Bergh K, Lyng RV, Ko D, Gilbert GL. Identification of surface proteins of
417 group B streptococci: Serotyping versus genotyping. *J Microbiol Methods.* 2009;78(3):363-5.
- 418 27. Maeland JA, Radtke A. Comparison of Z and R3 antigen expression and of genes encoding
419 other antigenic markers in invasive human and bovine *Streptococcus agalactiae* strains from Norway.
420 *Vet Microbiol.* 2013;167(3-4):729-33.
- 421 28. Gabrielsen C, Maeland JA, Lyng RV, Radtke A, Afset JE. Molecular characteristics of
422 *Streptococcus agalactiae* strains deficient in alpha-like protein encoding genes. *J Med Microbiol.*
423 2017;66(1):26-33.

- 424 29. Bevanger L, Maeland JA. Type classification of group B streptococci by the fluorescent
425 antibody test. *Acta Pathol Microbiol Scand B*. 1977;85B(6):357-62.
- 426 30. Lale R, Berg L, Stuttgen F, Netzer R, Stafnes M, Brautaset T, et al. Continuous Control of the
427 Flow in Biochemical Pathways through 5' Untranslated Region Sequence Modifications in mRNA
428 Expressed from the Broad-Host-Range Promoter Pm. *Appl Environ Microbiol*. 2011;77(8):2648-55.
- 429 31. Reischl U, Pulz M, Ehret W, Wolf H. Pcr-Based Detection of Mycobacteria in Sputum Samples
430 Using a Simple and Reliable DNA Extraction Protocol. *Biotechniques*. 1994;17(5):844-5.
- 431 32. Bevanger L. Ibc Proteins as Serotype Markers of Group-B Streptococci. *Acta Path Micro Im B*.
432 1983;91(4):231-4.
- 433 33. Heden LO, Frithz E, Lindahl G. Molecular Characterization of an Iga Receptor from Group-B
434 Streptococci - Sequence of the Gene, Identification of a Proline-Rich Region with Unique Structure and
435 Isolation of N-Terminal Fragments with Iga-Binding Capacity. *Eur J Immunol*. 1991;21(6):1481-90.
- 436 34. Flores AE, Chhatwal GS, Hillier SL, Baker CJ, Ferrieri P. Expression of group B protective surface
437 protein (BPS) by invasive and colonizing isolates of group B streptococci. *Curr Microbiol*.
438 2014;69(6):894-8.

439

440 **Supporting information captions.**

441

442 **S1 Table:** Overview of the R3 typing of the 140 GBS strains examined in this study, as
443 determined by Kvam *et al.* (18), and the *sar5* typing performed in the current study
444 (corresponding to the results of S1 Figure).

445

446 **S1 Materials and Methods:** Primers used and illustration of the cloning strategies of pKT1-
447 *sar5_F* and pKT1-*sar5D_F*.

448

449 **S1 Figure:** The *sar5* typing (PCR amplification) of the 140 GBS strains examined in this study.
450 Negative (-) controls were strain 94-51 (a known R3 negative strain) or H₂O, positive (+)
451 controls were the R3 reference strain CCUG 29784. A) amplification of the *sar5* gene from the
452 7 R3 positive GBS strains. B-E) amplification of *sar5* from the 133 R3 negative strains, in
453 pools of 3-5 strains. F) amplification of *sar5* from the strains within pool 7 and pool 9.

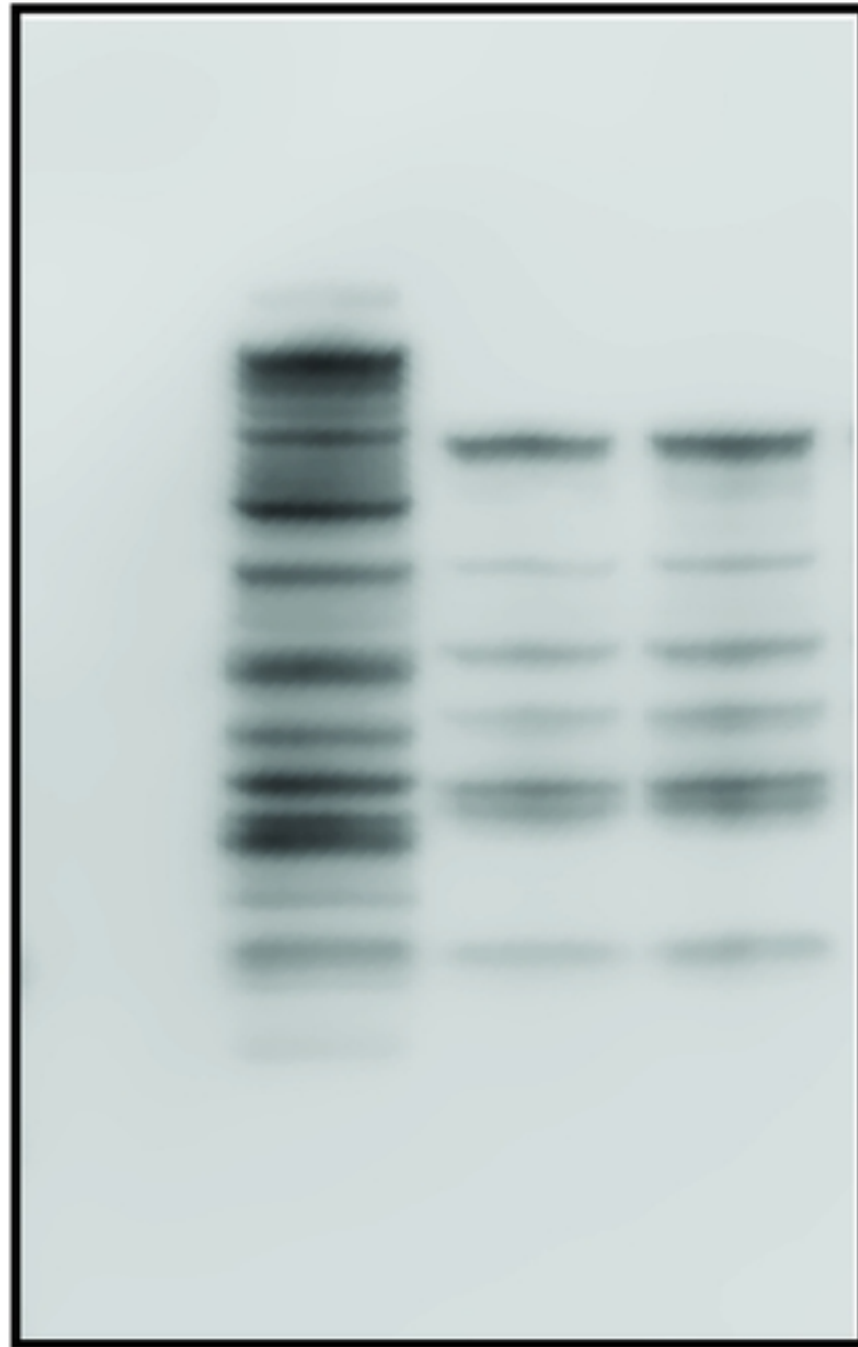
454

bioRxiv preprint doi: <https://doi.org/10.1101/2022.01.17.476664>; this version posted January 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

kDa

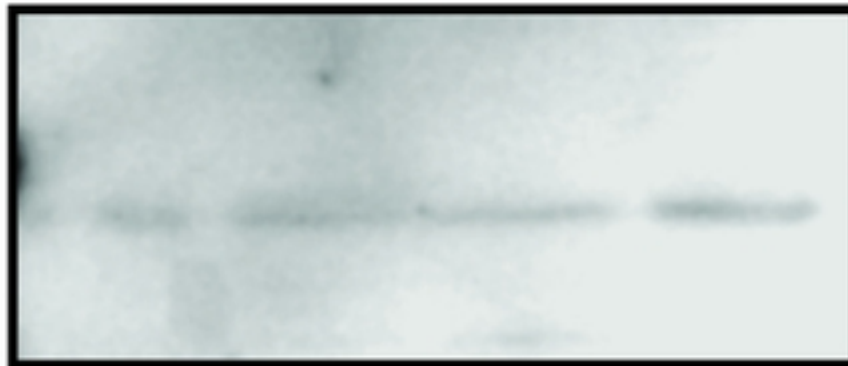
CCUG 29779
CCUG 29784
93-33
94-3

220 —
120 —
100 —
80 —
60 —
50 —
40 —
30 —



R3

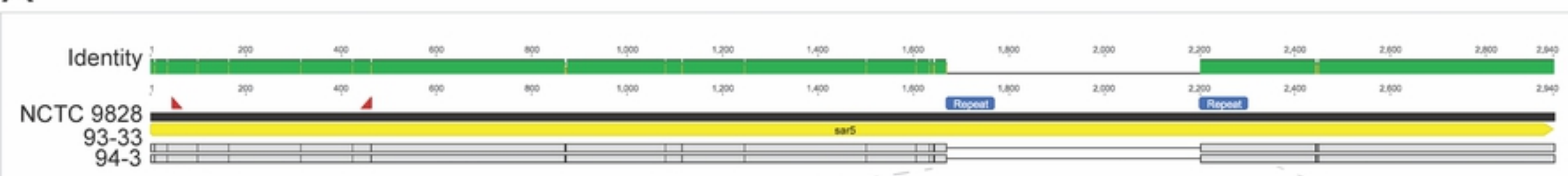
40 —



GAPDH

Figure 1

A



B

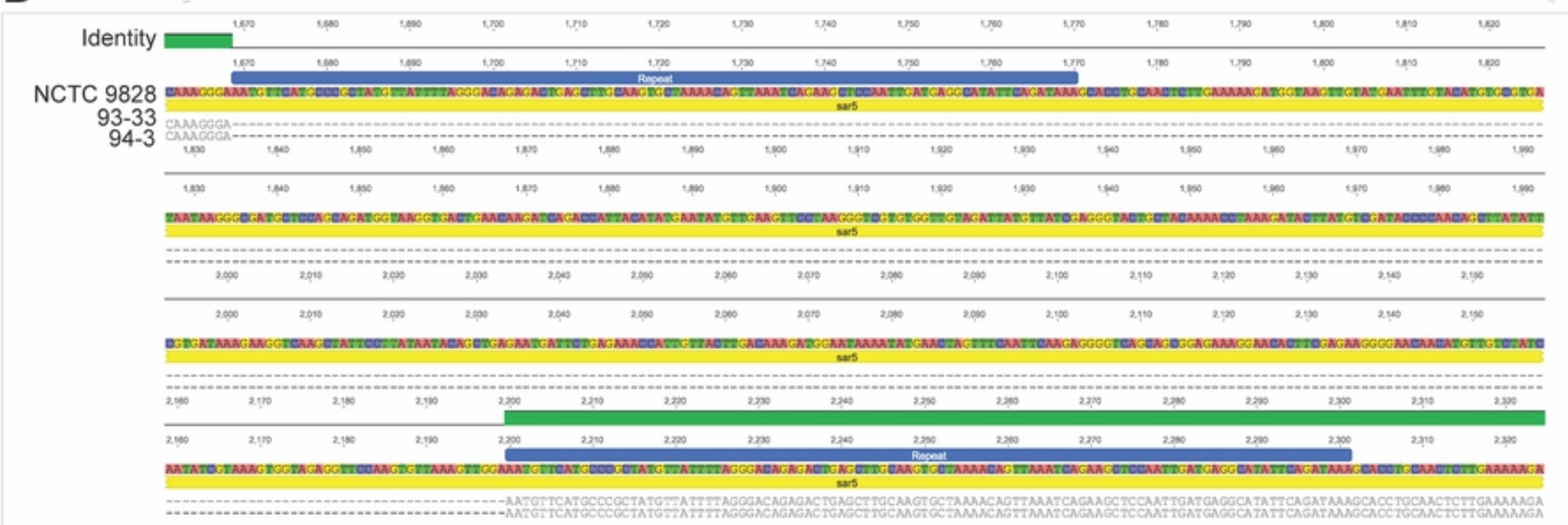


Figure 2

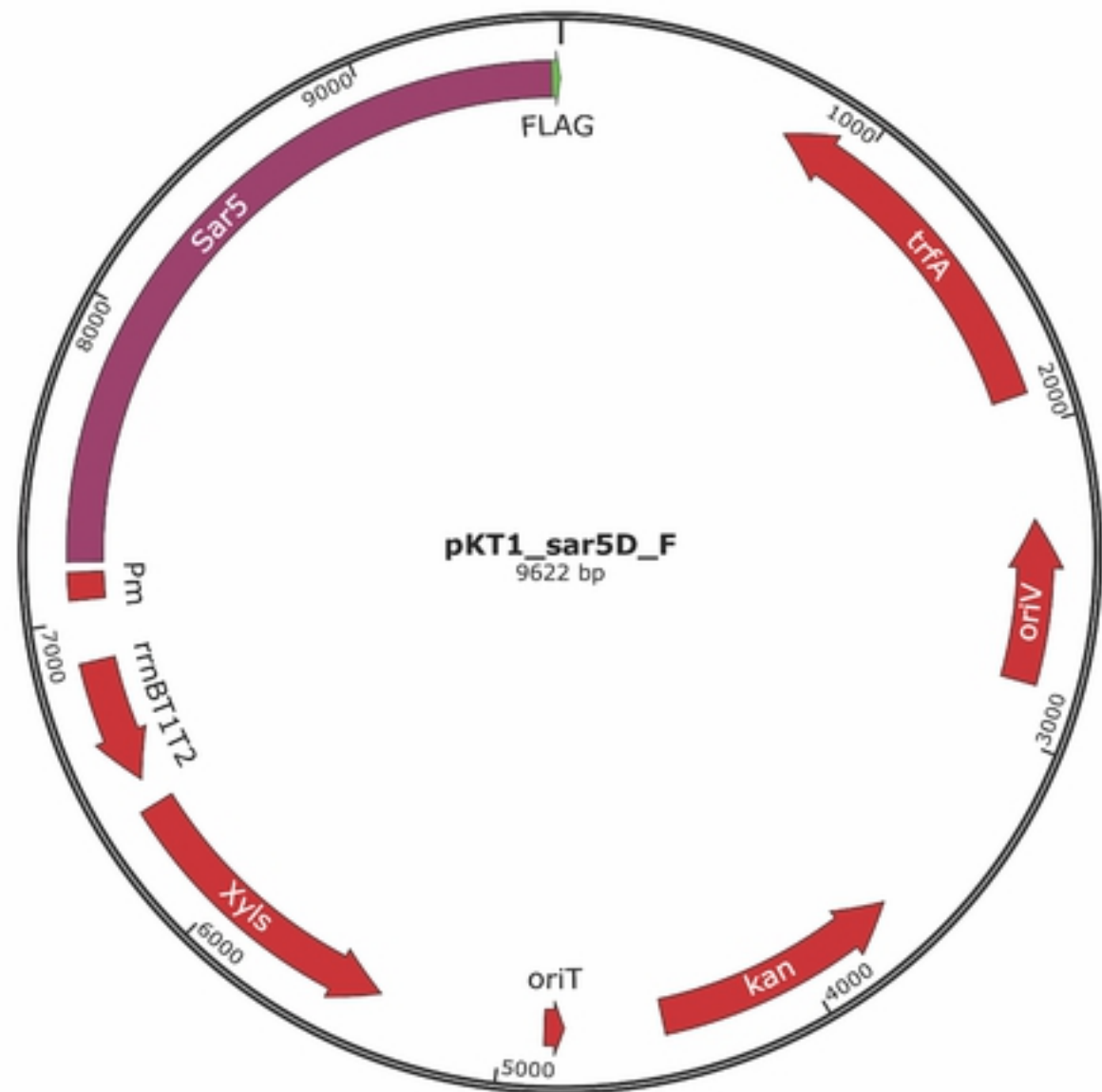
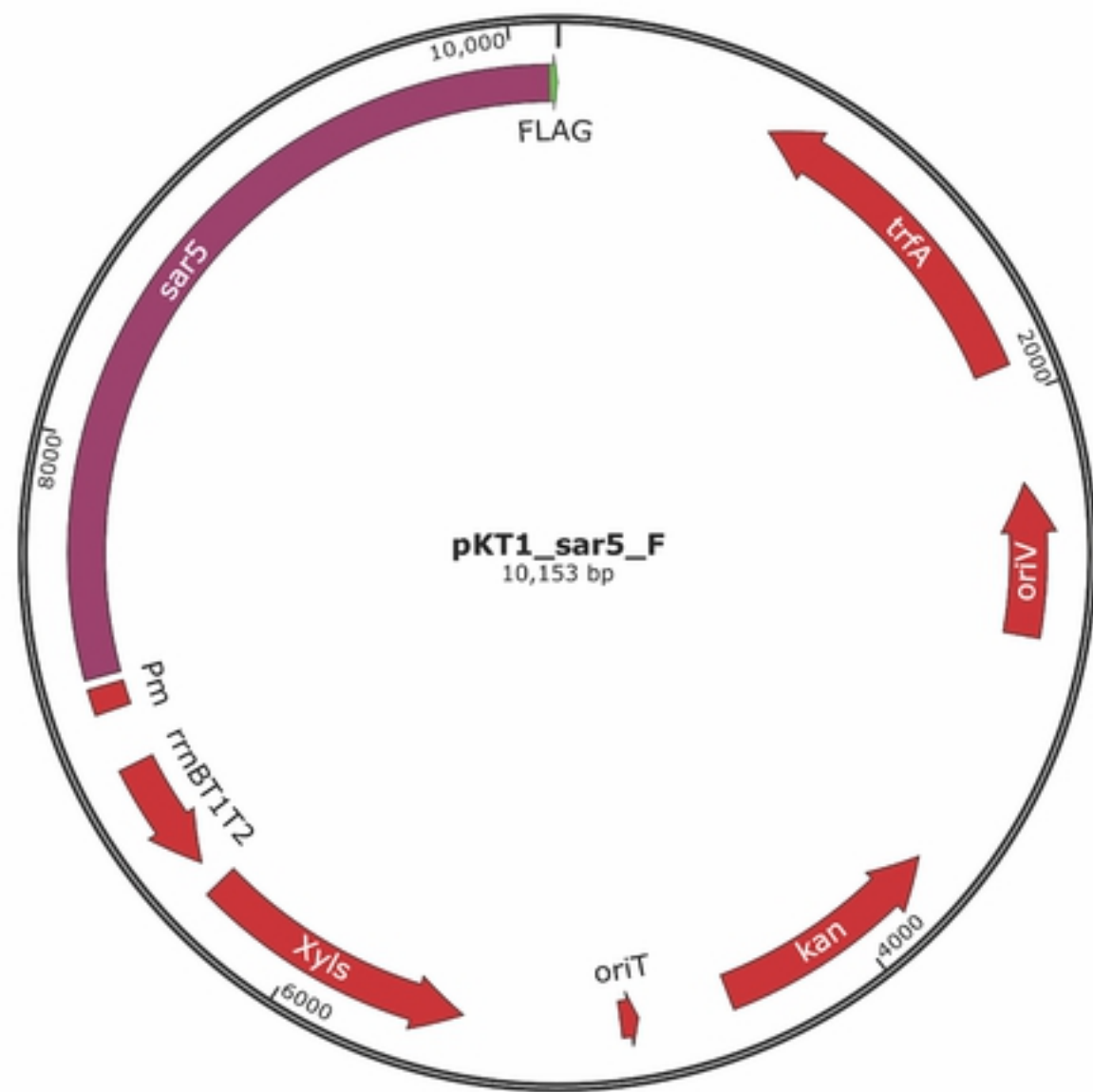


Figure 3

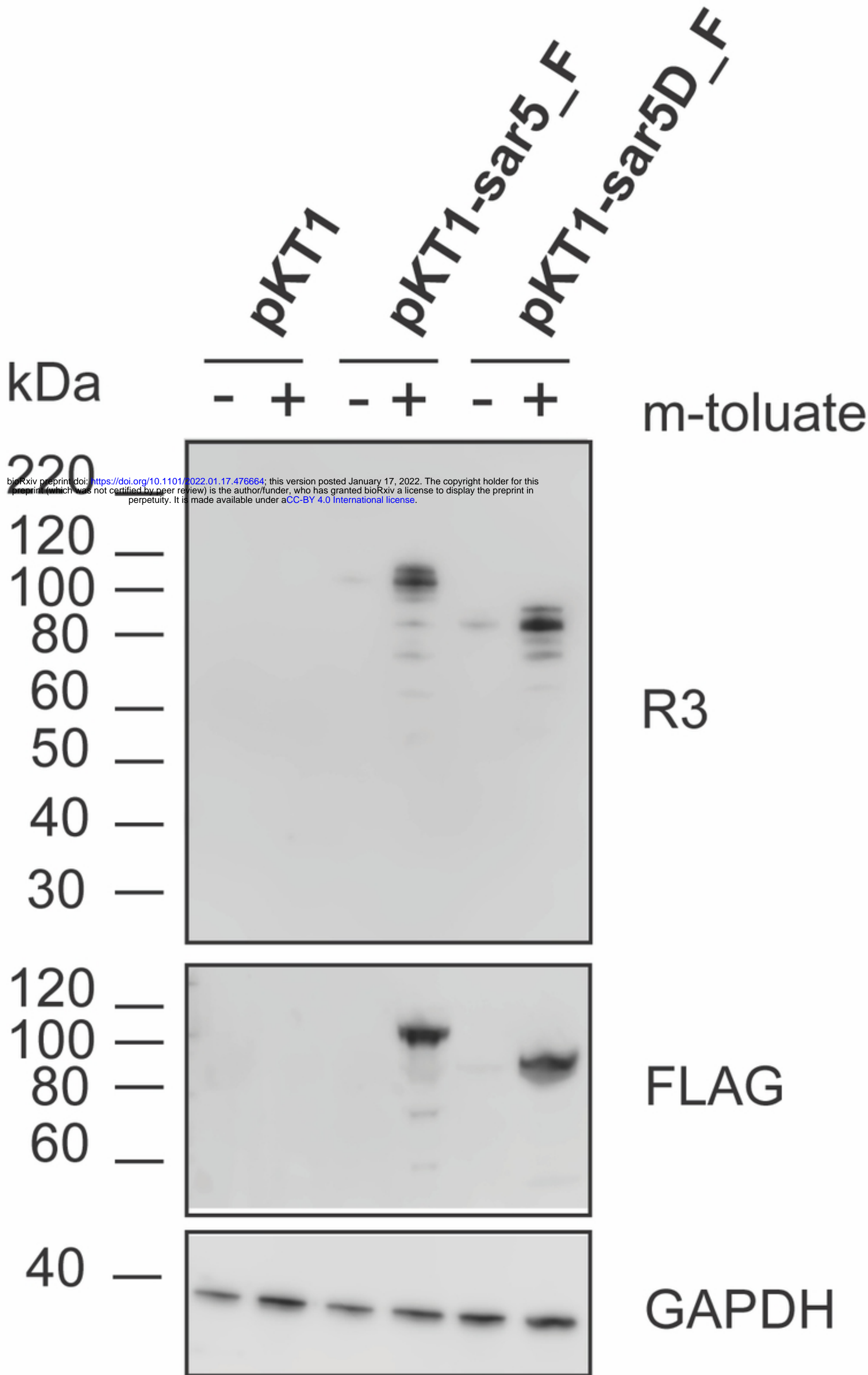


Figure 4