The binding mechanism of *Streptococcus suis* accessory virulence factor and adhesin SadP to globotetraosylceramide

Miralda Madar Johansson[1¶], Eva Bélurier[2¶#], Anastassios C. Papageorgiou[3], Anders P. Sundin[4], Jani Rahkila[5], Teemu Kallonen[6,9], Ulf J. Nilsson[4], Santeri Maatsola[2], Thomas K.M. Nyholm[7], Jarmo Käpylä[8], Jukka Corander[9], Reko Leino[10], Jukka Finne[11], Susann Teneberg[1], Sauli Haataja[2*]

[1]Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden

[2]Institute of Biomedicine, Research Center for Cancer, Infections and Immunity, University of Turku, Turku, Finland

[3]Turku Bioscience Centre, University of Turku and Åbo Akademi University, Tykistökatu 6, FI-20520 Turku, Finland

[4]Centre for Analysis and Synthesis, Department of Chemistry, Lund University, POB 124, SE-221 00 Lund, Sweden

[5]Instrument Centre, Åbo Akademi University, FI-20500 Turku, Finland

[6]Department of Biostatistics, University of Oslo, Blindern, 0317, Norway and Department of Clinical Microbiology, Turku University Hospital, Turku, Finland

[7]Biochemistry Faculty of Science and Engineering, Åbo Akademi University, FI-20500 Turku, Finland.

[8]Department of Biochemistry, University of Turku, FI-20014 Turku, Finland

[9]Parasites and Microbes, Wellcome Sanger Institute, Cambridge, United Kingdom, Department of Biostatistics, University of Oslo, Blindern, 0317, Norway and Helsinki Institute for Information Technology HIIT, Department of Mathematics and Statistics, University of Helsinki, 00014 Finland

[10]Laboratory of Molecular Science and Technology, Åbo Akademi University, FI-20500 Turku, Finland

[11]Molecular and Integrative Biosciences Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, P.O. Box 56, FI-00014, Helsinki, Finland

[#]Current address: Faculté des Sciences et Technologies (Sciences and Technology Faculty), University of Lille, Lille, France

[*]Corresponding author

E-mail: sauli.haataja@utu.fi

[¶]equal contribution

Abstract

Streptococcus suis is part of the pig commensal microbiome and a major pig pathogen causing pneumonia and meningitis, occasionally also in humans. The genomic analysis divides *S. suis* to asymptomatic carriage, respiratory and invasive systemic strains with distinct genomic signatures. The virulence factor *S. suis* adhesin P (SadP) recognizes the galabiose Galα1–4Gal-oligosaccharide. Based on the oligosaccharide fine specificity, SadP can be divided into subtypes P_N and P_O. We show that subtype P_N is distributed in the systemic strains. Both types of SadP are shown to predominantly bind to pig lung globotriaosylceramide (Gb3). SadP adhesin from virulent subtype P_N strain also binds to globotetraosylceramide (Gb4). Mutants of type P_N SadP adhesin were constructed into the galabiose-binding domain and analyzed to reveal the mechanism for Gb4 binding. Functional and structural analysis of type P_N SadP mutants showed that amino acid N285 of the galabiose-binding site was required for binding to Gb4 and strikingly was also required for interaction with the glycomimetic inhibitor phenylurea-galabiose. Molecular dynamics simulations provided further insight into the role of N285 for Gb4 and phenylurea-galabiose binding, suggesting additional hydrogen bonding to terminal GalNAc of Gb4 and urea-group. Importantly, type P_N SadP adhesin expressed by strains causing invasive systemic disease binds via a distinct amino acid N285 mediated mechanism to Gb4. Thus, this molecular mechanism of type P_N SadP binding is a candidate for selectively targeting *S. suis* meningitis without interfering with commensal strains, which opens up new venues for developing intervention strategies against this pathogen.

Author summary

Streptococcus suis is a major pig pathogen, which occasionally also causes meningitis in humans. Despite the wealth of existing genomic data, the mucosal surface colonizing mechanisms of these streptococci are poorly understood. *S. suis* population evolves rapidly and is prone to acquire antibiotic resistance factors and could hamper the efficacy of vaccines by serotype replacement. The accessory genomes of specific subpopulations are expected to encode factors that are associated with particular ecological niches and

enhanced pathogenicity. By identification of such accessory factors we could aim to develop clinical interventions targeting the pathovariants while leaving commensal bacteria unharmed, which would mitigate harmful serotype replacement. In this study we identified by both genomic and structure-function analysis the accessory *S. suis* virulence factor SadP as specific for population causing meningitis. The presented data provide an insight into a molecular mechanism how a single amino acid change can cause a functional divergence in host cell glycolipid receptor binding in emerged meningitis pathovariant of streptococci. The results can be utilized to design population- and ecology aware means to target infectious disease caused by this pathogen.

Introduction

Bacterial adhesion to host cell surfaces is a prerequisite for infectious disease. Host cell surfaces are heavily covered by surface carbohydrates that form the glycocalyx layer [1,2]. Pathogenic bacteria and their toxins can treacherously exploit carbohydrates to attach and invade cells. One example is a group of endogenous glycosphingolipids, known as globo series membrane glycosphingolipids, which are grouped based on the Galα1–4Gal (galabiose) structure. Globo series glycolipids are cellular receptors for toxic action of Verotoxins and for the attachment of bacteria, such as uropathogenic *Escherichia coli*, *Streptococcus suis* and *Pseudomonas aeruginosa* [3–6].

Since antibiotics are increasingly losing their power, alternative antimicrobial compounds need to be invented and developed [7,8]. Transformation of bacteria to antibiotic-resistant forms can be overcome by developing compounds that block *in vivo* colonization and virulence factors without killing the pathogens. Cell surface carbohydrate glycomimetics are the most tempting pipeline for the development of new generation of antimicrobials [9,10].

Streptococcus suis is an important pig pathogen causing septicemia, pneumonia and meningitis. It also causes severe zoonotic meningitis. Recent data show that it is a major pathogen of pigs and produces an alarmingly vast array of factors causing high antibiotic resistance [11]. The pathogenic types of *S. suis* are efficiently transmitted from host to host and the bacterium can be frequently isolated from the tonsils as a carrier strain. However, unfortunately predisposed by environmental factors such as overcrowding and

94 synergistic activities with other keystone pathogens [12], *S. suis* invades the host tissues and by expressing
95 both soluble and cell wall bound virulence factors aggravates tissue damage in the critical organs and causes
96 death [13,14].

97 *S. suis* strains expressing SadP adhesin bind specifically to Galα1–4Gal-containing oligosaccharides
98 [15]. Interestingly, galabiose-binding strains can be divided to subtypes P_N and P_O based on how they recognize
99 galabiose-containing carbohydrates in hemagglutination assays [16]. Type P_N strains recognize galactose, *N*-
100 Acetylgalactosamine, and the ceramide-linked oligosaccharides Gb3 (Galα1–4Galβ1–4Glc) and Gb4
101 (GalNAcβ1–3Galα1–4Galβ1–4Glc). Subtype P_O recognizes only galactose and preferentially binds to Gb3.
102 Recently, *S. suis* galabiose-binding adhesin SadP was identified as an LPNTG-anchored cell-wall protein [6].
103 The carbohydrate interaction of SadP has been shown to mediate *S. suis* binding to pharyngeal epithelium and
104 intestinal epithelial cells [16,17]. In an *S. suis* murine infection model, the mice deficient in globotriaosyl
105 ceramide (Gb3) expression (knockout in alpha-1,4-galactosyltransferase A4GALT) developed less severe
106 brain inflammation and injury [18]. Recently, low picomolar concentrations of glycomimetic 3-phenylurea-
107 galabiose-containing dendrimers were shown to block *S. suis* SadP (type P_N) adhesin binding activity [19].
108 However, the molecular basis of differences in galabiose-binding mechanisms of SadP adhesins have remained
109 elusive. In the present study, we have characterized the pig Galα1–4Gal-containing receptors in porcine lung
110 tissue using mass spectrometry and thin layer chromatography (TLC) overlay assay. The binding mechanism
111 of type P_N, P_O and site-specific mutants of type P_N SadP to globosyl oligosaccharides were studied. The results
112 show that *S. suis* SadP specifically recognizes globo series glycolipids isolated from the pig lungs and that
113 specific amino acid changes in the proximity of the galabiose-binding site have an effect on the SadP type P_N
114 and P_O specificity.

115 # Results

116 ## Sequence analysis of SadP subtypes P_N and P_O, structural modelling and distribution in *S.*
117 ## *suis* systemic, respiratory and non-clinical strains

118 SadP adhesins are 80 kDa LPNTG-anchored cell-wall proteins, which contain the signal sequence for
119 secretion, the N-terminal galabiose binding domain, tandem repeat domains and C-terminal LPNTG cell-wall

120 anchor domain [6]. The galabiose-binding domains have no significant sequence homologs to other bacterial
121 proteins or to other carbohydrate binding domains. The SadP N-terminal galabiose-binding domains (aa125-
122 328) of hemagglutinating type P_N and type P_O strains were cloned, sequenced and compared with multiple
123 alignment (Fig 1A). The structure consists of three α-helixes and ten β-strands (β1-β10) that form a β-
124 sandwich core domain. The sequences of the N-terminal galabiose-binding domain of type P_N SadP adhesins
125 are 100 % identical and are found in virulent serotype 2 strains. There is more sequence variation in the
126 galabiose-binding site of type P_O SadP. The carbohydrate binding sites of types P_N and P_O are highly conserved
127 (Fig 1B). Interestingly there is a conservative change at position 285 from asparagine to aspartate, which could
128 have an effect for the interaction of type P_N SadP with GalNAcβ1–3Galα1–4Gal-oligosaccharide.

129 The distribution of *sadP* genes in clinical and non-clinical strains analyzed with core genome phylogeny
130 analysis is shown in Figure 1C. The gene of *sadP* type P_N was found in 91.9 % of systemic *S. suis* isolates
131 (88.1 % of strains isolated from brain), whereas only less than 1 % of strains had the gene encoding *sadP* type
132 P_O. Based on the maximum-likelihood core genome phylogeny analysis the type P_N SadP gene is found in
133 clonal virulent *S. suis* strains, which mostly belong to serotype 2. Interestingly, also serotype 14 strains isolated
134 from human meningitis cases have type P_N gene. Type P_O *sadP* was found more frequently in respiratory and
135 non-clinical strains.

136

137 **Purification and analysis of glycosphingolipid composition of porcine lung**

138 The interaction of recombinant SadP with globo series glycolipids has not been thoroughly
139 characterized. Moreover, whether SadP binds to Gb3 or Gb4 isolated from the pig tissues has not yet been
140 directly shown. The glycolipids were isolated from porcine lung and erythrocytes, in order to control that the
141 glycolipids isolated from the lungs were not contaminants derived from the erythrocytes remaining in the
142 lungs. The total non-acid glycosphingolipid fractions were hydrolyzed with endoglycoceramidase II of
143 *Rhodococcus* sp., and the oligosaccharides thereby obtained were analyzed by LC-ESI/MS using graphitized
144 carbon columns. This method gives resolution of isomeric oligosaccharides, and the MS^2 analyses gives
145 complete sequence information and allows differentiation of linkage positions by diagnostic cross-ring $^{0,2}A$-
146 type fragment ions [20].

147 The base peak chromatograms from LC-ESI/MS of the oligosaccharides obtained by hydrolysis of the
148 total non-acid glycosphingolipid fractions of porcine erythrocytes and lung are shown in Figure 2A. By
149 comparison of the retention times and MS^2 spectra of oligosaccharides obtained from reference
150 glycosphingolipids, the major oligosaccharides obtained from porcine lung were tentatively identified as
151 globotriaosylceramide (Gb3, detected as a $[M-H^+]^-$ ion at *m/z* 503) and the blood group H type 2
152 pentasaccharide (H5-2, detected as a $[M-H^+]^-$ ion at *m/z* 852) (Figure 2A). The major oligosaccharides from
153 porcine erythrocytes were globotriaosylceramide (Gb3, detected as a $[M-H^+]^-$ ion at *m/z* 503),
154 globotetraosylceramide (Gb4, detected as a $[M-H^+]^-$ ion at *m/z* 706), and the blood group A type 2
155 hexasaccharide (A6-2, detected as a $[M-H^+]^-$ ion at *m/z* 1055) (Fig 2B). Thus, globotriaosylceramide was found
156 in both samples, but otherwise the two samples were significantly different. The total non-acid
157 glycosphingolipid fraction from porcine lung was thereafter separated by Iatrobeads chromatography. Aliquots
158 of the fractions obtained were analyzed by thin-layer chromatography, and fractions that were coloured green
159 by anisaldehyde were tested for binding of SadP using the chromatogram binding assay. The fractions were
160 thereafter pooled according to the mobility on thin-layer chromatograms and their SadP binding activity.
161

162 **The binding of SadP to neutral glycolipids isolated from porcine lung using TLC overlay**
163 **assay**

164 The separations gave two SadP binding fractions migrating in the triglycosylceramide region. One
165 fraction (0.2 mg) migrating as a single band was denoted fraction PL-1 (Fig 2B, lane 1), and one fraction (0.3
166 mg) migrating as a double band was denoted fraction PL-2 (Fig 2B, lane 2). In addition, one SadP binding
167 fraction (0.4 mg) with compounds migrating as tetraglycosylceramides was obtained (denoted fraction PL-4;
168 Fig 2B, lane 4).
169

170 **LC-ESI /MS of native SadP binding glycosphingolipids from porcine lung**

171 ESI/MS of the native fraction PL-1 gave a major $[M-H^+]^-$ ion at *m/z* 1132 (S1A Fig), indicating a
172 glycosphingolipid with three Hex, and sphingosine with non-hydroxy 24:1 fatty acid. MS^2 of the $[M-H^+]^-$ ion

173 at *m/z* 1132 gave a series of Y ions (Y_0 at *m/z* 646, Y_1 at *m/z* 808, and Y_2 at *m/z* 970) demonstrating a Hex-
174 Hex-Hex sequence combined with sphingosine with non-hydroxy 24:1 fatty acid (S1B Fig).

175 A $[M-H^+]^-$ ion at *m/z* 1132 was also obtained by ESI/MS of the native fraction PL-2 (S2 Fig). Here, the
176 major $[M-H^+]^-$ ion was seen at *m/z* 1022 (S2AB Fig), demonstrating a glycosphingolipid with three Hex, and
177 sphingosine with non-hydroxy 16:0 fatty acid. Also, here a Hex-Hex-Hex sequence was demonstrated by the
178 series of Y ions (Y_0 at *m/z* 536, Y_1 at *m/z* 698, and Y_2 at *m/z* 860) from MS^2 (S2D Fig).

179

180 Endoglycoceramidase digestion and LC-ESI /MS of SadP binding glycosphingolipids from
181 porcine lung

182 The base peak chromatograms from LC-ESI/MS of the oligosaccharides obtained by
183 endoglycoceramidase digestion of fractions PL-1 and PL-2 were very similar (S3AB Fig). Both had a $[M-H^+]^-$
184 ion at *m/z* 503 which eluted at the same retention time (17.4-18.4 min) as the saccharide obtained from
185 reference globotriaosylceramide (S3C Fig), while the saccharide from reference isoglobotriaosylceramide
186 eluted at 20.0-20.4 min (Figure S3D).

187 MS^2 of the ion at *m/z* 503 of fractions PL-1 and PL-2 gave in both cases two C-type fragment ions (C_1
188 at *m/z* 179 and C_2 at *m/z* 341) identifying a Hex-Hex-Hex sequence (Figure S3EF). A 4-substitution of the
189 internal Hex was demonstrated by the $^{0,2}A_2$ fragment ion at *m/z* 281 [20–22] . Taken together with the similarity
190 to the MS^2 spectrum of reference globotriaosyl saccharide (S3G Fig), this allowed identification of the
191 saccharides of fractions PL-1 and PL-2 as globotriaosyl saccharides (Galα1–4Galβ4Glc).

192 LC-ESI/MS of the oligosaccharides obtained by hydrolysis of fraction PL-4 with *Rhodococcus*
193 endoglycoceramidase II allowed a tentative identification of a globotetraosyl saccharide (GalNAcβ1–3Galα1–
194 4Galβ1–4Glc). This conclusion was based on the following spectral features: First, the base peak
195 chromatogram of fraction PL-4 had a $[M-H^+]^-$ ion at *m/z* 706 (Figure S4A), and MS^2 of this ion (S4B Fig) gave
196 a C-type fragment ion series (C_1 at *m/z* 220, C_2 at *m/z* 382, and C_3 at *m/z* 544), demonstrating a HexNAc-Hex-
197 Hex-Hex sequence. The $^{0,2}A_3$ fragment ion at *m/z* 484 demonstrated a 4-substituted Hex, while the $^{0,2}A_4$ ion at
198 *m/z* 646, and the $^{0,2}A_4$-H_2O ion at *m/z* 628, were derived from cross-ring cleavage of the 4-substituted Glc of

199 the lactose unit at the reducing end. The features of this MS^2 spectrum were very similar to the MS^2 spectrum
200 of the reference globotetraosyl saccharide (S4C Fig)

201 The base peak chromatogram of fraction PL-4 (S4A Fig) also had two $[M-H^+]^-$ ions at *m/z* 998, eluting
202 at 19.4 min and 21.4 min. In both cases MS^2 demonstrated a Fuc.Hex-(Fuc-)HexNAc-Hex-Hex sequence, and
203 the diagnostic ion at *m/z* 348 (S4D Fig) identified a Le^b hexasaccharide, whereas the diagnostic ion at *m/z* 510
204 (S4E Fig) identified a Le^y hexasaccharide [21].

205

206 **SadP subtype P_N and P_O binding to purified glycolipids**

207 TLC overlay assay is a powerful method to analyze glycolipid receptor function. The carbohydrate-
208 binding specificity of SadP to globo series isoreceptors was determined using purified glycolipids (all TLC
209 binding results are summarized in Table 1). Type P_N bound preferentially to Gb3 and P1 glycolipid, and also
210 weakly to Gb4 and the Forssman glycolipid (Fig 2C). SadP type P_O also bound strongly to Gb3 and P1, i.e. to
211 glycolipids with terminal Galα1–4Gal, whereas there was no binding to Gb4 and the Forssman glycolipid (Fig
212 2D). There was no binding to the non-acid glycolipid fraction from the pig intestine, containing the Globo H.

213 The type P_N SadP binding to dilutions of globo series glycolipids was analyzed with TLC overlay assay
214 to the detection limit for the various glycolipids (S5A Fig). The full length adhesin bound mainly to P1
215 glycolipid, with a detection limit at 0.4 μg, whereas the detection limit for the Gb3 glycolipid was 0.8 μg. The
216 detection limit of SadP(31-328) *N*-terminal domain was 0.2 μg for both Gb3 and P1.

217 Next, the SadP binding to 250 ng of glycolipids immobilized in plastic microtiter wells was
218 quantitatively evaluated by solid phase binding assay (S5B Fig). No binding to the Gb2 glycolipid,
219 galabiosylceramide was detected, whereas there was weak binding to Galα1–4Galβ1–O-*bis*-$(SO_2\text{-}C_{16}H_{33})_2$.
220 SadP was found to bind stronger to Gb3 with non-hydroxy ceramide than to Gb3 with hydroxy ceramide.
221 There was also binding to Gb4 and iGb3 isoglobotriaosylceramide. Binding to P1 pentaosylceramide was at
222 the same level as binding to Gb3 with non-hydroxy ceramide. The difference in the binding to iGb3 (Galα1–
223 3Gal- containing glycolipid) in the microtiter well assay compared to TLC overlay assay could be due to the
224 difference in the presentation of the binding epitope and the clustering of the glycolipid moieties.

225

226 **Galα1–4Gal-dependent endothelial cell binding activity of *S. suis* SadP**

227 *S. suis* type P_N WT strain D282 and the corresponding mutant strain D282Δ*sadP,* containing an insertion
228 mutation into SadP gene, were compared for their binding activity to cultured cell line EA.hy926, which is a
229 primary human umbilical vein cell line fused with a thioguanine-resistant clone of A549 (human lung
230 adenocarcinoma cell line). In addition, binding of purified fluorescently labelled SadP was analysed for the
231 binding to EA.hy926 cells. The cells were grown into glass coverslips for 48 h and the bacteria grown into
232 early log phase were added to cells and were let to adhere for 1 h. After washing and staining with DiffQuick
233 the bacteria were enumerated. The average adherence of wild-type bacteria was calculated to be 8.7 ± 0.5
234 bacteria/field, whereas the mean adhesion of 1.3 ± 0.5 of the insertion mutant Δ*sadP* was significantly less
235 (unpaired t test > 0.0001) (Fig 3A). In addition, the binding of the WT strain was inhibited with 10 μg/ml
236 pigeon ovomucoid, which contains *N*-linked glycans with terminal Galα1–4Gal [23]. The inhibitor reduced
237 the adherence of *S. suis* bacteria to the cells to 0.7 ± 0.1 bacteria/field.

238 For flow cytometry assay, the cells were grown in wells, washed and tested for the binding of FITC
239 labelled SadP. After washing, the cells were harvested with scraping and analyzed with flow cytometer. The
240 flow cytometry results (dot blot and the histogram) are presented in Figure 3BCD. The median SadP binding
241 with and without inhibitor (pigeon ovomucoid, 10 μg/ml) were 56.2 and 14.9 per 10 000 total events (median
242 of untreated cells was 12.8).

243

244 **Fine specificities toward Gb3 and Gb4 oligosaccharide structures**

245 Next, the ability of the type P_N SadP-D282(31-328) and P_O SadP-6407(31-328) recombinant adhesin
246 constructs to recognize galabiose glycoconjugates was analyzed using isothermal titration calorimetry (Fig
247 4AB). Both type P_N and P_O adhesins were titrated with oligosaccharides representing the terminal epitopes of
248 Gb3 and Gb4, the TMSEt (2-trimethylsilylethyl) glycosides of Galα1–4Gal and GalNAcβ1–3Galα1–4Gal
249 [24]. The dissociation constant values (K_D) of type P_N SadP-D282(31-328) and type P_O SadP-6407(31-328)
250 interaction with Galα1–4Gal were 13.0 ± 1.5 μM (n=0.85, ΔH=-69.0±2.0 and -TΔS=41.1) and 3.5 ± 1.6 μM
251 (n=0.85, ΔH=-40.0±2.7 and -TΔS=8.8) respectively. The K_D for interaction of P_N and P_O adhesins with

252 GalNAcβ1–3Galα1–4Gal compared to Galα1–4Gal were 2.6-fold higher with type P_N (K_D= 33.7±3.5 μM,
253 n=1.14, ΔH=-40.1±2.0 and -TΔS=14.0) and 269-higher with type P_O (K_D= 940±680 μM, n=0.13, ΔH=-
254 197±5500 and -TΔS=179).

255 Then, both adhesins were analyzed with AlphaScreen competitive inhibition assay (Fig 4C) for their
256 interaction with Galα1–4Gal in comparison to a glycomimetic inhibitor 3'-phenylurea-derivative (3'-
257 phenylurea-Galα1–4Gal β1–methoxyphenyl), which is a low nanomolar inhibitor of type P_N SadP [19,25].
258 The binding of His-tagged recombinant adhesins to biotinylated pigeon ovomucoid was inhibited with the
259 above oligosaccharides. The IC_{50} values of type P_N adhesin for Galα1–4Galβ1–methoxyphenyl and 3'-
260 phenylurea-Galα1–4Gal β1–methoxyphenyl were 1 μM and 0.030 μM, whereas with type P_O adhesin the
261 corresponding values were 0.3 μM and 70 μM respectively. Also 4 other recombinant type P_O SadP N-
262 terminal-domains were tested using 10 μM inhibitor concentrations (Fig 4D). For all type P_O adhesins the
263 inhibition obtained with the C3'-phenylurea-derivative was weaker than the inhibition obtained with Galα1–
264 4Gal.

265

266 **Site-specific mutation changes the specificity of type P_N recombinant adhesin to type P_O**

267 The binding mechanism of type P_N and P_O SadP was further studied by site-directed mutagenesis. Type
268 P_N adhesin SadP-D282(125-329) was mutated and the mutants were analyzed for the binding to galabiose
269 oligosaccharides with AlphaScreen competitive inhibition assay and ITC. Multiple alignment of SadP
270 homologues allowed us to predict amino acids from the galabiose-binding region that differed between type
271 P_N and P_O (Fig 1A). Based on that, the site-specific amino acid change N285D and E292Q, deletion Δ244-246
272 and N285D&Δ244-246 mutants were constructed into the type P_N SadP(125-328) cloned into the plasmid
273 pET28a. In addition, the 28 amino acid long region H216-D246 of the type P_N was replaced by the
274 corresponding region of type P_O.

275 The WT SadP and mutants were expressed and purified as described in the methods and analysed with
276 SDS-PAGE (S6 Fig). The galabiose-binding strengths of the above mutant constructs, full length SadP, SadP
277 *N*-terminal constructs SadP(31-328), SadP(125-328) and as a negative control site-specific mutant SadP(31-

278 328)W258A, were compared by AlphaScreen asssay (S7 Fig). The binding of full-length SadP to biotinylated
279 pigeon ovomucoid was weaker compared to the *N*-terminal domain constructs, which could be due to a larger
280 size of the full length SadP. The size of the protein could increase the distance of the acceptor and donor beads,
281 and hence reduce the transmission of the singlet oxygen during the assay [19]. As a negative control for the
282 assay, there was no binding of site-specific mutant W258A to pigeon ovomucoid. Of the mutants in the type
283 P_N SadP(125-328) background, mutants N285D and E292Q bound well whereas the binding of mutants Δ244-
284 246 and H216-D246delinsDELFNRFPNKVDSTNNGDGAPFRFFNKE was reduced. The binding of
285 N285D&Δ244-246 mutant was abolished.

286 Next the specificities of mutants were analyzed with AlphaScreen assay using phenylurea-derivative
287 and Galα1–4Gal (S7B - F Figs). All mutants were equally inhibited by the Galα1–4Gal oligosaccharide, but
288 only with the site-specific mutant N285D the inhibitory power of Phenylurea-Galα1–4Gal was weaker
289 compared to Galα1–4Gal, suggesting a specific switch from type P_N to P_O specificity.

290 Next, the Alphascreen inhibition assay was used to compare the effect of site-specific mutation in
291 N285D to the specificity between Galα1–4Gal (terminal Gb3 structure) and GalNAcβ1–3Galα1–4Gal
292 (terminal Gb4 structure) TMSEt glycosides. SadP(125-329) type P_N and site-specific mutant N285D were
293 similarly inhibited with Galα1–4Gal, whereas N285D was not inhibited with even higher concentrations of
294 GalNAcβ1–3Galα1–4Gal (Fig 5AB).

295 The K_D values of type P_N SadP(125-328) and site-specific mutant N285D for oligosaccharides was
296 analyzed with isothermal titration calorimetry (Fig 5C). The wild type SadP(125-329) K_D values for Galα1–
297 4Gal and GalNAcβ1–3Galα1–4Gal were 3.4±0.09 (n=0.76, ΔH=92.0±0.3 and -TΔS=60.8) and 36.0±4.8 μM
298 (n=0.56, ΔH=70.2±5.0 and -TΔS=44.7) respectively. The K_D of site specific mutant N285D with Galα1–4Gal
299 was 22.7±7.2 μM (n=0.52, ΔH=26.8±5.4 and -TΔS=20.5) (6-fold increase compared to wild-type SadP),
300 whereas there was no interaction between N285D and GalNAcβ1–3Galα1–4Gal. Thus, the above results
301 suggest that N285 has a specific role in SadP type P_N binding to GalNAcβ1–3Galα1–4Gal.

302 The effect of N285D mutation to SadP binding to glycolipids incorporated into POPC and POPC-
303 cholesterol liposomes was analyzed by dotting liposomes to PVDF membrane and binding of his-tagged
304 adhesins to immobilized liposomes (Fig 6). WT SadP type P_N bound to Gb4 of both non-cholesterol and

305 cholesterol-containing liposomes, whereas there was no binding by N285D mutant. The acyl chain length or
306 hydroxylation of ceramide had no major effects for Gb3 binding.
307

308 **Structural comparison of recombinant SadP type P_N and its N285D mutant**

309 The site-specific mutant SadP(125-329)N285D was crystallized with five molecules in the
310 crystallographic asymmetric unit (Fig 7A, Table S1). All molecules are similar with subtle differences as
311 suggested by the low root mean square deviation (r.m.s.d) between them (0.25-0.31 Å). The structure of the
312 N285D mutant is similar to that of the native SadP (PDB id 5BOB) with r.m.s.d. of 0.26 Å for 199 aligned
313 residues. The structure consists of three α-helixes and ten β-strands (β1-β10) that form a β-sandwich core
314 domain (Fig 7B). The first β-sheet of the β-sandwich is formed by antiparallel β1-β10-β9-β4-β7-β6 strands
315 and the second β-sheet by β2-β3-β8-β5 in antiparallel fashion.
316 Structural superposition with the Galα1–4Gal – Fhc structure revealed small differences (Fig 7C). The
317 side chain of D285 was found slightly rotated compared to the N285 side chain in the wild type.
318

319 **Resolution of the interaction mechanisms of SadP to globo series glycolipids and C3'-**
320 **Phenylurea–Galα1–4Gal**

321 The key amino acids of both types P_N and P_O are conserved. Y198, Q255, R238 and G233 forms
322 hydrogen bonds with the HO–3', HO-4' and HO-6' of the α'-Gal and the K316 with HO-2 and HO-3 of the β-
323 Gal pyranose ring. W258 most likely forms CH-π interactions with the hydrophobic face of the disaccharide
324 (Fig 7C). Type P_N SadP N285 is on the edge of the binding pocket, and it potentially interacts with the terminal
325 GalNAcβ1–3 of Gb4 or 3-phenylurea-group. To study the SadP interaction with the receptor saccharide
326 Galα1–4Galβ1–4Glc in solution, STD-NMR was applied using type P_N SadP(125-329) and the corresponding
327 mutant N285D (Fig 7D). STD-NMR results suggest that both the type P_N SadP and the mutant N285D interact
328 mainly with the two terminal galactose units of Gb3. With the mutant N285D the interactions of H-4′ and H-
329 4″ are, however, somewhat weaker compared to the interaction of H-6″. Additionally, the mutant N285D does

330 not seem to have a clear interaction with the β-glucoside unit at the reducing end of Gb3, whereas the type P_N
331 SadP appears to have interactions with H-2, H-3 and H-4.

332 In order to find a plausible hypothesis explaining the binding preferences of the galabiose, GalNAcβ1–
333 3Galα1–4Gal and phenylurea-galabioside for WT and the N285D mutant, molecular dynamic simulations of
334 their respective complexes were performed. Briefly, the ligand structures were built without the TMSEt
335 aglycon to simplify the simulations and were then placed in the binding site of SadP (pdb id 5BOA) with the
336 Galα1–4Gal disaccharide oriented as in the crystal structure. The terminal GalNAcβ1 residue was oriented
337 with the dihedral glycosidic bond angles minimized with the H1 parallel to H4' of Galα1–4Gal. The three
338 complexes with WT were mutated by exchanging N285 for D285. The complexes were unrestrainedly
339 subjected to 100 ns molecular dynamics simulations and all starting conformations converged towards similar
340 stable complex geometries (Fig 8A-E), except for the phenylurea derivative in complex with the N285D mutant
341 for which light constraints on the backbone atoms of protein secondary structures were applied (Fig 8F). The
342 α-face of the β-galactoside residue of all three ligands stacked to W258 forming CH-π interactions and the
343 galabiose disaccharide moiety in all simulations retained the interactions and hydrogen bonds observed in the
344 X-ray structure with the Gb3 trisaccharide (PDB id 5boa). The HO-4' of galabiose did not interact directly
345 with N285, but instead via N285 sidechain NH hydrogen bond donation to a network of hydrogen bonds
346 involving 3-5 water molecules and the G233 carbonyl oxygen (Fig 8A and S9 movie). Mutation of N285 to
347 D285 influenced the directionalities of the hydrogen bonding partners in this network (Fig 8D and S10 movie),
348 which may explain the different STD-NMR data with WT and the D285 mutant. The simulations with the
349 GalNAcβ1–3Galα1–4Gal and the phenylurea-galabioside in complex with the WT protein converged to
350 geometries that revealed a highly populated interaction between the ligands and N285. The ring oxygen of the
351 GalNAc residue of the trisaccharide and the urea carbonyl oxygen of the phenylurea derivative accepted
352 hydrogen bonds from the N285 side chain NH_2 (Fig 8B-C and S11 - S12 movies). The GalNAc HO-4 of the
353 trisaccharide formed transient hydrogen bonds to side chain amide groups of E286 and E234, while the phenyl
354 ring of the phenylurea ligand did not form any persistent interactions with any protein residue. The highly
355 populated hydrogen bonds of the trisaccharide GalNAc ring oxygen and the phenylurea carbonyl oxygen to
356 N285 could not form in the N285D mutant (Fig E-F and S13-14 movies). Instead, it was disrupted by the

357 formation of a water network between the D285 side chain carboxylate and the GalNAc ring oxygen of the
358 trisaccharide (Fig 8E), while the phenylurea moiety of the phenylurea galabioside rotated away from the D285
359 side chain carboxylate (Fig 8F). This lack of direct hydrogen bond interactions of the GalNAc ring oxygen
360 and urea carbonyl oxygen may at least partly explain the less efficient binding of the trisaccharide and phenyl
361 urea derivative to the N285D mutant.
362

363 **Discussion**

364 SadP carbohydrate specificity to neutral glycolipids from lungs was studied by TLC overlay assay. The
365 carbohydrate receptors for *S. suis* adhesion in respiratory tract and lungs have remained elusive. We studied
366 here how SadP type P_N and P_O recognize glycolipids isolated from pig lung. The major pig lung neutral
367 glycolipids were identified as Gb3, Gb4, the H type 2 pentaosylceramide, the Galili pentaosylceramide and
368 Le^b and Le^y hexaosylceramides (Fig 2). In TLC overlay assay, type P_N SadP bound to Gb3 and Gb4, whereas
369 type P_O SadP bound only to Gb3. Both SadP types showed no binding to Galα1–3Gal-containing Galili
370 pentaosylceramide. (Fig 2AB, Table 1). Therefore, our findings suggest that globo series glycolipids are the
371 major carbohydrate receptors for *S. suis* in pig lung. Although fucose-binding lectins have been described both
372 in bacterial and fungal lung pathogens, no fucose-specific binding activity was found in SadP. The globo series
373 glycolipids have been shown to be widely expressed in many tissues of pigs [26]; hence, they can potentially
374 serve as SadP receptors for other host organs and brain capillary veins critical for *S. suis* invasive infection.
375 We show that the binding of *S. suis* serotype 2 *sadP*-mutant to endothelial cell line was significantly reduced
376 and binding of both the WT *S. suis* bacteria and recombinant SadP adhesin were inhibited with the galabiose-
377 containing pigeon ovomucoid.

378 The oligosaccharide inhibition assays performed with both whole bacteria and recombinant adhesin
379 suggest that the reducing end β-Glc of Gb3 is required for optimal binding [16,19]. The TLC overlay assay
380 using purified glycolipids showed that SadP bound strongest to the P1 glycolipid (Fig 2BC and S5A Fig). The
381 preferential binding to the P1 glycolipid might also be due to a more optimal presentation of the epitope on
382 the longer core chain, whereas hindered presentation of binding epitope in galabiosyl ceramide might explain,
383 why SadP did not bind to Gb2 in TLC overlay assay. The presence of P1 structure is not well known in pig,

384 however it was found in α1,3GalT knockout pig kidney, since this glycosyltransferase is a direct competitor
385 of α1,4GalT [27]. This might be important for surveillance of pathogens in xenotransplantation.

386 The fine specificity of SadP subtypes was further analyzed with ITC and competitive inhibition assay.
387 The ratio of K_D values (Galα1–4Gal / GalNAcβ1–3Galα1–4Gal) for type P_N SadP was 0.4 whereas for type
388 P_O it was 0.004, which is in accordance to the corresponding hemagglutination inhibition MIC values observed
389 with bacterial hemagglutination assays [16]. Moreover, the IC_{50} values for Galα1–4Gal and 3'-phenylurea-
390 Galα1–4Gal are in accordance to previous hemagglutination inhibition MIC concentrations for type P_N and P_O
391 *S. suis* [25].

392 The mechanism of how the two SadP subtypes differ in their recognition of Gb3, Gb4 and glycomimetic
393 3-phenylurea-galabiose has not been studied before. The multiple-sequence alignment of SadP showed that
394 specific amino acids G233, E249 and N285 of type P_N in galabiose-binding domain are different compared to
395 type P_O SadP (Fig 1A). Of these, only substitution from asparagine to aspartate at position 285 is found in all
396 hemagglutinating type P_O adhesins. To analyze the mechanisms how SadP type P_N can bind to Gb4 terminal
397 trisaccharide (GalNAcβ1–3Galα1–4Gal) and 3'-phenylurea-galabiose, we designed mutations to type P_N SadP
398 and analyzed their binding to galabiose. The SadP type P_N region from H216 to D246 was swapped to the
399 corresponding region of type P_O (to contain substitution from amino acid G to N at position 233). The binding
400 of this mutant to galabiose was reduced compared to the wild type adhesin (S6A Fig), but was still sufficient
401 to further analyze the oligosaccharide specificity. Deletion Δ244-246 and site-specific mutations N285D and
402 E292Q bound also to galabiose at sufficient level (S6A Fig) allowing testing with 3'-phenylurea-derivative to
403 compare for type-specificity. Of these mutants, only N285D lost the ability to be efficiently inhibited by 3'-
404 phenylurea-derivative (S6B-F Fig), suggesting that N285 has an important role in interacting with the
405 phenylurea-group. The amino acid region 216-246 including G233N substitution have possibly no significant
406 role alone in the interaction with HO-3' substituted phenylurea-galabiose. However, G233 could still
407 synergistically interact with phenylurea. Therefore, the N285D mutant was chosen for further studies for the
408 type P_N binding mechanism. The WT and N285D mutant were compared for their binding to the soluble Gb4
409 terminal trisaccharide GalNAcβ1–3Galα1–4Gal and to globo series glycolipids Gb3 and Gb4 (Figs 5 and 6).

410 The results show consistently that N285 plays a specific role in Gb4 binding as demonstrated by competitive
411 inhibition assay, ITC and liposome binding assay.

412 Molecular dynamics calculations were performed with the structures representing type P_N SadP(125-
413 329) and its N285D site-specific mutant. Both proteins were analyzed with Galα1–4Gal, GalNAcβ1–3Galα1–
414 4Gal and 3'-phenylurea-Galα1–4Gal. In accordance to the co-crystal structure (PDB id. 5boa), the galabiose
415 HO-4', HO-6', HO-2 and HO-3 were hydrogen bonded to Y198, G233, R238, Q255 and K316. Hydrophobic
416 interaction with the α-face of the β-galactoside stacked to W258 forming CH-π interactions. The above
417 interactions were conserved in WT and N285 mutant. WT SadP N285 formed hydrogen bonds to GalNAc O-
418 5 ring oxygen of GalNAcβ1–3Galα1–4Gal and to urea's carbonyl group (Fig 8B). The amino acid D285 of
419 site-pecific mutant formed hydrogen bonds with solvent water molecules when simulated with GalNAcβ1–
420 3Galα1–4Gal and 3'-phenylurea-Galα1–4Gal. These results together with the AlphaScreen and ITC data give
421 us a plausible explanation for the molecular mechanism of how type P_N SadP interact with Gb4 terminal
422 GalNAc and with phenylurea-galabiose derivative. In addition, WT SadP amino acid G233 interacts with the
423 terminal GalNAc of Gb4 via hydrogen bond to N-Acetyl-group or to urea's -NH group (Fig 8C).

424 SadP does not have any sequence homologs in other galabiose-recognizing proteins, however
425 structurally *E. coli* PapG and SadP galabiose-binding sites share some similarities. They both contain a β-sheet
426 composed of antiparallel β-strands, which are adjacent to the galabiose-binding domain (PDB; *E. coli* 1j8r, *S.*
427 *suis* 5boa). In SadP, the binding site is formed by the alpha-helix and a non-helix loop, whereas in PapG the
428 galabiose binding site is located in the middle of two non-helix loops. Other galabiose binding proteins such
429 as *P. aeruginosa* LecA (4yw6), *E. coli* verotoxin (1bos) and *Lyophyllum decastes* fungal lectin (4ndv) show
430 different combining site architectures. As previously suggested, the galabiose-binding adhesin SadP from
431 gram-positive and *E. coli* P fimbrial adhesins from gram-negative bacteria are an example of convergent
432 evolution towards binding to the galabiose oligosaccharide. The structural similarities thus confirm this
433 hypothesis.

434 Hemagglutination positive strains expressing type P_N SadP from pig and zoonotic meningitis strains
435 belong to serotype 2 and clonal group CC1 [6,17]. *S. suis* hemagglutination type P_O strains were mostly found
436 from serotype 4 and were isolated from lungs of strains causing pneumonia [28]. The majority of strains that

437 have the gene encoding type P_N SadP belongs to a distinct population consisting of systemic isolates, which
438 were identified in a large BRaTP1T consortium functional genomic study [29]. *S. suis* strains that cause human
439 meningitis comprise a separate clade and are thought to have evolved when the pig farming was intensified in
440 1920s [29]. Zoonotic *S. suis* strains as a specific population have been thought to exponentially spread to
441 different geographical areas due to the selection of pigs that are optimal in terms of productivity. The BRaTP1T
442 consortium study of *S. suis* genomic signatures for pig and human infections has suggested that the systemic
443 infections are genetically determined and the genome size of highly virulent strains is reduced and enriched
444 with virulence genes. Since the type P_N SadP adhesin gene has not been lost in strains causing systemic
445 infections, it is tempting to speculate that the recognition of Gb4 may play a specific role in invasive infections.
446 In a previous study by Ferrando et al. (2017), type P_N SadP (designated SadP1) present in CC1 clonal group
447 of *S. suis* strains [17] was found to specifically adhere better to human intestinal cells (Caco-2) than to pig
448 intestinal cells. This could indicate that type P_N SadP has specific binding preferences to human intestine, thus
449 promoting zoonosis.

450 The results of the present study could have implications toward the therapeutic applications targeting
451 SadP and *S. suis* systemic infections (summarized in Fig 9). The gene encoding SadP type P_N has a downstream
452 homolog, which is a pseudogene (has a stop codon). It could be speculated, that the gene encoding SadP type
453 P_N is a putatively duplicated gene, which has evolved to bind Gb4 (Figure 9A). This study shows the molecular
454 mechanism how N285 contributes to the binding of terminal GalNAcβ1- saccharide of Gb4 by a specific
455 interaction with GalNAc hexose ring oxygen (Figure 9BC). The analogous N285-mediated binding
456 mechanisms to Gb4 and glycomimetic phenylurea-galabiose could be utilized in the design of therapeutics
457 against the invasive systemic infections without disturbing the commensal *S. suis* bacteria colonizing pig upper
458 respiratory tract.

459

460 # Materials and Methods

461 ## Bacterial strains

462 The chromosomal DNA was isolated from type P_N and P_O *S. suis* strains as described before [6] using
463 GelElute bacterial genomic DNA kit according to the manufacturer's instructions. *Streptococci* were grown in
464 Todd Hewitt broth supplemented with 0.5 % yeast extract or in Columbia agar plates supplemented with 5 %
465 sheep blood in 5 % CO_2 at 37°C and *E. coli* strain were grown in LB medium. Antibiotics used were 30 µg/ml
466 (pET28) and 100 µg/ml ampicillin pET46EkLIC) in *E. coli* and 500 µg/ml kanamycin for *S. suis* D282Δ*SadP*
467 mutant [6].
468 *Escherichia coli* NovaBlue (Novagen) was used for the cloning of SadP constructs and strain
469 BL21(DE3) for expression of the adhesin.
470

471 ## Cloning and construction of recombinant adhesins

472 The galabiose binding *N*-terminal domains of SadP from hemagglutination subtypes P_N and P_O were
473 cloned into pET46EkLIC vector as follows: The primer pairs SadPDER48
474 (gacgacgacaagatagaatcgctagaaccagatgtt) and SadPDER49 (gaggagaagcccggtttattcttctcaagggtaatctc) were
475 designed to clone the 882 bp fragment of the adhesin *N*-terminal galabiose binding domain. The fragments
476 were amplified with Phusion HotStart II DNA polymerase and were cloned into LIC-vector (LIC, ligation
477 independent cloning) pET46EkLIC (Novagen). The ligation products were transformed into NovaBlue
478 competent cells. This construct yielded a 33.4 kDa 6xHis-tagged fusion protein (SadP(31-328)). The sequence
479 of *N*-terminal domains of SadP homologs were verified by sequencing with T7 promoter and T7 terminator
480 primers. The vectors were transformed into expression strain BL21(DE3).
481 Site-specific mutant SadP(31-328) W258A was constructed by PCR with 5'-phosphorylated forward
482 primer tgg aac gca tct gct ggt caa gct and reverse primer ttg aga att aaa acc att ttc. The PCR product was
483 amplified with Phusion HotStart II DNA polymerase and the PCR fragment was ligated with T4 ligase
484 (Promega) and transformed into Novablue competent cells. The resulting plasmid SadP(31-328)W258A-
485 pET28a was sequenced to verify the mutation.

486 Synthetic gene construct and the corresponding site-specific mutants for WT SadP(125-328) type P_N
487 were obtained from Genscript. Site-specific mutants constructed were Δ244-246 (PSAD-1), N285D (PSAD-
488 2), E292Q (PSAD-3), N285D&Δ244-246 (PSAD-4) and H216-
489 D246delinsDELFNRFPNKVDSTNNGDGAPFRFFNKE (PSAD-5) (replacement of aa 216-242 into
490 corresponding sequence of type P_O of strain 6107).
491

492 **Bioinformatics**

493 To compare the distribution of SadP alleles in *S. suis* from different infections e.g. invasive, respiratory
494 and non-clinical samples, the presence of either P_N or P_O allele was detected from 374 previously published *S.*
495 *suis* isolates [29] using SRST2 software version 0.2.0 with default settings [30]. The SadP genes from *S. suis*
496 P1/7 (AM946016.1) and 6407 (NZ_CP008921.1) were used as reference sequences for P_N and P_O, respectively.
497 The distribution of P_N and P_O alleles in the *S. suis* phylogeny was investigated by creating a core-genome
498 maximum-likelihood tree using Roary version 3.12.0 [31] and RAxML version 8.2.8 [32] from the 374
499 isolates.
500

501 **Expression and purification of SadP proteins**

502 The bacteria were grown at 30°C, 250 rpm, to an OD600 of 0.5, and the protein expression was induced
503 with 0.2 mM IPTG for 3.5 hours. Bacteria were harvested by centrifugation with 3000 x g, at +4°C and stored
504 at -84°C. The recombinant protein was purified with Ni-NTA affinity chromatography. Briefly, bacteria were
505 lysed with 0.4 mg/ml hen egg lysozyme (Sigma) in 50 mM sodium phosphate buffer pH 8.0, containing 0.5 M
506 NaCl, EDTA-free protease inhibitor cocktail (Pierce), 20 mM imidazole, 20 µg/ml deoxyribonuclease and 1
507 mM $MgCl_2$ on ice for 30 mins. The lysate was sonicated to further homogenize the cell debris and was
508 centrifuged at 20000 x g, +4°C for 30 min. The filtered lysate was purified with Ni-NTA affinity
509 chromatography using HiPrep FF column connected to Äktaprime plus, GE Healthcare at +25°C. Further
510 purification was done with a gel filtration HiLoad 16/60 Superdex 200 column using Tris-HCl, pH 7.5, 0.15
511 M NaCl as running buffer. The purity of the recombinant proteins was analyzed with SDS-PAGE

512 chromatography. For native gel electrophoresis, the recombinant proteins were diluted to sample buffer
513 without SDS and were separated with 8 % polyacrylamide gel.
514

515 **Crystallization**

516 SadP N285D mutant (15 mg/ml) was crystallized with the hanging-drop vapor-diffusion method. The
517 well condition contained 1.3-1.5 M sodium citrate tribasic, 0.1 M sodium cacodylate, pH 6.5. Equal volumes
518 of well solution and protein (2 μL+2μL) were mixed and the crystals grew at 16 °C, initially as haystacks of
519 very thin needles. Single crystals suitable for data collection were subsequently grown by using the seeding
520 technique. A small part of the haystack was crushed with a metallic needle in the crystallization drop and 0.2
521 μL were diluted in 20 μL of mother liquor to create a seeding stock. Pre-equilibrated (1-d old) crystallization
522 drops were seeded with 0.1 μL of the seeding stock. Single crystals started to grow after ~3 days.
523

524 **Data collection and crystal structure determination**

525 Diffraction data were collected on the BioMAX beamline at MAX IV (Lund, Sweden) from a single
526 crystal under cryogenic temperatures (100K). The crystal diffracted to 1.85 Å but the resolution was later
527 truncated during processing to 2.05 Å owing to completeness. The crystal was found to belong to the *C*2 space
528 group. Data processing was carried out with automated procedures in EDNA [33] and scaling was done with
529 AIMLESS [34].
530 Molecular replacement with PHASER [35] was carried out to obtain initial phases using the native
531 structure as a template. Five molecules in the crystallographic asymmetric unit were located (Matthews
532 coefficient 2.67 $Å^3$/Da corresponding to ~53.9% solvent content). Initial building of the structure was carried
533 out using the automated building procedure in BUCCANEER [36] using the ccp4i2 interface [37]. After the
534 initial building that produced an almost complete structure, the refinement continued in PHENIX v.1.17.1-
535 3660 [38] using simulated annealing at 1000K with maximum likelihood as target function. The first round of
536 simulated annealing resulted in R_{work}/R_{free} of 0.227/0.269. Inclusion of waters and rounds of manual rebuilding
537 using COOT [39] alternated with PHENIX refinement resulted in the final structure with R_{work}/R_{free} of

538 0.174/0.221 (Table S1). A glycerol molecule was found bound in each of the A, C, and D subunits. Two
539 glycerol molecules were located in the binding site of subunit B. The structure has been deposited to the Protein
540 Data Bank under the accession code 6YRO
541

542 **Isolation of SadP binding glycosphingolipids from porcine lung**

543 Non-acid glycosphingolipids were isolated from porcine lung as described [40]. Briefly, the lung tissue
544 was lyophilized, and then extracted in two steps in a Soxhlet apparatus with chloroform and methanol (2:1 and
545 1:9, by volume, respectively). The material obtained was subjected to mild alkaline hydrolysis and dialysis,
546 followed by separation on a silicic acid column. Acid and non-acid glycosphingolipid fractions were obtained
547 by chromatography on a DEAE-cellulose column. In order to separate the non-acid glycolipids from alkali-
548 stable phospholipids, this fraction was acetylated and separated on a second silicic acid column, followed by
549 deacetylation and dialysis. Final purifications were done by chromatographies on DEAE-cellulose and silicic
550 acid columns.

551 The total non-acid glycosphingolipid fraction (9 mg) from porcine lung was first separated on an
552 Iatrobeads (latrobeads 6RS-8060; Iatron Laboratories, Tokyo) column (1.0 g) and eluted with increasing
553 volumes of methanol in chloroform. Aliquots of the fractions obtained were analyzed by thin-layer
554 chromatography. Fractions that were coloured green by anisaldehyde were tested for binding of SadP using
555 the chromatogram binding assay. SadP bound to a fraction containing glycosphingolipids migrating in the tri-
556 and tetraglycosylceramide regions on thin-layer chromatograms. This fraction (1.0 mg) was further separated
557 on an Iatrobeads column (1.0 g), eluted with chloroform/methanol/water 65:25:4 (by volume), 30 x 0.5 ml,
558 followed by chloroform/methanol/water 65:25:4, 10 mi. The SadP binding triglycosylceramide was eluted in
559 fractions 12-17, and these fractions were pooled into one fraction migrating as a single band (0.2 mg, denoted
560 fraction PL-1), and fraction migrating as a double band (0.3 mg, denoted fraction PL-2) on thin-layer
561 chromatograms. In addition, a SadP tetraglycosylceramide was present in fractions 19-31, and pooling of
562 these fractions gave 0.4 mg (denoted fraction PL-4).
563

564 **Reference glycosphingolipids**

565 Total acid and non-acid glycosphingolipid fractions were isolated as described [40], and the individual
566 glycosphingolipids were obtained by repeated chromatography on silicic acid columns, and by HPLC, and
567 identified by mass spectrometry [20,41] and 1H NMR spectroscopy [42].
568

569 **Thin-layer chromatography**

570 Aluminum- or glass-backed silica gel 60 high performance thin-layer chromatography plates
571 (Merck, Darmstadt, Germany) were used for thin-layer chromatography, and chromatographed with
572 chloroform/methanol/water (60:35:8 by volume) as solvent system. The different glycosphingolipids
573 were applied to the plates in quantities of 4 μg of pure glycosphingolipids, and 20-40 μg of
574 glycosphingolipid mixtures. Chemical detection was done with anisaldehyde [43].
575

576 **Radiolabeling**

577 Aliquots of 100 μg of the different SadP protein preparations were labeled with ^{125}I by the Iodogen
578 method according to the manufacturer's instructions (Pierce/Thermo Scientific), giving approximately 2000
579 cpm/μg protein.
580

581 **Chromatogram binding assays**

582 Binding of radiolabeled proteins to glycosphingolipids on thin-layer chromatograms was done as
583 described [44]. Chromatograms with separated glycosphingolipids were dipped for 1 min in diethylether/n-
584 hexane (1:5, by volume) containing 0.5% (w/v) polyisobutylmethacrylate (Aldrich Chern. Comp. Inc.,
585 Milwaukee, WI). After drying, the chromatograms were soaked in PBS containing 2% (w/v) bovine serum
586 albumin, 0.1% (w/v) NaN_3 and 0.1% (w/v) Tween 20 (Solution A), for 2 h at room temperature. Thereafter
587 the plates were incubated with ^{125}I- labeled SadP protein (l -5 x 10^6 cpm/ml) for 2 h at room temperature. After

588 washing six times with PBS, and drying, the thin-layer plates were autoradiographed for 12 h using XAR-5
589 x-ray films (Eastman Kodak, Rochester, NY).
590

591 **Microtiter well assay**

592 Binding of radiolabeled SadP to glycosphingolipids in microtiter wells was performed as described [44].
593 In short, 250 µg of pure glycosphingolipids in methanol were applied to microtiter wells (Falcon 3911, Becton
594 Dickinson Labware, Oxnard, CA). When the solvent had evaporated, the wells were blocked for 2 h at room
595 temperature with 200 µl of BSA/PBS. Thereafter, the wells were incubated for 4 h at room temperature with
596 50 µl of ^{125}I-labeled SadP (2 x 10^3 cpm/µl) diluted in BSA/PBS. After washing 6 times with PBS, the wells
597 were cut out and the radioactivity was counted in a gamma counter.
598

599 **LC-ESI/MS of native glycosphingolipids**

600 Native glycosphingolipids were analyzed by LC-ESI/MS as described [45]. Glycosphingolipids were
601 dissolved in methanol:acetonitrile in proportion 75:25 (by volume) and separated on a 200x0.150 mm column,
602 packed in-house with 5 µM polyamine II particles (YMC Europe GmbH, Dinslaken, Germany). An
603 autosampler, HTC-PAL (CTC Analytics AG, Zwingen, Switzerland) equipped with a cheminert valve (0.25
604 mm bore) and a 2 µl loop, was used for sample injection. An Agilent 1100 binary pump (Agilent technologies,
605 Palo Alto, CA) delivered a flow of 250 µl/min, which was split down in an 1/16" microvolume-T (0.15 mm
606 bore) (Vici AG International, Schenkon, Switzerland) by a 50 cm x 50 µm i.d. fused silica capillary before the
607 injector of the autosampler, allowing approximately 2-3 µl/min through the column. Samples were eluted with
608 an aqueous gradient (A:100% acetonitrile to B: 10 mM ammonium bicarbonate). The gradient (0-50% B) was
609 eluted for 40 min, followed by a wash step with 100% B, and equilibration of the column for 20 min. The
610 samples were analyzed in negative ion mode on a LTQ linear quadropole ion trap mass spectrometer (Thermo
611 Electron, San José, CA), with an IonMax standard ESI source equipped with a stainless steel needle kept at –
612 3.5 kV. Compressed air was used as nebulizer gas. The heated capillary was kept at 270 °C, and the capillary
613 voltage was –50 kV. Full scan (*m/z* 500-1800, two microscans, maximum 100 ms, target value of 30,000) was

614 performed, followed by data-dependent MS^2 scans (two microscans, maximun 100 ms, target value of 30.000)
615 with normalized collision energy of 35%, isolation window of 2.5 units, activation q= 0.25 and activation time
616 30 ms). The threshold for MS^2 was set to 500 counts. Data acquisition and processing were conducted with
617 Xcalibur software version 2.0.7 (Thermo Fisher Scientific). Manual assignment of glycosphingolipid
618 sequences was done with the assistance of the Glycoworkbench tool (Version 2.1), and by comparison of
619 retention times and MS^2 spectra of reference glycosphingolipids.
620

621 **Endoglycoceramidase digestion and LC-ESI/MS**

622 Endoglycoceramidase II from *Rhodococcus* spp. (Ito) (Takara Bio Europe S.A., Gennevilliers, France)
623 was used for hydrolysis of glycosphingolipids. Briefly, 50 µg of glycosphingolipids were resuspended in 100
624 µl 0.05 M sodium acetate buffer, pH 5.0, containing 120 µg sodium cholate, and sonicated briefly. Thereafter,
625 1 mU of enzyme was added, and the mixture was incubated at 37°C for 48 h. The reaction was stopped by
626 addition of chloroform/methanol/water to the final proportions 8:4:3 (by volume). The oligosaccharide-
627 containing upper phase thus obtained was separated from detergent on a Sep-Pak QMA cartridge (Waters).
628 The eluant containing the oligosaccharides was dried under nitrogen and under vacuum.
629 The glycosphingolipid-derived oligosaccharides were resuspended in 50 µl of water and analyzed by
630 LC-ESI/MS as described [20]. The oligosaccharides were separated on a column (200 x 0.180 mm) packed in-
631 house with 5 µm porous graphite particles (Hypercarb, Thermo-Hypersil, Runcorn, UK). An autosampler,
632 HTC-PAL (CTC Analytics AG, Zwingen, Switzerland) equipped with a cheminert valve (0.25 mm bore) and
633 a 2 µl loop, was used for sample injection. An Agilent 1100 binary pump (Agilent technologies, Palo Alto,
634 CA) delivered a flow of 250 µl/min, which was split down in an 1/16" microvolume-T (0.15 mm bore) (Vici
635 AG International, Schenkon, Switzerland) by a 50 cm x 50 µm i.d. fused silica capillary before the injector of
636 the autosampler, allowing approximately 2-3 µl/min through the column. The oligosaccharides (3 µl) were
637 injected on to the column and eluted with an acetonitrile gradient (A: 10 mM ammonium bicarbonate; B: 10
638 mM ammonium bicarbonate in 80% acetonitrile). The gradient (0-45% B) was eluted for 46 min, followed by

639 a wash step with 100% B, and equilibration of the column for 24 min. A 30 cm x 50 μm i.d. fused silica
640 capillary was used as transfer line to the ion source.

641 The oligosaccharides were analyzed in negative ion mode on an LTQ linear quadrupole ion trap mass
642 spectrometer (Thermo Electron, San José, CA). The IonMax standard ESI source on the LTQ mass
643 spectrometer was equipped with a stainless-steel needle kept at –3.5 kV. Compressed air was used as nebulizer
644 gas. The heated capillary was kept at 270 °C, and the capillary voltage was –50 kV. Full-scan (*m/z* 380-2 000,
645 2 microscans, maximum 100 ms, target value of 30 000) was performed, followed by data dependent MS^2
646 scans of the three most abundant ions in each scan (2 microscans, maximum 100 ms, target value of 10 000).
647 The threshold for MS^2 was set to 500 counts. Normalized collision energy was 35%, and an isolation window
648 of 3 u, an activation q = 0.25, and an activation time of 30 ms, was used. Data acquisition and processing were
649 conducted with Xcalibur software (Version 2.0.7).

650 Manual assignment of glycan sequences was done on the basis of knowledge of mammalian biosynthetic
651 pathways, with the assistance of the Glycoworkbench tool (Version 2.1), and by comparison of retention times
652 and MS^2 spectra of oligosaccharides from reference glycosphingolipids [20].

653

654 **S. suis cell binding assay**

655 EA.hy926 cells were maintained and grown in DMEM, 10 % FCS medium at 37°C, 5 % CO_2. For
656 bacterial binding assay the cells were detached with Trypsin-EDTA, counted and 15 000 cells/well added into
657 round glass coverslips in 24 well-plates. Cells were grown for 24 - 48 h to subconfluency.

658 *S. suis* D282 WT strain and D282-Δ*sadP* were grown in THY overnight at 37°C, 5 % CO_2. Bacteria
659 were diluted 1/20 into prewarmed 37°C THY and were grown to OD550 of 0.2 and were diluted 1/100 into
660 prewarmed DMEM, 10 % FCS without antibiotics. 500 μl of bacterial dilution was pipetted into the wells
661 (MOI 100:1) and the plate was centrifuged with 800 x g, 15 min at 20°C.

662 The plate was incubated at 37°C for 1 h. The wells were washed 4 x 1000 μl of PBS. The cells were
663 stained with DiffQuick kit, washed and the coverslips were mounted on glass slides with Permount.

664 The bacterial binding was quantitated by microscopy with 100 x objective with immersion oil. The
665 bound bacteria were enumerated by counting two wells for each sample. The results were expressed as average
666 of bacteria / optical field.

667 For flow cytometry, N-terminal domain of SadP was labelled with FITC (Sigma-Aldrich) in 0.2 M
668 borate buffer, pH 8.0, 0.16 M NaCl (50:1 molar ratio of FITC/SadP, 83 nmol of SadP and 1.7 µmol of FITC)
669 and the labelling was stopped with Tris-buffer. The labelled SadP was purified from the free FITC with PD-
670 10 (GE Healthcare) desalting column. The EA.hy926 cells were seeded into 6 well culture dishes (1 x 10^5
671 cells/well) and were grown for 48 h. The cells were washed with DMEM and were incubated with 400 ng/ml
672 of labelled SadP with or without 10 µg/ml pigeon ovomucoid. After binding, the wells were washed with
673 phosphate Buffered Saline (PBS, 0.15 M NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 1.5 mM KH_2PO_4). and were
674 fixed with 2 % (w/v) paraformaldehyde in PBS for 15 min. The cells were washed twice with PBS and the
675 cells were scraped of the wells and suspended into the PBS. 10 000 cells were analyzed for SadP binding with
676 FACSCalibur.

677

678 **Amplified luminescent proximity homogeneous assay (AlphaScreen)**

679 The AlphaScreen assay was optimized as described before [19]. Briefly, the assay was performed using
680 AlphaScreen Streptavidin Donor beads and NiNTA Acceptor beads (PerkinElmer). The molar concentrations
681 of galabiose-containing biotinylated ovomucoid (receptor) and adhesin were optimized by finding the hook
682 point for the interaction by setting up a matrix of proteins in 96 well AlphaPlates (Perkin Elmer) in 10 mM
683 Tris-HCl, pH 7.5, 0.15 M NaCl, 0.2% BSA, 0.05 % Tween 20 0.2% (TBST-0.2%BSA). For the optimized
684 assay dilutions of 5 µl of the His-tagged adhesin and 5 µl of biotinylated ovomucoid were pipetted into the
685 wells and the plates were centrifuged for 1 min, 1000 x g. The plates were incubated at +25°C for 1 hr. This
686 was followed by addition of Ni-NTA- beads (20 µg/ml) and incubation for 1.5 hrs. Finally, streptavidin donor
687 beads were added (20 µg/ml) and the mixtures were incubated for 30 mins and were then measured with the
688 Ensight multimode reader (Perkin Elmer) using excitation wavelength of 680 nm with donor beads and
689 measurement of emission wavelength of 615 nm from AlphaLISA anti-HIS acceptor beads. For inhibition
690 assays the oligosaccharides, whose structures have been described before [25] or depicted in Figure 4 for

691 TMSEt- glycosides of GalNAcβ1–3Galα1–4Gal and 3Galα1–4Gal, were diluted into TBST-0.2%BSA buffer
692 and were mixed with the His-tagged adhesin and biotinylated ovomucoid. The binding was measured as
693 described above. The binding inhibition data was fitted using Prism with settings of log(inhibitor) vs. response
694 slope (four parameters).
695

696 **Isothermal titration calorimetry**

697 SadP-D282(31-328) (P_N), SadP-6107(31-328) (P_O), SadP(125-329) type P_N and site-specific mutant
698 SadP(125-328)N285D were desalted into PBS using PD-10 desalting column (GE Healtcare). The synthetic
699 oligosaccharides were diluted to the same buffer. 0.1 - 0.2 µM of SadP proteins in a 350 µl vessel was titrated
700 with 1.5 - 2 mM solutions of galabiose oligosaccharides using MicroCal (Malvern). The protein solution was
701 stirred with 750 rpm at 25°C and 16 injections of a volume of 2.49 µl were injected at 180 s intervals. The data
702 from single determinations was analyzed with Origin.
703

704 **Preparation of liposomes and SadP binding assay**

705 Vesicles were prepared as described before [46]. Briefly, POPC (1-palmitoyl-2-oleoyl-glycero-3-
706 phosphocholine, Avanti Polar lipids), cholesterol and glycolipids were dissolved and mixed into the
707 chloroform : methanol (2:1, vol/vol) to contain 30 % cholesterol and 2 mM glycolipids. The solvent was
708 evaporated and the lipid film was suspended for 30 min at 60 °C in 10 mM Tris-Cl, 140 mM NaCl buffer. The
709 lipid-buffer suspension was briefly vortexed followed by the extrusion procedure (Avanti mini extruder using
710 0.1 µm polycarbonate membranes filter, (Avanti Polar Lipids, Alabaster, AL, USA)) to form large unilamellar
711 vesicles.
712 PVDF membrane was wetted (wetting in 100 % MeOH, washing 4 x 5 min MilliQ water) and was
713 overlaid onto wet Whatman filter paper. Liposomes were serially diluted 1/10 into Tris-buffered saline and 2
714 µl of liposomes containing 2 mM, 0.2 mM and 0.02 mM of glycolipid was pipetted onto the membrane and
715 the membrane was not allowed to dry during the pipetting. The membrane was saturated with 2 % BSA (w/v)
716 in TBS for 1 h at 20°C and was incubated with 10 µg/ml recombinant SadP for 2 h at 20°C. The membrane

717 was washed 4 x 5 min with TBS and the bound SadP was detected with 1:10000 dilution of anti-His primary
718 antibody (Sigma) and 1:20000 dilution of HRP-labelled secondary rabbit anti-mouse antibody
719 (DakoCytomation). The membrane was incubated with ECL substrate (WesternBright Quantum, Labtech) and
720 the SadP binding was imaged with Fuji LAS-4000.
721

722 **STD-NMR**

723 STD-NMR measurements were carried out on a Bruker AVANCE III HD spectrometer (Bruker BioSpin
724 GmbH, Rheinstetten, Germany) operating at 600.13 MHz, equipped with a nitrogen-cooled triple-channel TCI
725 inverse CryoProbe™. The pulse sequence used for the experiments was the stddiffesgp provided by Bruker.
726 A 1x PBS buffer with a pH of 7.45 with 10% D_2O, containing 50 µM protein and 2500 µM Gb3 trisaccharide
727 was used for the experiments. The on-resonance pulse was set to the methyl groups at 0.9 ppm and the off-
728 resonance pulse was set at 20 ppm. A saturation time of 2 s was used and the experiments were carried out at
729 313 K to somewhat improve the weak signals, likely resulting from slow k_{off}. A blank experiment with only
730 Gb3 and no protein was done to rule out any self-STD.
731

732 **Molecular dynamics simulations**

733 Molecular dynamics simulations were performed with the OPLS3 force field in Desmond (Schrödinger
734 Release 2019-4: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2017;
735 Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2017) using default settings except
736 for the length of the simulation. Simulations were performed with periodic boundary conditions using
737 orthorhombic simulation boxes with SPC water model. Counter ions were used to neutralize each other.
738 Starting conformations of ligands in complex with SadP (pdb id 5BOA) were built by manually placing ligand
739 galabiose residues in an identical position to that in the crystal structure (pdb id 5BOA). Complexes with the
740 mutants were generated by exchanging N285 for D285. The complexes were then subjected to 100 ns
741 molecular dynamics simulations. The complex of the phenylurea derivative with the N285D mutant was not

742 stable during the 100 ns simulations. Therefore, light constraints (1 kcal·mol^{-1}·Å^{-1}) on the protein secondary
743 structures were applied. Molecular images were generated using PyMOL v2.30 (Schrodinger LLC).
744

745 ## Acknowledgements

746 The synthetic glycolipid Galα1–4Galβ1–O-bis-(SO_2-$C_{16}H_{33}$)$_2$ was a kind gift from Professor Göran
747 Magnusson, Lund University, Sweden. ACP thanks Noushin Madani for help during crystallographic data
748 collection. Infrastructure support from Biocenter Finland is acknowledged.
749

750 ## References

751

752 1. Pizarro-Cerda J, Cossart P. Bacterial adhesion and entry into host cells. Cell. 2006;124: 715–727.
753 doi:10.1016/j.cell.2006.02.012
754 2. Weinbaum S, Tarbell JM, Damiano ER. The Structure and Function of the Endothelial Glycocalyx
755 Layer. Annu Rev Biomed Eng. 2007;9: 121–67. doi:10.1146/annurev.bioeng.9.060906.151959
756 3. Lingwood CA, Binnington B, Manis A, Branch DR. Globotriaosyl ceramide receptor function - where
757 membrane structure and pathology intersect. FEBS Lett. 2010;584: 1879–1886.
758 doi:10.1016/j.febslet.2009.11.089
759 4. Stromberg N, Marklund BI, Lund B, Ilver D, Hamers A, Gaastra W, et al. Host-specificity of
760 uropathogenic *Escherichia coli* depends on differences in binding specificity to Gal alpha 1-4Gal-
761 containing isoreceptors. EMBO J. 1990;9: 2001–2010.
762 5. Chemani C, Imberty A, de Bentzmann S, Pierre M, Wimmerova M, Guery BP, et al. Role of LecA and
763 LecB lectins in *Pseudomonas aeruginosa*-induced lung injury and effect of carbohydrate ligands. Infect
764 Immun. 2009;77: 2065–2075. doi:10.1128/IAI.01204-08
765 6. Kouki A, Haataja S, Loimaranta V, Pulliainen ATAT, Nilsson UJUJJ, Finne J. Identification of a novel
766 streptococcal adhesin P (SadP) protein recognizing galactosyl-α1-4-galactose-containing

767 glycoconjugates: Convergent evolution of bacterial pathogens to binding of the same host receptor. J
768 Biol Chem. 2011;286: 38854–38864. doi:10.1074/jbc.M111.260992
769 7. Rello J, Parisella FR, Perez A. Alternatives to antibiotics in an era of difficult-to-treat resistance: new
770 insights. Expert Review of Clinical Pharmacology. 2019. pp. 12(7):635-642.
771 doi:10.1080/17512433.2019.1619454
772 8. Alves M, Ferreira IFR, Dias J, Teixeira V, Martins A, Pintado M. A review on antimicrobial activity
773 of mushroom (basidiomycetes) extracts and isolated compounds. Planta Medica. 2012. pp.
774 78(16):1707-1718. doi:10.1055/s-0032-1315370
775 9. Bernardi A, Jimenez-Barbero J, Casnati A, Castro C De, Darbre T, Fieschi F, et al. Multivalent
776 glycoconjugates as anti-pathogenic agents. Chem Soc Rev. 2013;42: 4709–4727.
777 doi:10.1039/c2cs35408j [doi]
778 10. Cozens D, Read RC. Anti-adhesion methods as novel therapeutics for bacterial infections. Expert Rev
779 Anti Infect Ther. 2012;10: 1457–1468. doi:10.1586/eri.12.145
780 11. Zhang B, Ku X, Yu X, Sun Q, Wu H, Chen F, et al. Prevalence and antimicrobial susceptibilities of
781 bacterial pathogens in Chinese pig farms from 2013 to 2017. Sci Rep. 2019;9: 9908.
782 doi:10.1038/s41598-019-45482-8
783 12. Vötsch D, Willenborg M, Weldearegay YB, Valentin-Weigand P. *Streptococcus suis* - The "two faces"
784 of a pathobiont in the porcine respiratory tract. Frontiers in Microbiology. 2018. p. 9:480.
785 doi:10.3389/fmicb.2018.00480
786 13. Fittipaldi N, Segura M, Grenier D, Gottschalk M. Virulence factors involved in the pathogenesis of the
787 infection caused by the swine pathogen and zoonotic agent *Streptococcus suis*. Future Microbiol.
788 2012;7: 259–279. doi:10.2217/fmb.11.149; 10.2217/fmb.11.149
789 14. Baums CG, Valentin-Weigand P. Surface-associated and secreted factors of *Streptococcus suis* in
790 epidemiology, pathogenesis and vaccine development. Anim Health Res Rev. 2009;10: 65–83.
791 doi:10.1017/S146625230999003X; 10.1017/S146625230999003X
792 15. Kouki A, Pieters RJRJ, Nilsson UJUJ, Loimaranta V, Finne J, Haataja S. Bacterial Adhesion of
793 *Streptococcus suis* to Host Cells and Its Inhibition by Carbohydrate Ligands. Biology (Basel). 2013;2:
794 918–935. doi:10.3390/biology2030918

795 16. Haataja S, Tikkanen K, Nilsson U, Magnusson G, Karlsson K-AA, Finne J. Oligosaccharide-receptor
796 interaction of the Gal alpha 1-4Gal binding adhesin of *Streptococcus suis*. Combining site architecture
797 and characterization of two variant adhesin specificities. J Biol Chem. 1994;269: 27466–27472.
798 17. Ferrando ML, Willemse N, Zaccaria E, Pannekoek Y, Van Der Ende A, Schultsz C. Streptococcal
799 Adhesin P (SadP) contributes to *Streptococcus suis* adhesion to the human intestinal epithelium. PLoS
800 One. 2017;12: e0175639. doi:10.1371/journal.pone.0175639
801 18. Kong D, Chen Z, Wang J, Lv Q, Jiang H, Zheng Y, et al. Interaction of factor H-binding protein of
802 *Streptococcus suis* with globotriaosylceramide promotes the development of meningitis. Virulence.
803 2017;8: 1290–1303. doi:10.1080/21505594.2017.1317426
804 19. Haataja S, Verma P, Fu O, Papageorgiou AC, Pöysti S, Pieters RJ, et al. Rationally Designed
805 Chemically Modified Glycodendrimer Inhibits *Streptococcus suis* Adhesin SadP at Picomolar
806 Concentrations. Chem - A Eur J. 2018;24: 1905–1912. doi:10.1002/chem.201704493
807 20. Karlsson H, Halim A, Teneberg S. Differentiation of glycosphingolipid-derived glycan structural
808 isomers by liquid chromatography/mass spectrometry. Glycobiology. 2010.
809 doi:10.1093/glycob/cwq070
810 21. Chai W, Lawson AM, Piskarev V. Branching pattern and sequence analysis of underivatized
811 oligosaccharides by combined MS/MS of singly and doubly charged molecular ions in negative-ion
812 electrospray mass spectrometry. J Am Soc Mass Spectrom. 2002. doi:10.1016/S1044-0305(02)00363-
813 X
814 22. Robbe C, Capon C, Coddeville B, Michalski JC. Diagnostic ions for the rapid analysis by nano-
815 electrospray ionization quadrupole time-of-flight mass spectrometry of O-glycans from human mucins.
816 Rapid Commun Mass Spectrom. 2004. doi:10.1002/rcm.1352
817 23. Suzuki N, Khoo KH, Chen HC, Johnson JR, Lee YC. Isolation and characterization of major
818 glycoproteins of pigeon egg white: ubiquitous presence of unique N-glycans containing Galalpha1-
819 4Gal. J Biol Chem. 2001;276: 23221–23229. doi:10.1074/jbc.M101379200
820 24. Nilsson U, Ray AK, Magnusson G. Synthesis of the globotetraose tetrasaccharide and terminal tri- and
821 di-saccharide fragments. Carbohydr Res. 1994. doi:10.1016/0008-6215(94)90010-8
822 25. Ohlsson J, Larsson A, Haataja S, Alajaaski J, Stenlund P, Pinkner JSS, et al. Structure-activity

823 relationships of galabioside derivatives as inhibitors of *E. coli* and *S. suis* adhesins: nanomolar
824 inhibitors of S. suis adhesins. OrgBiomolChem. 2005;3: 886–900. doi:10.1039/b416878j
825 26. Steil D, Bonse R, Meisen I, Pohlentz G, Vallejo G, Karch H, et al. A topographical atlas of shiga toxin
826 2e receptor distribution in the tissues of weaned piglets. Toxins (Basel). 2016.
827 doi:10.3390/toxins8120357
828 27. Diswall M, Ångström J, Karlsson H, Phelps CJ, Ayares D, Teneberg S, et al. Structural characterization
829 of α1,3-galactosyltransferase knockout pig heart and kidney glycolipids and their reactivity with human
830 and baboon antibodies. Xenotransplantation. 2010. doi:10.1111/j.1399-3089.2009.00564.x
831 28. Kurl DNN, Haataja S, Finne J. Hemagglutination activities of group B, C, D, and G streptococci:
832 Demonstration of novel sugar-specific cell-binding activities in *Streptococcus suis*. Infect Immun.
833 1989;57: 384–389.
834 29. Weinert LA, Chaudhuri RR, Wang J, Peters SE, Corander J, Jombart T, et al. Genomic signatures of
835 human and animal disease in the zoonotic pathogen *Streptococcus suis*. Nat Commun. 2015.
836 doi:10.1038/ncomms7740
837 30. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid genomic
838 surveillance for public health and hospital microbiology labs. Genome Med. 2014.
839 doi:10.1186/s13073-014-0090-6
840 31. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: Rapid large-scale
841 prokaryote pan genome analysis. Bioinformatics. 2015. doi:10.1093/bioinformatics/btv421
842 32. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large
843 phylogenies. Bioinformatics. 2014. doi:10.1093/bioinformatics/btu033
844 33. Monaco S, Gordon E, Bowler MW, Delagenière S, Guijarro M, Spruce D, et al. Automatic processing
845 of macromolecular crystallography X-ray diffraction data at the ESRF. J Appl Crystallogr. 2013.
846 doi:10.1107/S0021889813006195
847 34. Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta Crystallogr Sect D
848 Biol Crystallogr. 2013. doi:10.1107/S0907444913000061
849 35. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser
850 crystallographic software. J Appl Crystallogr. 2007. doi:10.1107/S0021889807021206

851 36. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta
852 Crystallogr Sect D Biol Crystallogr. 2006. doi:10.1107/S0907444906022116
853 37. Potterton L, Agirre J, Ballard C, Cowtan K, Dodson E, Evans PR, et al. CCP 4 i 2: The new graphical
854 user interface to the CCP 4 program suite. Acta Crystallogr Sect D Struct Biol. 2018.
855 doi:10.1107/S2059798317016035
856 38. Adams PD, Afonine P V., Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: A
857 comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr Sect D
858 Biol Crystallogr. 2010. doi:10.1107/S0907444909052925
859 39. Emsley P, Cowtan K. Coot: Model-building tools for molecular graphics. Acta Crystallogr Sect D Biol
860 Crystallogr. 2004. doi:10.1107/S0907444904019158
861 40. Karlsson KA. Preparation of Total Nonacid Glycolipids for Overlay Analysis of Receptors for Bacteria
862 and Viruses and for Other Studies. Methods Enzymol. 1987. doi:10.1016/0076-6879(87)38018-8
863 41. Samuelsson BE, Pimlott W, Karlsson KA. Mass spectrometry of mixtures of intact glycosphingolipids.
864 Methods Enzymol. 1990. doi:10.1016/0076-6879(90)93442-N
865 42. Koerner TAW, Prestegard JH, Demou PC, Yu RK. High-Resolution Proton NMR Studies of
866 Gangliosides. 1. Use of Homonuclear Two-Dimensional Spin-Echo J-Correlated Spectroscopy for
867 Determination of Residue Composition and Anomeric Configurations. Biochemistry. 1983.
868 doi:10.1021/bi00280a014
869 43. E S. Dünnschicht-Chromatographie. Berlin: Springer-Verlag; 1962.
870 44. Jansson L, Ångström J, Lebens M, Imberty A, Varrot A, Teneberg S. Carbohydrate binding specificities
871 and crystal structure of the cholera toxin-like B-subunit from *Citrobacter freundii*. Biochimie. 2010.
872 doi:10.1016/j.biochi.2010.02.010
873 45. Johansson MM, Dedic B, Lundholm K, Branzell FB, Barone A, Benktander J, et al. Characterization
874 of moose intestinal glycosphingolipids. Glycoconj J. 2015. doi:10.1007/s10719-015-9604-8
875 46. Halling KK, Ramstedt B, Nyström JH, Slotte JP, Nyholm TKM. Cholesterol interactions with fluid-
876 phase phospholipids: Effect on the lateral organization of the bilayer. Biophys J. 2008.
877 doi:10.1529/biophysj.108.133744
878 47. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server.

Nucleic Acids Res. 2014. doi:10.1093/nar/gku316

Tables

Table 1.

Table 1. Binding of SadP to glycosphingolipids on thin-layer chromatograms

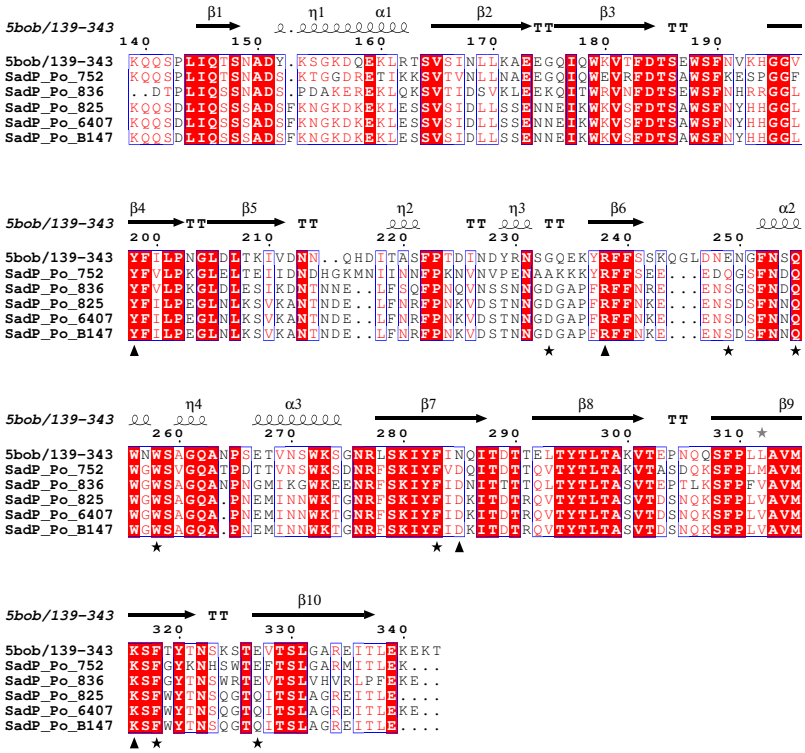
No. Trivial name	Structure	SadP type P_N/P_O	Source
1. Galabiaosylcer (Gb2)	Galα1–4Galβ–O–Cer	-	Human meconium
2. Galabiaosylcer (Gb2-S)	Galα1–4Galβ1–O–*bis*–$(SO_2\text{-}C_{16}H_{33})_2$	-	Synthetic
3. Isoglobotri (iGb3)	Galα1–3Galβ1–4Glcβ1–O–Cer	+/-	Cat intestine
4. Globotri (Gb3)	Galα1–4Galβ1–4Glcβ1–O–Cer	+++/+++	Human erythrocytes
5. Globotetra (Gb4)	GalNAcβ1–3Galα1–4Galβ1–4Glcβ1–O–Cer	+/-	Human erythrocytes
6. Forssman	GalNAcα1–3GalNAcβ1–3Galα1–4Galβ1–4Glcβ1–O–Cer	+/-	Dog intestine
7. para-Forssman	GalNAcβ1–3GalNAcβ1–3Galα1–4Galβ1–4Glcβ1–O–Cer	-	Human erythrocytes
8. P1	Galα1–4Galβ1–4GlcNAcβ1–3Galβ1–4Glcβ1–O–Cer	+++/+++	Human erythrocytes
9. H type 4/Globo H hexa	Fucα1–2Galβ1–3GalNAcβ1–3Galα1–4Galβ1–4Glcβ1–O–Cer	-	Pig intestine
10. H type 2 penta	Fucα1–2Galβ1–4GlcNAcβ1–3Galβ1–4Glcβ1–O–Cer	-	Pig lung
11. Galili penta	Galα1–3Galβ1–4GlcNAcβ1–3Galβ1–4Glcβ1–O–Cer	-	Pig lung
12. Le^b hexa	Fucα1–2Galβ1–3(Fucα1-4)GlcNAcβ1–3Galβ1–4Glcβ1–O–Cer	-	Pig lung
13. Le^y hexa	Fucα1–2Galβ1–4(Fucα1-3)GlcNAcβ1–3Galβ1–4Glcβ1–O–Cer	-	Pig lung
14. A type 2 hexa	Galα1–3(Fucα1–2)Galβ1–4GlcNAcβ1–3Galβ1–4Glcβ1–O–Cer	-	Pig erythrocytes
15. A type 4/Globo A hepta	GalNAcα1-3(Fucα1-2)Galβ1-3GalNAcβ1–3Galα1-4Galβ1-4Glcβ1–O–Cer	-	Pig intestine

[a] Binding is defined as follows: +++ denotes a highly reproducible binding of SadP type P_N/P_O when 4 μg of the glycosphingolipid was applied on the thin-layer chromatogram, while + denotes an occasional binding, and - denotes no binding even at 4 μg.

Figures

Fig 1.

A.



B

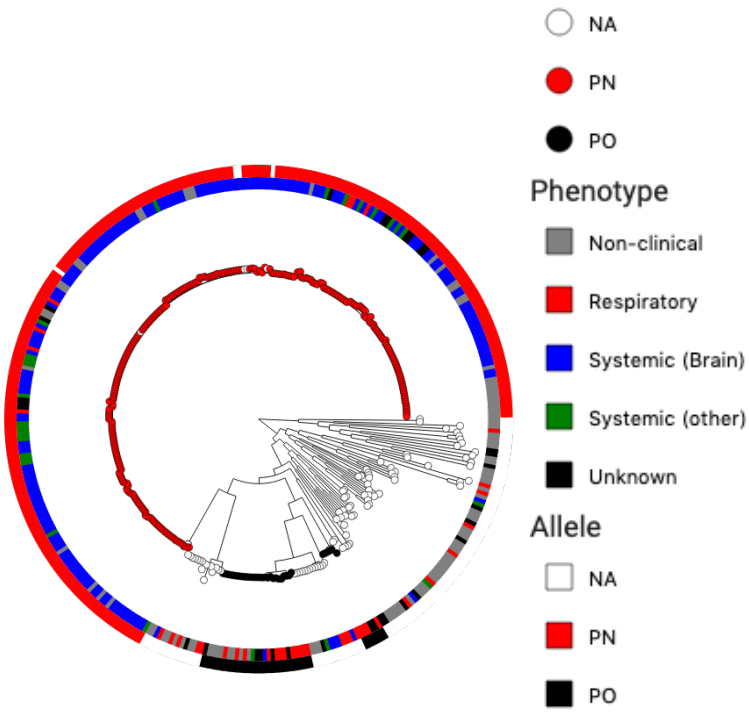


Fig 2.

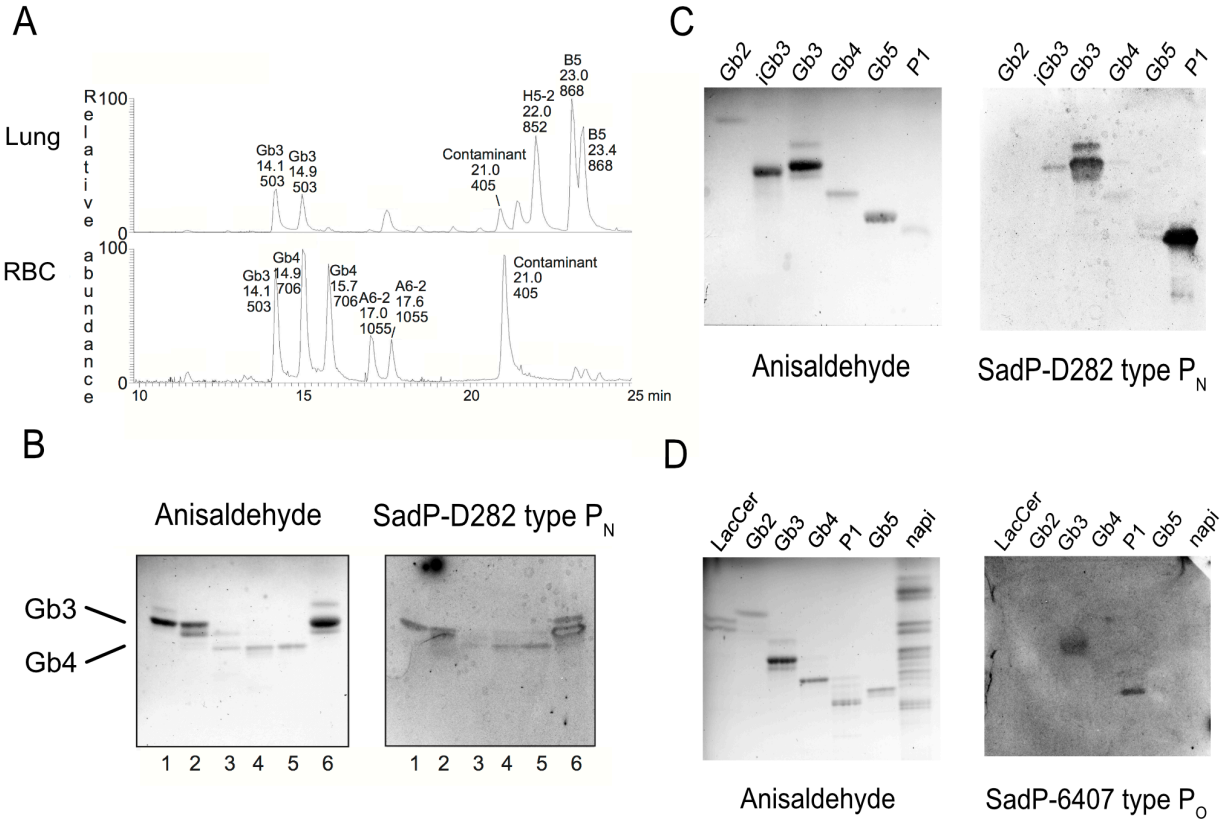

A
Lung
RBC
Relative abundance
100
0
Gb3 14.1 503
Gb3 14.9 503
Contaminant 21.0 405
H5-2 22.0 852
B5 23.0 868
B5 23.4 868
Gb4 14.9 706
Gb4 15.7 706
A6-2 17.0 1055
A6-2 17.6 1055
10
15
20
25 min
B
Anisaldehyde
SadP-D282 type P_N
Gb3
Gb4
1 2 3 4 5 6
C
Gb2 iGb3 Gb3 Gb4 Gb5 P1
Anisaldehyde
SadP-D282 type P_N
D
LacCer Gb2 Gb3 Gb4 P1 Gb5 napi
Anisaldehyde
SadP-6407 type P_O

Fig. 3.

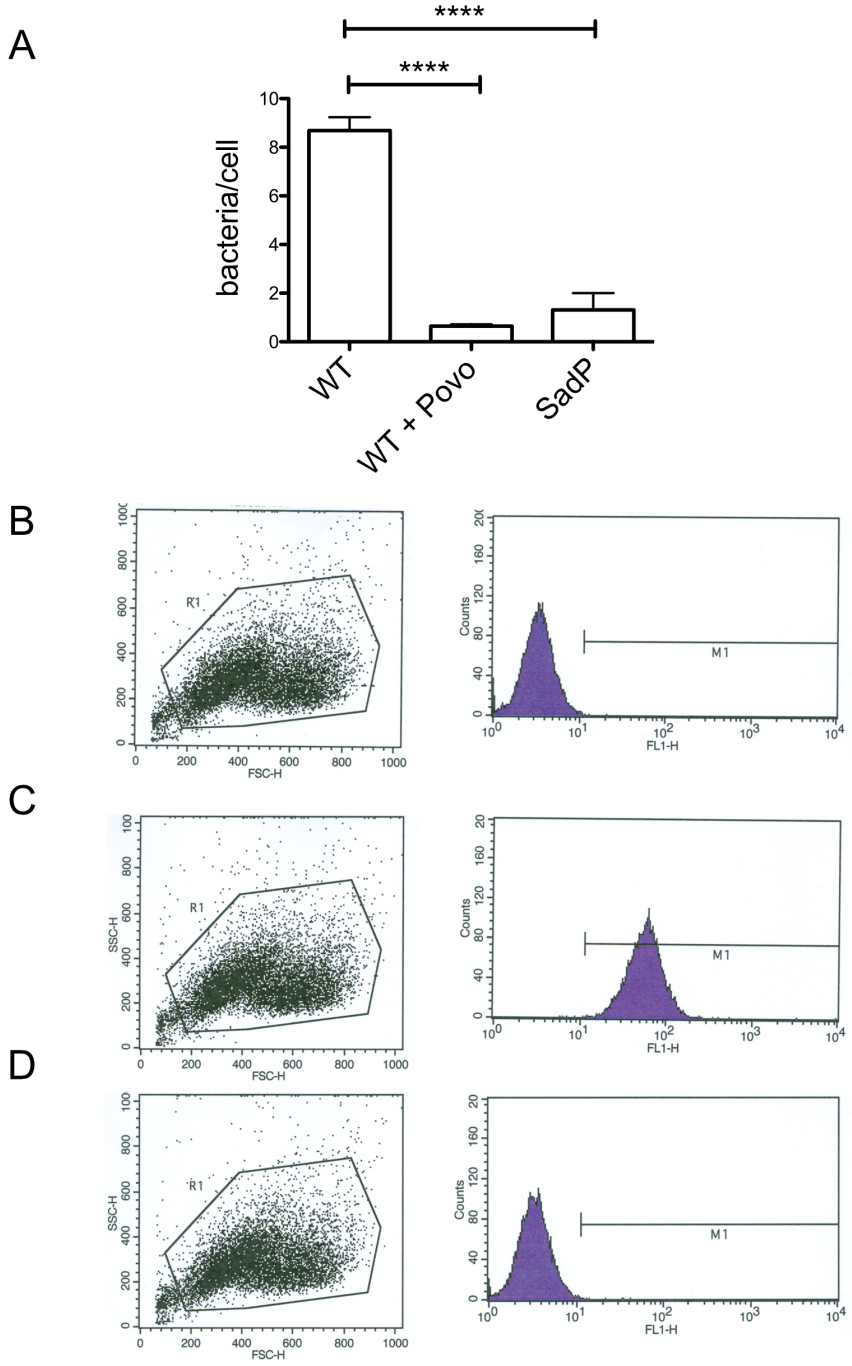

A
****
****
bacteria/cell
10
8
6
4
2
0
WT
WT + Povo
SadP
B
C
D
R1
FSC-H
SSC-H
Counts
M1
FL1-H

Fig 4.

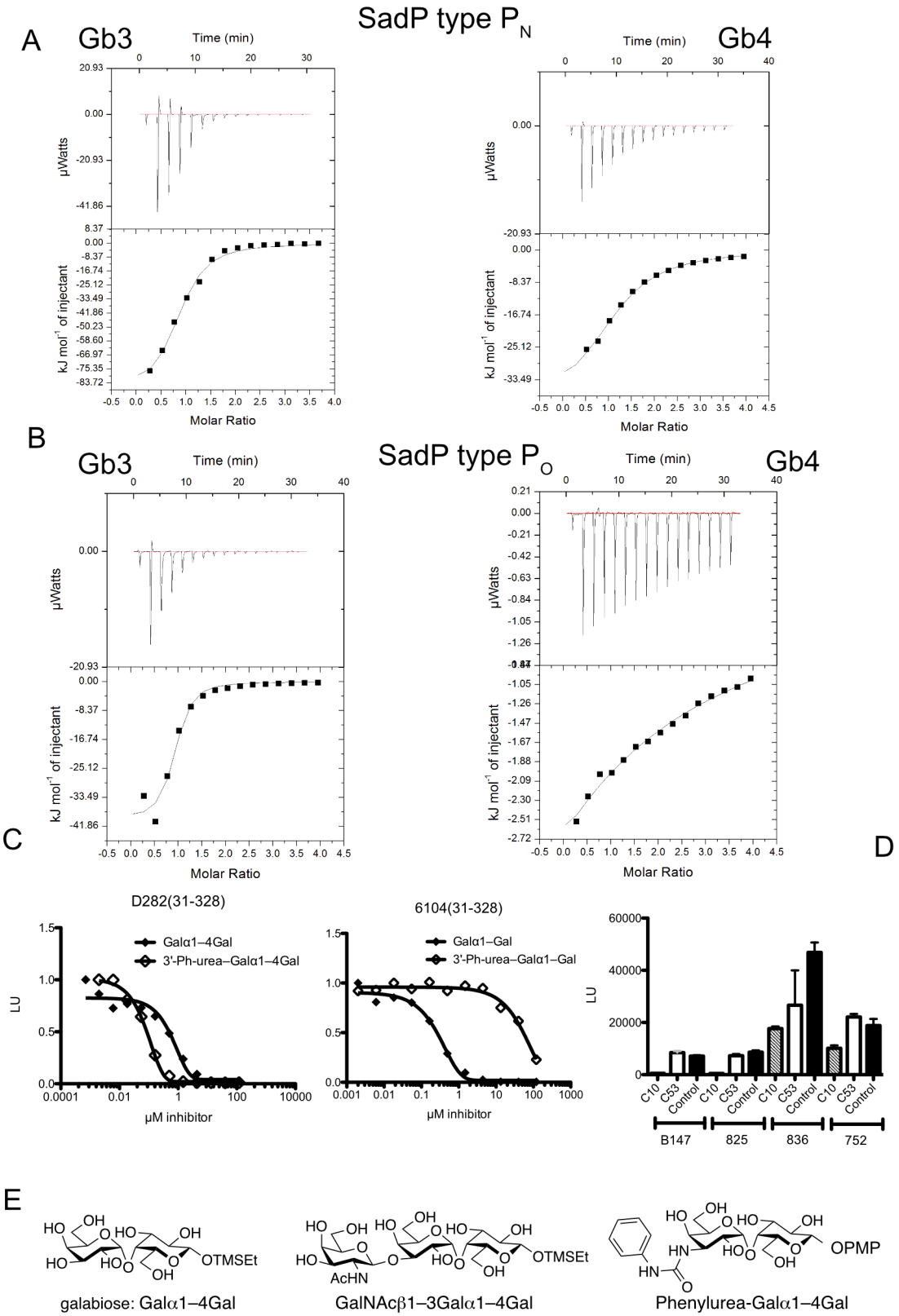

SadP type P_N
A
Gb3
Gb4
Time (min)
µWatts
kJ mol⁻¹ of injectant
Molar Ratio
B
SadP type P_O
C
D
D282(31-328)
6104(31-328)
Galα1–4Gal
3'-Ph-urea–Galα1–4Gal
Galα1–Gal
3'-Ph-urea–Galα1–Gal
LU
µM inhibitor
C10
C53
Control
B147
825
836
752
E
galabiose: Galα1–4Gal
GalNAcβ1–3Galα1–4Gal
Phenylurea-Galα1–4Gal
OTMSEt
OPMP
AcHN

Fig 5.

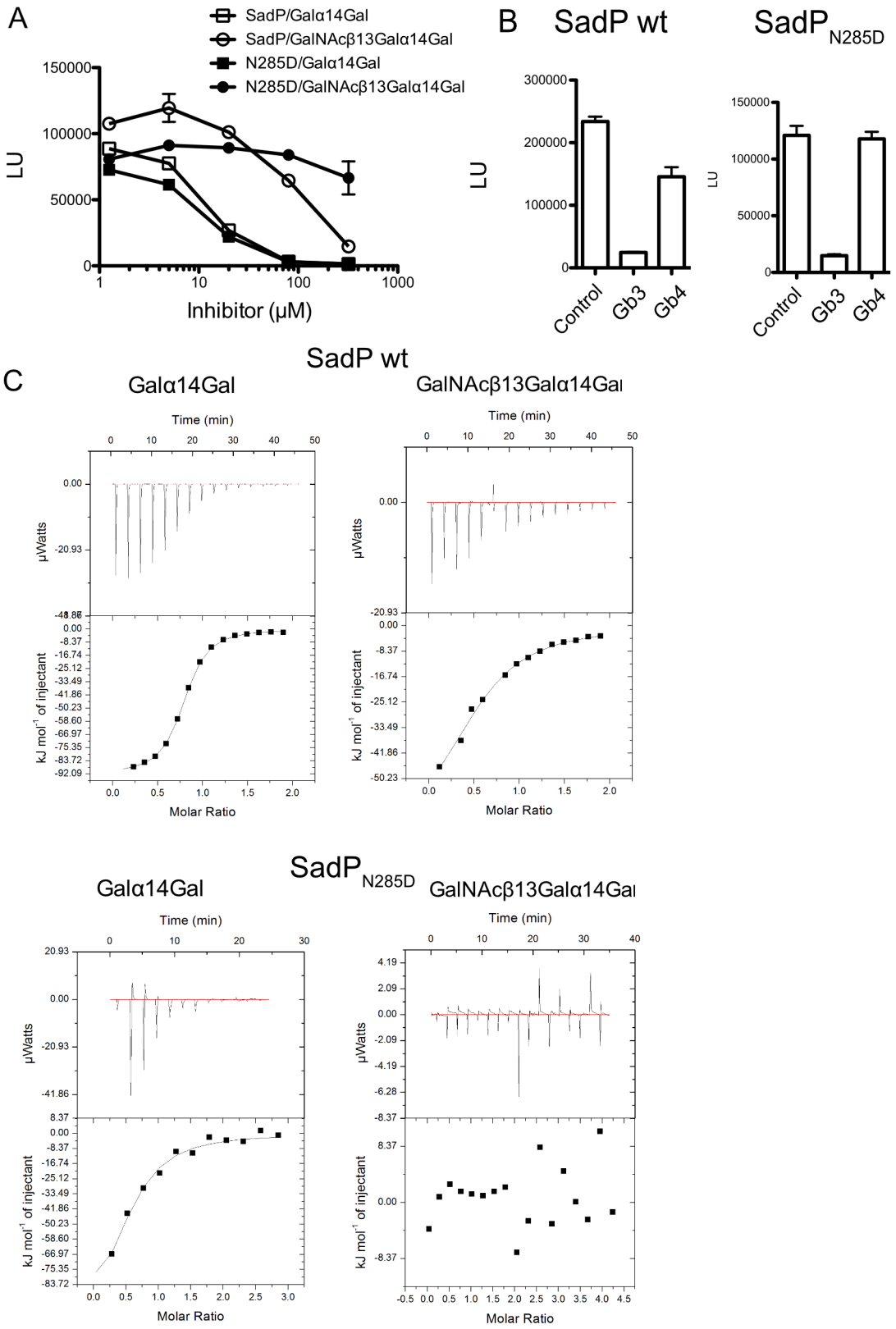

A
SadP/Galα14Gal
SadP/GalNAcβ13Galα14Gal
N285D/Galα14Gal
N285D/GalNAcβ13Galα14Gal
LU
Inhibitor (μM)
B
SadP wt
SadP N285D
Control
Gb3
Gb4
C
SadP wt
Galα14Gal
GalNAcβ13Galα14Gal
SadP N285D
Time (min)
μWatts
kJ mol-1 of injectant
Molar Ratio

Fig 6.

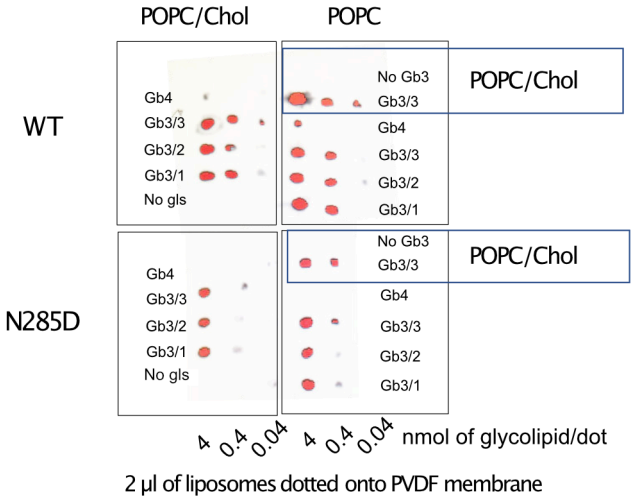

POPC/Chol
POPC
WT
N285D
Gb4
Gb3/3
Gb3/2
Gb3/1
No gls
No Gb3
Gb3/3
Gb4
Gb3/3
Gb3/2
Gb3/1
POPC/Chol
POPC/Chol
4
0.4
0.04
4
0.4
0.04
nmol of glycolipid/dot
2 µl of liposomes dotted onto PVDF membrane

Fig 7.

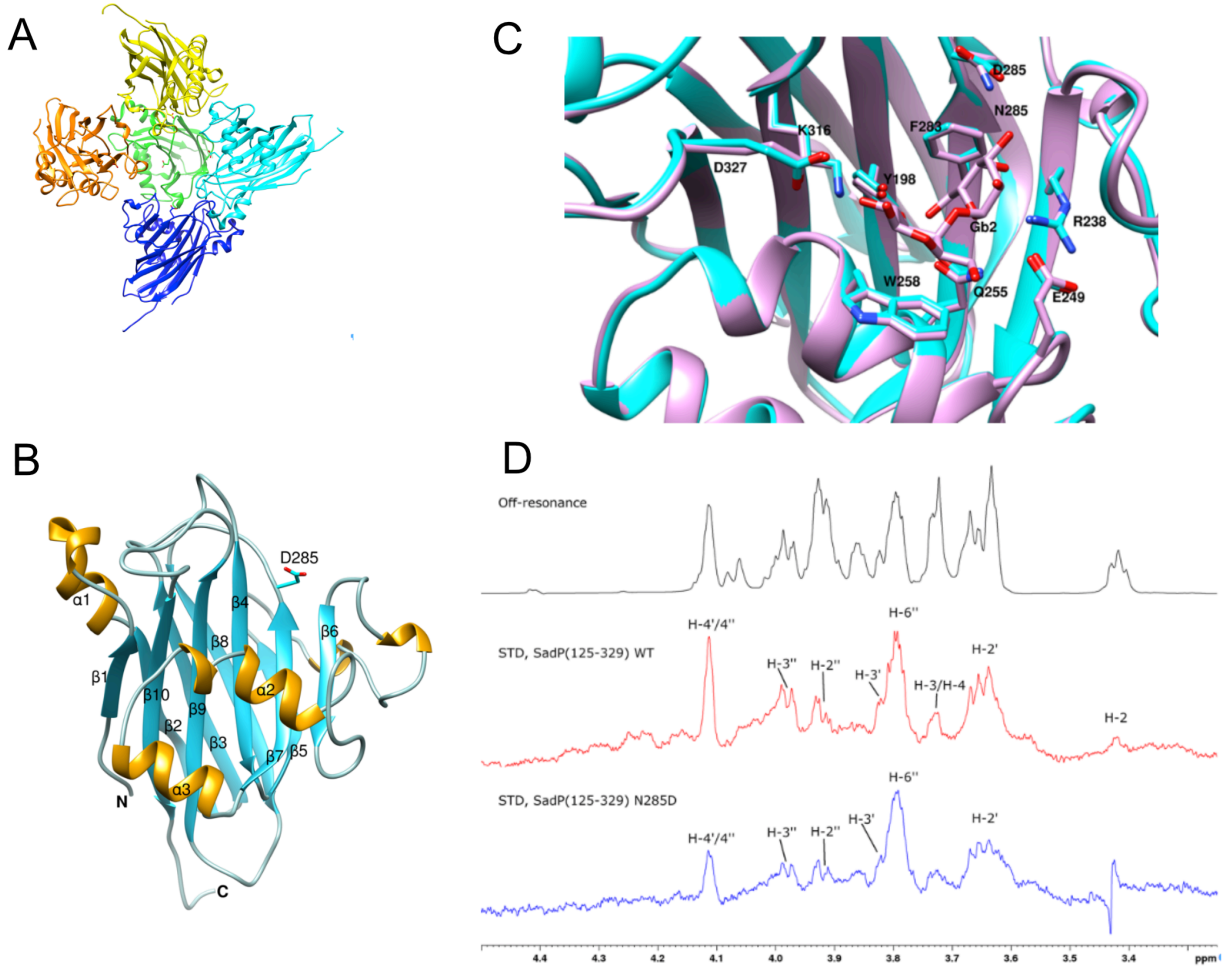

A
B
C
D
D285
N285
K316
F283
D327
Y198
Gb2
R238
W258
Q255
E249
α1
β1
β10
β2
β9
β3
β8
β4
β6
α2
β7
β5
α3
N
C
Off-resonance
STD, SadP(125-329) WT
STD, SadP(125-329) N285D
H-4'/4''
H-3''
H-2''
H-3'
H-6''
H-3/H-4
H-2'
H-2
4.4
4.3
4.2
4.1
4.0
3.9
3.8
3.7
3.6
3.5
3.4
ppm

Fig 8.

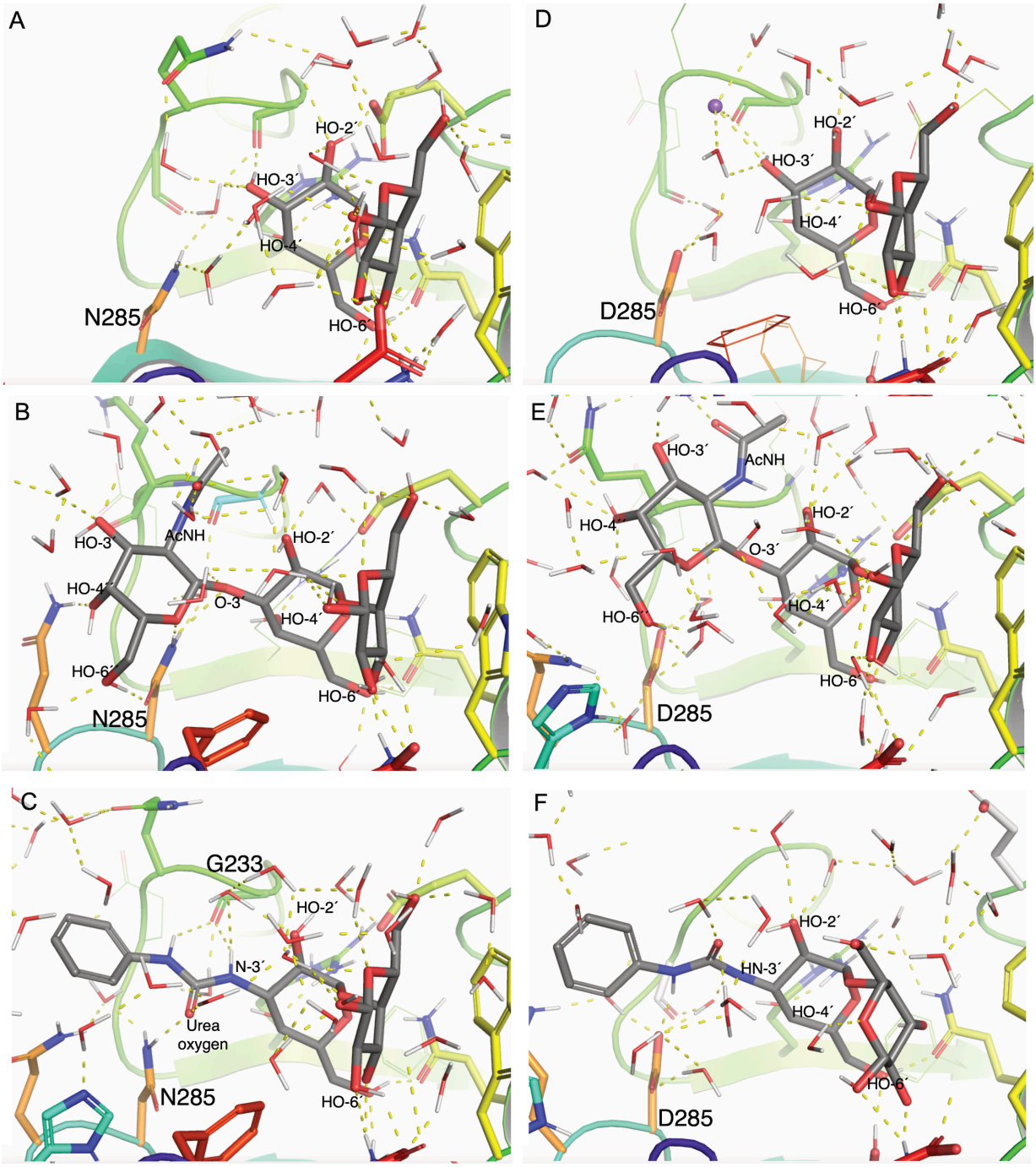

A
HO-2´
HO-3´
HO-4´
N285
HO-6´
D
HO-2´
HO-3´
HO-4´
D285
HO-6´
B
HO-3´
AcNH
HO-2´
HO-4´
O-3´
HO-4´
HO-6´´
HO-6´
N285
E
HO-3´
AcNH
HO-4´´
HO-2´
O-3´
HO-6´´
HO-4´
HO-6´
D285
C
G233
HO-2´
N-3´
Urea
oxygen
N285
HO-6´
F
HO-2´
HN-3´
HO-4´
HO-6´
D285

Fig 9.

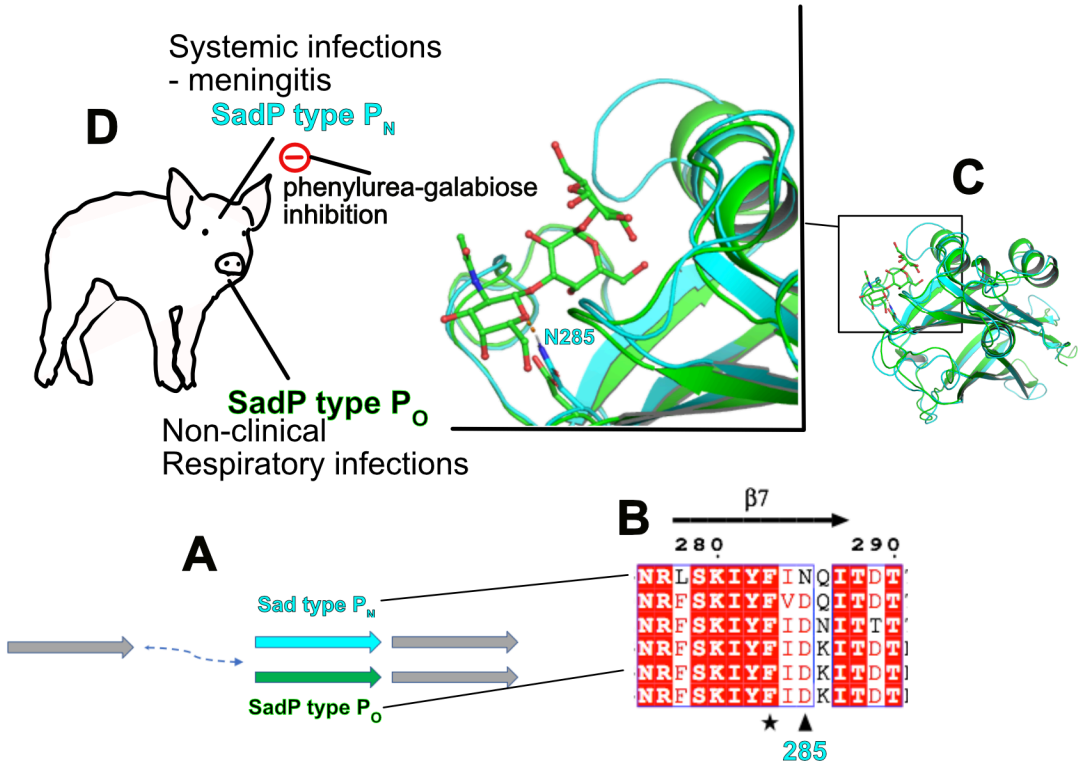

Systemic infections
- meningitis
D
SadP type P_N
phenylurea-galabiose
inhibition
SadP type P_O
Non-clinical
Respiratory infections
N285
C
A
Sad type P_N
SadP type P_O
B
β7
280
290
285

Figure legends

Fig 1. Homology of SadP types P_N and P_O galabiose-binding domains and their distribution in systemic, respiratory and non-clinical *S. suis* isolates. A. Multiple alignment of SadP galabiose binding domains of type P_N and type P_O strains. Conserved amino acids are shown in shaded red rectangles and white letters. Residues with >70% similarity according to their physicochemical properties are framed in white background and shown with red letters. Residues involved in hydrogen-bond formation with galabiose (Galα1–4Gal) are shown with (▲) and those involved in hydrophobic interactions with a star (★). Residues 139-343 of SadP are included as found in the crystal structure (pdb id 5BOB). Secondary structure elements are shown on the top. Figure was created with ESpript [47]. **B. Distribution of SadP type P_N and P_O alleles in *S. suis* clinical and non-clinical isolates.** Maximum-likelihood core genome phylogeny of 374 *S. suis* isolates from Weinert et al. 2015. The phylogeny was run with RAxML using SNP sites. The nucleotide substitution model used was GTR GAMMA. Tips of the tree are coloured according to the SadP allele detected with SRST2. NA means that with the default parameters neither allele was detected. The inner circle shows the phenotype according to Weinert et al. 2015. Grey non-clinical, red respiratory, blue systemic (brain), green systemic (other) and black unknown. The outer circle is coloured according to the SadP allele detected with SRST2 with the same colours as the tree tips. White NA, red P_N and black P_O.

Fig 2. Different glycolipid binding specificities of SadP type P_N from systemic *S. suis* strains and type P_O SadP from respiratory strains towards pig lung glycolipids and purified globo series glycolipids . A. Comparison of the total non-acid glycosphingolipid fractions from porcine lung and erythrocytes. Base peak chromatogram from LC-ESI/MS of the oligosaccharides derived the total non-acid glycosphingolipid fraction from porcine lung by digestion with *Rhodococcus* endoglycoceramidase II. B. Base peak chromatogram from LC-ESI/MS of the oligosaccharides derived the total non-acid glycosphingolipid fraction from porcine erythrocytes by digestion with *Rhodococcus* endoglycoceramidase II. The identification of individual glycosphingolipid-derived oligosaccharides was based on their determined molecular masses and subsequent MS^2 sequencing. The oligosaccharides identified in the chromatograms are: Gb3, Galα1–4Galβ1–4Glc; H5-2, Fucα2–3Galβ1–4GlcNAcβ1–3Galβ1–4Glc; B5, Galα1–3Galβ1–4GlcNAcβ1–3Galβ1–4Glc;

Gb4, GalNAcβ1–3Galα1–4Galβ1–4Glc; A6-2, GalNAcα1–3(Fucα2–3)Galβ1–4GlcNAcβ1–3Galβ1–4Glc. **B. SadP binding tri- and tetraglycosylceramides isolated from porcine lung.** Thin-layer chromatogram after detection with anisaldehyde (A), and autoradiograms obtained by binding of SadP (B), followed by autoradiography for 12 h, as described under "Methods". The solvent system used was chloroform/methanol/water (60:35:8, by volume). The lanes were: Lane 1, fraction PL-1 isolated from pig lung, 1 μg; Lane 2, fraction PL-2 isolated from pig lung, 1 μg; Lane 3, fraction PL-3 isolated from pig lung, 1 μg; Lane 4, fraction PL-4 isolated from pig lung, 1 μg; Lane 5, reference globotetraosylceramide (GalNAcβ1–3Galα1–4Galβ1–4Glcβ1Cer), 2 μg; Lane 6, reference globotriaosylceramide (Galα1–4Galβ1–4Glcβ1Cer), 2 μg. **C. TLC overlay assay with purified glycolipids**, **SadP-D282 type P_N** 1. Gb2, galabiaosylceramide 2 μg, 2. iGb3, isoglotriosylceramide 4 μg, 3 Gb3, globotriosylceramide 4 μg , 4. Gb4, globotetraosylceramide 4 μg, 5. Gb5, Forssman 4 μg, 6. P1 pentaosylceramide 4 μg. **D. SadP-6407 type P_O** 1. Lactosylceramide 4 μg, 2. Galabiaosylceramide 4 μg, 3. Globotriaosylceramide 4 μg, 4. Globotetraosylceramide 4 μg, 5. P1, pentaosylceramide 4 μg, 6. Forssman pentaosylceramide 4 μg, 7. Non-acid glycosphingolipids of porcine small intestine (blood group O) 40 μg.

Fig 3. SadP binding to EA.hy926 cell line. A. Binding of S. suis D282 wild type serotype 2 strain and the SadP insertion mutation to EA.hy926 cells grown on coverslips. The binding of the wild type bacteria was inhibited with 10 μg/ml of pigeon ovomucoid. Flow cytometry analysis of SadP(31-328) type P_N binding to EA.hy926 cells. **B.** Buffer control. **C.** SadP adhesin (0.4 μg/ml) and **D.** SadP adhesin in the presence of 10 μg/ml of pigeon ovomucoid.

Fig 4. Oligosaccharide specificity of SadP from systemic (type P_N) and respiratory (type P_O) *S. suis* strains analyzed with glycans representing the terminal saccharide structure of Gb3 and Gb4. A. and B. ITC of SadP type P_N and type P_O with Galα1–4Galβ1– and GalNAcβ1–3Galα1–4Galβ1– TMSEt derivatives **C.** AlphaScreen inhibition assay with Galα1–4Galβ1– and 3'Phenylurea-Galα1–4Galβ1– methoxyphenyl (-PMP = 4-methoxyphenyl). **D**. AlphaScreen inhibition assay of Galα1–4Galβ1– (compound

C10) and 3'Phenylurea-Galα1–4Galβ1– methoxyphenyl (compound C53) with recombinant N-terminal domains cloned from 4 different type P_O *S. suis* strains **E**. Structures of the ligands in this study.

Fig 5. A. Single amino acid site-specific mutation N285D of SadP from systemic strains changes its Gb4 binding specificity to type P_O Gb3 binding specificity of respiratory strains. A. Inhibition of WT SadP(125-329) and its site-specific mutant SadP(125-329)N285D with the dilutions series (1.25 to 320 μM, duplicate determinations) of TMSEt (2-trimethylsilylethyl) glycosides of Galα1–4Gal and GalNAcβ1–3Galα1–4Gal representing the Gb3 and Gb4 oligosaccharides. **B.** Inhibition assay with 5 μM of Galα1–4Gal and GalNAcβ1–3Galα1–4Gal TMSEt glycosides. **C.** ITC measurements of WT SadP and N285D mutant with the TMSEt glycosides of Galα1–4Gal and GalNAcβ1–3Galα1–4Gal representing the Gb3 and Gb4 oligosaccharides.

Fig 6. SadP and $SadP_{N285D}$ binding to liposomes. POPC/glycolipid liposomes (4, 0.4 and 0.04 nmol/dot) and POPC/cholesterol/globosylceramide liposomes (4, 0.4 and 0.04 nmol/dot) were applied onto PVDF membrane. WT, His-tagged SadP(125-329) type P_N; N285D, site-specific mutant of SadP(125-329); Replicate membranes containing dotted liposomes with or without cholesterol (see Methods) were probed with **SadP and $SadP_{N285D}$**. The bound proteins were detected with anti-His and HRP-labelled antibodies as described in the methods and membrane was imaged with FujiLAS-4000. The globo series glycolipids used were: Gb4, d18:1-24:0 + d18:1 – 16:0; Gb3/1, Gb3 d18:1-24:0; Gb3/2, Gb3 d18:1-16:0; Gb3/3, Gb3 d18:1-h16:0.

Fig 7. A. Arrangement of the five site-specifically mutated SadP(125-329)N285D molecules in the crystallographic asymmetric unit. Each molecule is shown as ribbon and colored differently. **B.** Ribbon diagram of SadP(125-329)N285D structure. α-Helices are shown in yellow color and β-strands in cyan. D285 is depicted as sticks. **C.** Structural superposition of Galα1–4Gal -bound Fhc (PDB 5BOA) structure (identical to SadP type P_N (pink) onto $SadP(125\text{-}329)_{N285D}$ (cyan). Residues involved in binding interactions are shown

in stick representation. **D.** STD-NMR of binding of wild type SadP(125-329) type P_N and its site-specific mutant N285D. Peaks in the STD spectra indicate hydrogens (C-H) that are in close proximity to the protein.

Fig 8. Molecular dynamics of SadP type P_N binding to Gb4 saccharide and glycomimetic phenylurea-galabiose inhibitor. Representative MD snapshots (at 90 ns) of SadP. WT in complex with **A**. Galα1–4Gal, **B**. GalNAcβ1–3Galα1–4Galβ, and **C.** C3´-phenylurea-Galα1–4Galβ. Representative MD snapshots of SadP N285D in complex with **D.** Galα1–4Gal, **E.** GalNAcβ1–3Galα1–4Galβ, and **F.** C3´-phenylurea-Galα1–4Galβ.

Fig 9. Summary of the results. A. SadP type P_N (cyan) and type P_O (green) are putative duplicates of the downstream SadP homolog (grey). **B.** Conservative change of type P_N SadP asparagine-285 to aspartic acid of SadP type P_O. **C.** This study has shown the mechanism how N285 contributes to the binding of terminal GalNAcβ1- saccharide of Gb4 by a specific interaction with GalNAc hexose ring oxygen (orange dashed line). The 3D-structure of SadP type P_O from type 4 *S. suis* 6407 was obtained by homology modelling (Swiss Prot) and superimposed structures of type P_N (cyan) in complex with GalNAcβ1–Galα1–4Gal and type P_O (green) were created with Pymol. The structures suggest that the abolishment of Gb4 binding is based on the same mechanism as shown with N285D mutant. **D.** The results of the binding mechanisms of SadP type P_N with Gb4 and the distribution of SadP subtypes in *S. suis* clinical and non-clinical strains suggest that SadP Gb4 binding is associated with strains causing systemic disease. Therefore, the molecular mechanism of SadP type P_N binding to Gb4 could be targeted to prevent *S. suis* systemic diseases.

Supporting information

S1 Fig. LC-ESI/MS of the native fraction PL-1 from pig lung. (A) Base peak chromatogram from LC-ESI/MS of fraction PL-2 from pig lung. (B) MS^2 of the ion at *m/z* 1132 (retention time 14.1 min). The interpretation formula shows the deduced glycosphingolipid structure.

S2 Fig. LC-ESI/MS of the native fraction PL-2 from pig lung. (A) Base peak chromatogram from LC-ESI/MS of fraction PL-2 from pig lung. (B) Mass chromatogram of *m/z* 1022. (C) Mass chromatogram of *m/z* 1132. (D) MS^2 of the ion at *m/z* 1022 (retention time 13.1 min). (E) MS^2 of the ion at *m/z* 1132 (retention time 12.5 min). (F) Interpretation formulas showing the deduced glycosphingolipid structures.

S3 Fig. LC-ESI/MS of the oligosaccharides obtained by digestion of fraction PL-1, PL-2, reference globotriaosylceramide, and reference isoglobotriaosylceramide with *Rhodococcus* endoglycoceramidase II. A. Base peak chromatogram from LC-ESI/MS of the oligosaccharides derived from fraction PL-1. B. Base peak chromatogram from LC-ESI/MS of the oligosaccharides derived from fraction PL-2. C. Base peak chromatogram from LC-ESI/MS of the oligosaccharides derived from reference globotriaosylceramide (Galα1–4Galβ1–4Glcβ1–O–Cer). D. Base peak chromatogram from LC-ESI/MS of the oligosaccharides derived from reference isoglobotriaosylceramide (Galα1–3Galβ1–4Glcβ1–O–Cer). E. MS^2 of the ion at *m/z* 503 (retention time 18.4 min) in (A). F. MS^2 of the ion at *m/z* 503 (retention time 17.7 min) in (B). G. MS^2 of the ion at *m/z* 503 (retention time 17.4 min) in (C). H. MS^2 of the ion at *m/z* 503 (retention time 20.2 min) in (D). The interpretation formula shows the deduced oligosaccharide structure. The identification of oligosaccharides was based on their retention times, determined molecular masses and subsequent MS^2 sequencing.

S4 Fig. LC-ESI/MS of the oligosaccharides obtained by digestion of fraction PL-4, and reference globotetraosylceramide with *Rhodococcus* endoglycoceramidase II. A. Base peak chromatogram from LC-ESI/MS of the oligosaccharides derived from fraction PL-4. B. MS^2 of the ion at *m/z* 706 (retention time 18.4 min) in (A). C. MS^2 of the ion at *m/z* 706 (retention time 18.6 min) from LC-ESI/MS of the oligosaccharides

derived from reference globotetraosylceramide (GalNAcβ1–3Galα1–4Galβ1–4Glcβ1–O–Cer). D. MS^2 of the ion at *m/z* 998 (retention time 19.4 min) in (A). E. MS^2 of the ion at *m/z* 998 (retention time 21.4 min) in (A). The interpretation formula shows the deduced oligosaccharide structure. The identification of oligosaccharides was based on their retention times, determined molecular masses and subsequent MS^2 sequencing.

S5 Fig. SadP types P_N and P_O glycolipid binding specificity analysed with TLC overlay analysis and microwell binding assay: A. Decreasing glycolipid dilutions from 2 to 0.2 µg per lane probed with SadP type P_N full length and it's N-terminal domain SadP(31-328). **B.** Microwell binding assay with radiolabelled SadP type P_N to 250 ng of glycolipids, Gb2 galabiosylceramide, Gb2-S Galα1–4Galβ1–O-*bis*-$(SO_2\text{-}C_{16}H_{33})_2$, Gb3 non-OH globotriaosylceramide with non-hydroxy ceramide, Gb3 OH globotriaosylceramide with hydroxy ceramide, iGb3 isoglobotriaosylceramide, Gb4, globotetraosylceramide, P1 pentaosylceramide. Leb Lewis b hexaosylceramide.

S6 Fig. SDS-PAGE of the type P_N WT SadP(125-329) and the site-specific mutants N285D and E292Q, deletion Δ244-246 and N285D&Δ244-246 mutants were constructed into the type P_N SadP(125-328) cloned into the plasmid pET28a. In addition, the 28 amino acid long region H216-D246 of the type P_N was replaced by the corresponding region of type P_O.

S7 Fig. A. Binding of SadP wild type and site-specific mutants to Galα1–4Gal analysed with AlphaScreen**. B - F.** Inhibition assay for the binding of site-specific mutants of PSAD1-5 (PSAD-1 Δ244-246, PSAD-2 N285D, PSAD-3 E292Q, PSAD-4 N285D&Δ244-246 and PSAD-5 H216-D246delinsDELFNRFPNKVDSTNNGDGAPFRFFNKE (replacement of aa 216-242 into corresponding sequence of type P_O of strain 6107).

S8 Fig. Structure analysis of TMSEt (2-trimethylsilylethyl) glycosides

S9 movie. MD simulation of type P_N WT SadP(125-329) and Galα1–4Gal

S10 movie. MD simulation of N285D SadP(125-329) and Galα1–4Gal

S11 movie. MD simulation of of type P_N WT SadP(125-329) and GalNAcβ1–3Galα1–4Gal

S12 movie. MD simulation of of type P_N WT SadP(125-329) and 3'-phenylurea- Galα1–4Gal

S13 movie. MD simulation of N285D SadP(125-329) and GalNAcβ1–3Galα1–4Gal

S14 movie. MD simulation of N285D SadP(125-329) and 3'-phenylurea- Galα1–4Gal