A novel and highly specific Forssman antigen-binding protein from sheep polyomavirus

Nils H. Rustmeier[1], Lisete M. Silva[2], Antonio Di Maio[2], Joshua C. Müller[1], Alexander Herrmann[1], Ten Feizi[2], Yan Liu[2], Thilo Stehle[1]

[1]Interfaculty Institute of Biochemistry, University of Tübingen, Tübingen, Germany

[2]Glycosciences Laboratory, Faculty of Medicine, Imperial College London, London, UK

Manuscript Information:

Word count, Abstract: 245

Word count, Author summary: 167

Word count, Text: 4492

Corresponding author:

Thilo Stehle: thilo.stehle@uni-tuebingen.de

Phone: +49-7071-2978090

Abstract

Polyomaviruses are small, non-enveloped double-stranded DNA viruses of humans and other mammals, birds, and fish. Infections are usually asymptomatic and result in latency, however, some polyomaviruses can induce severe diseases, including cancer, in immunocompromised individuals. Established cellular receptors for polyomavirus infection are sialylated glycolipids (such as gangliosides), membrane proteins, and glycosaminoglycans. Polyomaviruses are usually highly host specific, but the exact

principles that govern host tropism remain unknown in many cases. Here, glycan array screening shows that the major capsid protein VP1 of sheep polyomavirus (ShPyV) binds to the Forssman glycolipid, an antigen of many vertebrates and a potential tumor marker in humans. Following closer investigation, we can report for the first time that a neutral, non-sialylated glycolipid acts as a polyomavirus receptor. Concurrently, we present the first report of a viral protein that specifically engages the Forssman antigen. We demonstrate that ShPyV VP1 binds to Forssman-positive erythrocytes but not those of human A, B and O blood groups, which is a clear distinction from features thus far described for Forssman lectins. X-ray crystallography and structural analysis of the VP1-Forssman glycan complex define the terminal Forssman disaccharide as the determinant of this protein-receptor interaction. These results strongly suggest that the sheep polyomavirus can use Forssman antigen for infectious cell entry. Furthermore, the ability of ShPyV VP1 to distinguish Forssman-positive from -negative cells may prove useful for monitoring the Forssman-‘status’ of normal, preneoplastic and neoplastic cells and tissues and establishing the antigen level as a biomarker.

Author summary

Elucidation of host cell receptor specificities of viral infection is crucial to understand the pathobiology of associated diseases and develop treatments. However, for many polyomaviruses the receptor engagement as the initial event in infection is poorly understood. In only a few cases polyomavirus tropism has been pinned down to a single type of glycan receptor. While many polyomaviruses utilize sialyl glycans to attach to host cells, the role of non-sialylated glycans as receptors is so far underestimated. Here, we show for the first time that a glycan of neutral charge, in this case the carbohydrate portion of the Forssman antigen, acts as an attachment receptor for a polyomavirus capsid protein and may thus contribute to host tropism and infective

52 cell entry. These results represent a significant addition to knowledge on polyomavirus-
53 glycan interactions and complement general principles of carbohydrate engagement
54 by viruses. Furthermore, as a specific binding protein of Forssman antigen, VP1 may
55 help to determine levels of this antigen in healthy and malignant tissues in humans.
56
57 **Introduction**
58
59 The Forssman antigen (FA), a neutral globo-series glycosphingolipid (GSL, see Table
60 1), is produced by the glycosyltransferase activity of Forssman synthetase (FS)[1,2].
61 Encoded by the *GBGT1* gene, FS catalyzes the addition of αGalNAc to the non-
62 reducing end of globoside (see Table 1). This reaction results in the key immunogenic
63 motif of the Forssman antigen, the terminal GalNAcα1-3GalNAcβ Forssman
64 disaccharide[3]. With few exceptions in humans, *GBGT1* is transcribed but encodes for
65 an inactive FS, resulting in a predominantly Forssman-negative human population[4,5].
66 So far, only a small number of healthy individuals worldwide were found to carry an
67 activating mutation within *GBGT1.* In these individuals, the resulting allele gives rise to
68 a Forssman antigen-positive (FA+) erythrocyte phenotype, which is now recognized as
69 the human FORS blood group[6–8]. Other reports of healthy human FA+ tissues are
70 scarce, and human sera often contain natural anti-FA antibodies[9–11]. More frequently,
71 Forssman antigenicity was described in cancerogenesis and a few other diseases of
72 humans[12–21]. Yet, both the genetics and the precise role of FA expression in disease
73 remain poorly understood.
74
75 The Forssman glycolipid also serves as a receptor for different bacterial infections in
76 animals[22–28], however, it is so far unknown if viruses also utilize FA to target their hosts.
77 In this study, by means of glycan microarray screening, NMR, surface plasmon

78 resonance (SPR), and cell binding assays, we show that FA, which is the predominant
79 glycolipid in ovine erythrocytes[29], serves as a cellular attachment receptor for the VP1
80 of a polyomavirus originally identified in sheep[30]. This designated sheep polyomavirus
81 (ShPyV) is a member of a non-enveloped, double-stranded DNA virus family with
82 icosahedral capsids, scaffolded by the major capsid protein VP1. To initiate an infection
83 many polyomaviruses rely on the interaction between their VP1 and sialic acid-
84 equipped glycolipids in the target cell membranes[31–36]. Although other infection
85 receptors such as glycoproteins and glycosaminoglycans have also been reported for
86 some polyomaviruses[37–42], detailed structural information is so far only available for
87 the interactions between VP1s and sialylated oligosaccharides. Thus, our studies
88 present the first example of a polyomavirus VP1 protein interacting with a non-sialyl
89 neutral glycan, complementing the principles of how polyomaviruses recognize their
90 targets.
91
92 **Results**
93
94 <u>Sheep polyomavirus VP1 binds the Forssman epitope with a strong preference over</u>
95 <u>other glycans</u>
96
97 To identify potential attachment factors of ShPyV, we prepared recombinant VP1
98 pentamers as previously described[43]. The protein was analyzed using a neoglycolipid-
99 based glycan screening array encompassing a broad spectrum of lipid-linked glycan
100 probes[44]. A full list of 672 glycan probes is in Table S1, and these include major types
101 of mammalian sequences found on glycoproteins (N- and O-linked), glycolipids, blood
102 group and Lewis antigen related, oligosaccharide fragments of glycosaminoglycan
103 heparin, as well as those derived from polysaccharides of bacteria, fungi, and plants.

104 Among these are 204 sialylated glycan probes variously bound by several
105 polyomavirus VP1s described previously[45–49]. We detected a signal for the Forssman
106 glycolipid that surpasses all other signals by at least one order of magnitude (Fig. 1).
107 A dominance of the terminal GalNAcα1-3GalNAcβ Forssman disaccharide is apparent
108 as the VP1 did not give binding signals with globoside glycolipid (P-antigen, position
109 25, Table S1), the precursor of the Forssman glycolipid which lacks the terminal
110 α-linked GalNAc residue. Furthermore, there were no signals with other GalNAc-
111 terminating glycan probes, including the blood group A related glycan probes, and the
112 two Tn antigen probes GalNAcα-Ser and -Thr (positions 221 and 222, Table S1). The
113 disaccharide probe GalNAcα1-3GalNAc (position 223) with a ring-opened GalNAc
114 core (reduced by reductive amination) was also not bound. These suggest the
115 importance of adjoining monosaccharides beyond the αGalNAc residue for ShPyV VP1
116 binding. Among the sialylated glycans analyzed only one short α2,6-linked sialyl probe
117 with a 9-O-acetylated N-acetylneuraminic acid residue (Neu5,9Ac_2) yielded a weak
118 signal (position 607). No binding signals were detected in the negative control
119 experiment with detection antibodies alone in the absence of ShPyV VP1 (Table S1).
120
121 To assess if the glycan binding revealed by the glycan microarray also occurs in
122 solution, we evaluated the interactions between ShPyV VP1 and the Forssman
123 pentaose (F_P) in the presence of α2,6-sialyllactose (6'SL) by saturation transfer
124 difference NMR (STD NMR) spectroscopy, which reveals protein-to-ligand
125 magnetization transfer in the case of short-term binding (10^{-8} M < K_D < 10^{-3} M)[50,51]. In
126 the first instance, the ^{1}H reference resonances of isolated F_P and 6'SL were collected
127 (Fig. 2a and 2b), which allowed us to unambiguously assign the oligosaccharide
128 signals[43,52] (except for the overlaps within 3.5-4.0 ppm) in the ^{1}H spectrum of a sample
129 comprising VP1, F_P, and 6'SL (Fig. 2c). Subsequently, we used the same sample to

130 perform the STD NMR experiment. Evaluation of the resulting spectrum reveals
131 magnetization transfer to both F_P and 6'SL, which infers that VP1 interacts with both
132 glycans in the solution (Fig. 2b). However, while resonances derived from all
133 monosaccharides in F_P (two GalNAc, two Gal, and a Glc) received magnetization
134 transfer, only the sialic acid methyl group of 6'SL re-emerged in the STD spectrum.
135 Yet, these results are in good agreement with the findings from the glycan array
136 analysis of ShPyV VP1, which indicated a much higher affinity for F_P than sialylated
137 glycans (Fig. 1). Strikingly, the methyl group of the αGalNAc in F_P gives a particularly
138 strong peak in the STD spectrum (Fig. 2b), which implies a prominent (and potentially
139 specificity-conferring) role in the VP1-F_P interaction.
140
141 <u>ShPyV VP1 has a nanomolar affinity towards Forssman pentaose-decorated surfaces</u>
142
143 After assessing the complex formation between ShPyV VP1 and F_P in the context of
144 the solid-supported glycan array and solution NMR, we sought to evaluate the
145 underlying binding kinetics by surface plasmon resonance (SPR) and compare the
146 performance of VP1 with the albumen gland agglutinin of *Helix pomatia* (HPA), an
147 αGalNAc-specific lectin with high affinity towards the Forssman antigen[53,54]. Therefore,
148 we prepared an SPR chip decorated with F_P (see Fig. S3) and measured binding
149 responses using different dilutions of VP1 and HPA. The resulting response curves
150 show that within the given association time the pentavalent VP1 and hexavalent HPA
151 bind to the immobilized F_P without reaching a complex formation equilibrium, which
152 suggests that the proteins do not readily dissociate from the surface (Fig. 3a). Indeed,
153 fitting of the kinetic parameters results in very low dissociation rates (k_{off}) for both
154 proteins, which is a common phenomenon for multivalent binding systems. However,
155 the curves of VP1 result in a higher rate for complex formation (k_{on}), determining an

156 overall approximately three-fold lower dissociation constant (K_D) compared to HPA
157 (Fig. 3a).
158
159 ShPyV VP1 specifically binds to Forssman-positive erythrocytes
160
161 To confirm the functional relevance of the VP1-F_P complex *in vitro*, we performed cell
162 binding assays using different erythrocytes and compare the cell binding of VP1 with
163 commercial reagents for Forssman antigen detection, i.e., HPA and the monoclonal
164 anti-FA IgM antibody M1/87[55]. We incubated VP1 with Forssman-positive red blood
165 cells (RBCs) from sheep and human RBCs of the different ABO phenotypes and found
166 that ShPyV VP1 bound sheep RBCs with saturation but did not attach to any of the
167 human erythrocytes (Fig. 3b left). In contrast, the αGalNAc-specific lectin HPA bound
168 to the FA on sheep RBCs and also to the antigens on human blood group A1 and A2
169 RBCs[56] (Fig. 3b left). To confirm that the carbohydrate portions of the antigens are
170 responsible for binding, we incubated the two proteins with excessive amounts of the
171 terminal trisaccharides (trioses) of FA (GalNAcα1-3GalNAcβ1-3Gal) and A substance
172 (GalNAcα1-3[Fucα1-2]Gal) and assessed the remaining sheep RBCs attachment. As
173 predicted, FA triose blocked VP1 and HPA, whereas A antigen triose could only inhibit
174 HPA attachment (Fig. 3b right). Lastly, we determined the effective protein
175 concentrations (EC50, see Methods) for sheep RBCs attachment using different
176 dilutions of ShPyV VP1, HPA, and M1/87. Evaluation of the response curves yielded
177 the lowest EC50 for ShPyV VP1 with a value of 1.2 nM, followed by M1/87 and HPA
178 with 3.3 nM and 5.2 nM, respectively (Fig. 3c). Collectively, these results demonstrate
179 the remarkable specificity and affinity of ShPyV VP1 towards the Forssman antigen.
180

181 Crystal structure of Forssman pentaose bound to VP1 reveals an intricate network of
182 interactions
183
184 To characterize the binding of Forssman antigen by ShPyV VP1 with atomic detail, we
185 solved the crystal structure of the VP1 in complex with 10 mM of Forssman pentaose
186 (F_P) at 1.92 Å resolution (Table S4). Initial electron density calculation using
187 unliganded ShPyV VP1 model (PDB 6Y61) shows that five molecules of the
188 oligosaccharide (GalNAcα1-3GalNAcβ1-3Galα1-4Galβ1-4Glc) associate with the
189 outer surface of the pentameric VP1, at a position that does not overlap with the
190 previously described sialic acid binding site[43] (Fig. S5), in a five-fold symmetric array.
191 All five Forssman binding sites are identical and located between two neighboring VP1
192 monomers (Fig. 4a). In the binding sites without artificial crystal contacts, electron
193 density provides an unambiguous outline of the last four monosaccharides of F_P (Fig.
194 4b). The intermolecular contacts show that the terminal trisaccharide of F_P contributes
195 to binding, whereas the remaining βGal and Glc divert away from VP1 and do not
196 interact with the protein surface. Closer inspection of the binding site reveals that the
197 methyl group of the terminal αGalNAc invades a hydrophobic surface depression
198 constituted by phenylalanine residues of two neighboring VP1 monomers, establishing
199 van der Waals interactions (Fig. 4c). Additionally, the *N*-acetyl moiety of αGalNAc
200 interacts with the Q266 and F66cw residues via hydrogen bonds (Fig. 4c). In βGalNAc,
201 the carbonyl oxygen forms two hydrogen bonds with the side chains of Q266 and N268
202 (Fig. 4c). Apart from these contacts, numerous interactions involve the hydroxyl groups
203 of F_P that are directed towards the protein surface. An overview of all contacts and their
204 distances is provided in Table 2. We also solved the crystal structure of VP1 in complex
205 with the FA glycan precursor, globo-*N*-tetraose, which reveals negligible binding only

206 (Fig. S5). We therefore conclude that the terminal GalNAcα1-3GalNAc disaccharide is
207 essential for the high affinity of the ShPyV VP1-FP complex.
208
209 **Discussion**
210
211 Sheep polyomavirus (ShPyV) was originally identified as a genomic contaminant in
212 processed lamb meat and it remains unknown if sheep are the natural hosts of this
213 virus[30]. Our functional data show that ShPyV engages Forssman antigen (FA)-
214 decorated cells with high affinity and specificity, making it plausible that FA initiates
215 sheep polyomavirus infection. However, in sheep the FA was thus far only detected on
216 erythrocytes but not on leukocytes, bone marrow or other tissues[29,57,58]. However, as
217 the studies about the distribution of Forssman antigen in sheep tissues are rather old,
218 the more sensitive techniques available today could help to elucidate the potential
219 targets cells for viral replication as well as the precise role of Forssman antigen during
220 ShPyV infection.
221
222 The antigenic determinant of Forssman antigen is the GalNAcα1-3GalNAcβ
223 disaccharide. In sheep erythrocytes it is predominantly linked to globotriasolyceramide,
224 giving rise to the linear Forssman pentaosylceramide (see Table 1), but it also appears
225 on the (unbranched) glycans of neolacto-series or Galili glycolipids[29]. In other animals,
226 GalNAcα1-3GalNAcβ occurs on branched structures such as extended H-like antigens
227 (dog gastric mucosa and mucus)[59,60], GM1 ganglioside (English sole fish)[61], isoglobo-
228 ganglio-neolacto glycans (swine)[62], as well as in the form of mucin core 5 (GalNAcα1-
229 3[NeuAca2-6]GalNAc) in bovine submaxillary mucin[63]. We believe that ShPyV VP1
230 can bind a range of different Forssman glycans, providing that the structural deviations
231 do not result in clashes with the protein. By the assessment of our ShPyV VP1-

232 Forssman pentaose (F_P) complex structure, the designated minimal glycan binding
233 motif is the GalNAcα1-3GalNAc disaccharide with few additional contacts to the
234 subsequent αGal. From structural perspective, this epitope recognition clearly
235 disallows branching of the glycan at certain positions, e.g., O3 of αGalNAc, N2 of
236 βGalNAc or O4 of αGal. For example, in the blood group A glycan fucose replaces the
237 *N*-acetyl group of the βGalNAc, which would result in overlaps with the protein and
238 accounts for the lack of binding (Fig. S5b). This feature distinguishes VP1 from many
239 other Forssman glycan-binding proteins such as Galectin-9, *Helix pomatia* agglutinin
240 (HPA), *Sinularia lochmodes* lectin-2, and *Dolichos biflorus* agglutinin, where
241 recognition is based on the subterminal or terminal GalNAc (see Fig. S6), resulting in
242 a much broader glycan binding capacity[64–67].
243
244 Previously we have shown that ShPyV VP1 interacts with sialylated trisaccharides in
245 crystal structures[43]. Also, our newly presented results from STD NMR and
246 crystallography reveal that Forssman and sialyl glycan binding occurs on separate
247 sites of the protein surface (Fig. S5b), most likely independently from each other. Yet,
248 in our flow cytometry assays we did not detect any binding to human erythrocytes
249 although they are known to be decorated with sialic acids. However, the preferred form
250 of sialic acid for ShPyV VP1, as evinced by our glycan array data, is 9-*O*-acetylated
251 *N*-acetylneuraminic acid (Neu5,9Ac_2), which in human erythrocytes is present in too
252 low amounts to result in detectable binding[68]. These findings imply that human cells
253 are generally incompatible targets for the sheep virus.
254
255 Humans are considered a largely Forssman-negative species and the presence of
256 Forssman antigenicity is mostly associated with human cancer[69]. During the last
257 decade, the elucidation of the genetic basis behind FA reactivation and the

258 establishment of the human FORS blood group underline the continual relevance of
259 Forssman antigen[5,6]. We propose that ShPyV VP1 complements and expands the
260 available resources for the comprehensive detection of Forssman antigen in cellular
261 and histological samples of healthy or malignant provenance. We also anticipate
262 refining mutant VP1 protein to precisely target the Forssman antigen for biomedical
263 use.
264
265 **Summary**
266
267 We show that the Forssman antigen serves as a cellular attachment receptor for the
268 major capsid protein VP1 of ShPyV. After initial detection by glycan array screening,
269 in-solution NMR spectroscopy and cell binding assays demonstrate functional
270 relevance. Structural analysis of VP1 in complex with the carbohydrate moiety of FA,
271 Forssman pentaose (F_P), characterizes the determinants of the underlying interactions
272 with high accuracy. Our findings have significance for the fields of structural virology,
273 glycobiology, and pathobiology as follows:
274 (i) We report the first structural and functional data of a viral entity to specifically
275 interact with the Forssman antigen.
276 (ii) The ShPyV VP1 – F_P complex provides the first detailed information on how
277 a polyomavirus binds to a non-sialylated ligand.
278 (iii) The high specificity and affinity of ShPyV VP1 toward Forssman antigen
279 establishes this protein as a potential biomedical reagent.
280
281 **Materials and methods**
282
283 <u>Protein expression and purification</u>

284 The VP1 of sheep polyomavirus was recombinantly produced as previously.[43] Briefly,
285 transformed *E. coli* BL21 (DE3) were grown in LB (Miller) medium supplemented with
286 100 µg/ml ampicillin at 37 °C. Expression was induced at an OD_{600} of 0.7 with 400 µM
287 IPTG and carried out at 20 °C for 16 h. ShPyV VP1 was extracted from bacterial pellets
288 by sonification and subsequent centrifugation (17,000 rpm at 4 °C). The His_6-tagged
289 VP1 was purified by immobilized-metal affinity chromatography (IMAC) and
290 subsequent gel filtration (20 mM HEPES pH 7.5, 150 mM NaCl). For crystallization of
291 ShPyV VP1 the His_6-tag was cleaved off by incubation with 10 U/mg thrombin
292 overnight at room temperature. Tags and thrombin were separated from ShPyV VP1
293 by IMAC and gel filtration. If not immediately used, the protein was stored at -20 °C.
294
295 Glycan array
296 The His_6-tagged ShPyV VP1 was analyzed in the neoglycolipid (NGL)-based
297 microarray system[70]. A broad-spectrum screening glycan microarray set
298 encompassing 672 sequence-defined lipid-linked glycan probes was used (the glycan
299 probes included, and their sequences are given in Supplemental Table S1). Details of
300 the preparation of the glycan probes and the generation of the microarrays are in
301 Supplementary Glycan Microarray Document (Supplemental Table S2) in accordance
302 with the MIRAGE (Minimum Information Required for A Glycomics Experiment)
303 guidelines for reporting of glycan microarray-based data[71]. The microarray analyses
304 were performed essentially as described[47,48]. In brief, after blocking the slides for 1h
305 with HBS buffer (10 mM HEPES, pH 7.4, 150 mM NaCl) containing 0.33% (w/v) blocker
306 Casein (Pierce), 0.3% (w/v) Bovine Serum Albumin (Sigma-Aldrich) and 5 mM $CaCl_2$,
307 the microarray was overlaid with the VP1 protein for 90 minutes as a protein-antibody
308 complex that was prepared by preincubating VP1 with mouse monoclonal anti-
309 polyhistidine and biotinylated anti-mouse IgG antibodies (both from Sigma) at a ratio

310 of 4:2:1 (by weight) and diluted in the blocking solution to provide a final VP1
311 concentration of 150 µg/mL. Binding was detected with Alexa Fluor-647-labelled
312 streptavidin (Molecular Probes) at 1 µg/mL for 30 minutes. Apart from the protein-
313 antibody precomplexation step, which was performed on ice, all other steps were
314 carried out at ambient temperature. Microarray imaging and data analysis are
315 described in the Supplementary MIRAGE document (Supplemental Table S2).
316
317 NMR
318 All nuclear magnetic resonance (NMR) data were collected in a Bruker AVIII-600 MHz
319 spectrometer equipped with a TXI-z probe head at 298 K using 3 mm NMR tubes.
320 Samples were provided in 99 % D_2O-PBS containing 20 mM KH_2PO_4/K_2HPO_4 (pH 7.4)
321 and 150 mM NaCl. ^1H reference spectra of isolate oligosaccharides samples were
322 recorded using pure solutions of 4 mM of Forssman pentaose and 2 mM of 6'SL.
323 Subsequently, the ^1H oligosaccharide resonances of a sample containing 1 mM
324 Forssman pentaose and 1 mM 6'SL in the presence of 20 µM ShPyV VP1 were
325 collected under suppression of protein signals. From the same sample, the saturation
326 difference transfer (STD) NMR experiment was performed using previously published
327 pulse protocols[43]. NMR spectra were processed using TOPSPIN 4 (Bruker). The
328 carbohydrates used here were purchased from Biosynth, UK.
329
330 VP1 mutagenesis
331 Site-directed mutagenesis of VP1 was performed in samples of 20 µl containing
332 1x ReproFast amplification buffer (Genaxxon), 20 ng template vector DNA, 50 ng
333 primers, 2 mM dNTP, and 1 unit of ReproFast polymerase (Genaxxon). Initial
334 denaturation was performed at 95 °C for 2 minutes, followed by 18 cycles of 95 °C for
335 1 minute (denaturation), 55-70 °C for 1 minute (annealing), and 72 °C for 6 minutes

336 (elongation). Subsequently, DNA synthesis was completed at 72 °C for 10 minutes.
337 Parental template DNA was digested with 2 units of DpnI for two hours at 37 °C. The
338 PCR product was directly transformed into *E. coli* DH5α.
339
340 DNA primers for the site-directed mutagenesis of ShPyV VP1 were ordered from
341 ThermoFisher Scientific and are displayed from 5'- to 3' ends:
342
343 S95C_fwd: GCTGAACCAAGACATGACCTGCGATACCATCCTGATGTGGGAGG
344 S95C_rev: CCTCCCACATCAGGATGGTATCGCAGGTCATGTCTTGGTTCAGC
345
346 Fluorescence labeling
347 VP1 was expressed and purified as described above. Surficial cysteine residues were
348 reduced in degassed SEC buffer (20 mM HEPES pH 7.4, 150 mM NaCl) supplemented
349 with 80 mM DTT at 4 °C overnight. DTT was removed from the protein solution using
350 a PD-10 desalting column (Cytiva) with DTT-free SEC buffer. A 20-fold molar excess
351 of cysteine reactive AFDye 488 Maleimide (Fluoroprobes) was added to the protein
352 solutions (0.5 - 0.7 mg/ml) and the reaction was incubated in the dark at 4 °C for 16-
353 20 h. Excess dye was quenched and removed by desalting in SEC buffer
354 supplemented with 10 mM DTT. VP1-488 conjugate was eluted in the same buffer.
355 The degree of labeling was photometrically determined according to the
356 manufacturer's protocol.
357
358 Flow cytometry
359 Defibrinated blood from sheep was purchased from TCS Biosciences Ltd and
360 BioTrading Benelux B.V., and human reference erythrocytes from Immucor. Prior to
361 analysis, all blood was diluted to 0.25 % in 1x PBS + 2 % fetal calf serum (FC buffer).
362 For staining, labeled protein was added to samples of ~$1x10^6$ cells in a total volume of

363 100 µL. After incubation in the dark for 15 minutes, cells were precipitated by
364 centrifugation (1.5 min at 1900 rpm) and resuspended in fresh FC buffer. Unstained
365 reference samples were prepared accordingly without the addition of labeled protein.
366 For cell attachment inhibition assays, proteins were incubated with oligosaccharides
367 (Forssman antigen trisaccharide: Elicityl, France; Blood group A trisaccharide:
368 Biosynth, UK) for 5 minutes prior to their addition to the cellular samples (~$1x10^6$ cells
369 in 100 µL). To determine the concentration for half maximal cell attachment (effective
370 concentration, EC50), samples comprising serially diluted concentrations of labeled
371 protein, starting from 10 µg/mL for M1/87-FITC (Santa Cruz Biotechnology) and HPA
372 Alexa Fluor™ 488 conjugate (Invitrogen), and 5 µg/mL for VP1-488, were prepared as
373 described above. Here, the wash step was omitted to guarantee consistent protein
374 concentrations and to prevent agglutination. All samples were transferred to the flow
375 cytometer (LSR II, BD Biosciences) and fluorescent fractions were captured at 520 nM.
376 100,000 events were recorded for all samples except those with agglutination, in which
377 case 10,000-20,000 events were recorded. All data were analyzed using FACSdiva
378 software (BD Biosciences). EC50 values were calculated using R[72].
379
380 Surface plasmon resonance
381 Assessment of the kinetic binding parameters were performed via surface plasmon
382 resonance (SPR) in a Biacore X-100 system. Using an adapted protocol[73], BSA
383 decorated with Forssman pentaose (~76 kDa) as the ligand was prepared via reductive
384 amination (see Fig. S3). After rebuffering into 10 mM acetate buffer at pH 4,
385 approximately 100 response units (RU) of ligand were immobilized onto the CM5
386 sensor chip surface using amine coupling as described in the manufacturer's protocol
387 (Cytiva). For the reference flow cell, native BSA (Sigma-Aldrich) was immobilized
388 similarly. Purified VP1 (162 kDa) and native HPA (79 kDa, Sigma-Aldrich) buffered in

389 SPR assay buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 0.005 % Polysorbate 20)
390 were used as analytes in different concentrations between 0.2 nM and 20 µM. All
391 measurements were performed at 25 °C. During binding (120 s) and dissociation
392 (180 s), a flow rate of 30 µL/min was applied. After each cycle, the surfaces were
393 regenerated applying 10 mM glycine pH 1.5 for 120 s. Experimental curves were
394 interpreted using BIAevaluation software (Cytiva).
395
396 Crystallization and data collection
397 Purified ShPyV VP1 was concentrated to 3.5 mg/ml. Crystallization droplets comprised
398 1 µl protein and 1 µl reservoir solution (150 mM KSCN and 20 % (w/v) PEG 3,350).
399 The droplets were equilibrated against the 500 µl reservoir using the hanging drop
400 vapor diffusion method at 20 °C. For ligand derivatization, VP1 crystals were soaked
401 in crystallization solution supplemented with the respective oligosaccharides (10 mM
402 Forssman pentaose (F_P), 10 mM globo-*N*-tetraose, 5 mM F_P + 20 mM 3'SLN, 5 mM F_P
403 + 20 mM 6'SLN, 20 mM A antigen trisaccharide (all purchased from Biosynth, UK), and
404 20 mM Tf antigen (TCI Europe), at 20 °C for half an hour. Prior to flash freezing, crystals
405 were transferred to oligosaccharide soaking solutions supplemented with 20 % (v/v)
406 MPD for cryo-protection. X-ray diffraction experiments were performed at the Swiss
407 Light Source (SLS) beamline X06DA (PSI in Villigen, Switzerland). Data were collected
408 at a wavelength of 1 Å with an excitation time and angle of 0.1 s and 0.1° per image.
409
410 Structure determination and refinement
411 X-ray diffraction images were integrated and scaled using *XDS*[74]. For each data set,
412 2,000-3,000 reflections were reserved as a test set for structure validation. Phases
413 were determined by molecular replacement using the native sheep polyomavirus VP1
414 structure as search model (PDB 6Y61) in Phaser (CCP4)[75]. Phase refinement and

415 model building were performed until validation parameters converged using Phenix
416 and Refmac5, and Coot, respectively[76–78]. Oligosaccharides conformations are in
417 accordance with the restraints from the CCP4 libraries. All protein structure
418 representations were rendered in PyMOL (Schrödinger).

419

420 **Data availability**

421 Atomic coordinates of ShPyV VP1 in complex with carbohydrate ligands are deposited
422 in the RCSB protein data bank (PDB) with the accession codes 7B6S (Forssman
423 pentaose, F_P), 7B6T (globo-*N*-tetraose), 7B6U (F_P + 6'SLN), and 7B6V (F_P + 3'SLN).
424 Other data can be provided by the authors upon request.

425

426 **References**

427 1. B Siddiqui, S. H. A revised structure for the Forssman glycolipid hapten. *J Biol Chem* **246**,
428 5766–5769 (1971).
429 2. Fraser, B. A. & Mallette, M. F. Structure of forssman hapten glycosphingolipid from sheep
430 erythrocytes. *Immunochemistry* **11**, 581–593 (1974).
431 3. Haslam, D. B., Baenziger, J. U., Baenzigert, J. U. & Baenziger, J. U. Expression cloning of
432 Forssman glycolipid synthetase: A novel member of the histo-blood group ABO gene family.
433 *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10697–10702 (1996).
434 4. Xu, H., Storch, T., Yu, M., Elliott, S. P. & Haslam, D. B. Characterization of the human
435 Forssman synthetase gene. An evolving association between glycolipid synthesis and host-
436 microbial interactions. *J. Biol. Chem.* **274**, 29390–29398 (1999).
437 5. Yamamoto, M., Cid, E. & Yamamoto, F. Molecular genetic basis of the human Forssman
438 glycolipid antigen negativity. *Sci. Rep.* **2**, 975 (2012).
439 6. Svensson, L. *et al.* Forssman expression on human erythrocytes: Biochemical and genetic
440 evidence of a new histo-blood group system. *Blood* **121**, 1459–1468 (2013).
441 7. Storry, J. R. *et al.* International Society of Blood Transfusion Working Party on red cell
442 immunogenetics and blood group terminology: Cancun report (2012). *Vox Sang.* **107**, 90–96
443 (2014).
444 8. Hult, A. K. & Olsson, M. L. The FORS awakens: review of a blood group system reborn.
445 *Immunohematology* **33**, 64–72 (2017).
446 9. Ono, K. *et al.* Expression of Forssman antigen in human large intestine. *J. Histochem.*
447 *Cytochem.* **42**, 659–665 (1994).
448 10. Jesus, C. *et al.* Prevalence of antibodies to a new histo-blood system: the FORS system. *Blood*
449 *Transfus.* **16**, 178 (2018).
450 11. Hult, A. K., McSherry, E., Möller, M. & Olsson, M. L. GBGT1 is allelically diverse but
451 dispensable in humans and naturally occurring anti-FORS1 shows an ABO-restricted pattern.
452 *Transfusion* **58**, 2036–2045 (2018).
453 12. Kawanami, J. The appearance of Forssman hapten in human tumor. *J. Biochem.* **72**, 783–785
454 (1972).
455 13. Yoda, Y., Ishibashi, T. & Makita, A. Isolation, Characterization, and Biosynthesis of Forssman
456 Antigen in Human Lung and Lung Carcinoma. *J. Biochem.* **88**, 1887–1894 (1980).
457 14. Taniguchi, N., Yokosawa, N., Narita, M., Mitsuyama, T. & Makita, A. Expression of forssman
458 antigen synthesis and degradation in human lung cancer. *J. Natl. Cancer Inst.* **67**, 577–583
459 (1981).

15. Yokota, M., Warner, G. A. & Hakomori, S. Blood Group A-like Glycolipid and a Novel Forssman Antigen in the Hepatocarcinoma of a Blood Group O Individual. *Cancer Res.* **41**, (1981).
16. Mori, T., Sudo, T. & Kano, K. Expression of Heterophil Forssman Antigen on Cultured Malignant Cell Lines. *JNCI J. Natl. Cancer Inst.* **70**, 811–818 (1983).
17. Hakomori, S. *Tumor-Associated Carbohydrate Antigens*. *Annual review of immunology* vol. 2 103–126 (Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA , 1984).
18. Uemura, K., Hattori, H., Ono, K., Ogata, H. & Taketomi, T. Expression of Forssman glycolipid and blood group-related antigens A, Le(x) and Le(y) in human gastric cancer and in fetal tissues. *Jpn. J. Exp. Med.* **59**, 239–249 (1989).
19. Säljö, K., Thornell, A., Jin, C., Norlén, O. & Teneberg, S. Characterization of human medullary thyroid carcinoma glycosphingolipids identifies potential cancer markers. *Int. J. Mol. Sci.* **22**, 10463 (2021).
20. Ariga, T., Yoshida, T., Mimori, T. & Yu, R. K. Autoantibodies against Forssman glycolipids in Graves' disease and Hashimoto's thyroiditis. *Clin. Exp. Immunol.* **86**, 483–488 (1991).
21. Hakomori, S., Wang, S. M. & Young, W. W. Isoantigenic expression of Forssman glycolipid in human gastric and colonic mucosa: Its possible identity with 'A like antigen' in human cancer. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 3023–3027 (1977).
22. Stromberg, N. *et al.* Host-specificity of uropathogenic Escherichia coli depends on differences in binding specificity to Galα1-4Gal-containing isoreceptors. *EMBO J.* **9**, 2001–2010 (1990).
23. Stromberg, N., Nyholm, P. G., Pascher, I. & Normark, S. Saccharide orientation at the cell surface affects glycolipid receptor function. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 9340–9344 (1991).
24. Johanson, I., Lindstedt, R. & Svanborg, C. Roles of the pap- and prs-encoded adhesins in Escherichia coli adherence to human uroepithelial cells. *Infect. Immun.* **60**, 3416–3422 (1992).
25. Lanne, B. *et al.* Glycoconjugate receptors for P-fimbriated Escherichia coli in the mouse: An animal model of urinary tract infection. *J. Biol. Chem.* **270**, 9017–9025 (1995).
26. Elliott, S. P., Yu, M., Xu, H. & Haslam, D. B. Forssman synthetase expression results in diminished Shiga toxin susceptibility: A role for glycolipids in determining host-microbe interactions. *Infect. Immun.* **71**, 6543–6552 (2003).
27. Müthing, J. *et al.* Promiscuous Shiga toxin 2e and its intimate relationship to Forssman. *Glycobiology* **22**, 849–862 (2012).
28. Nasir, W., Nilsson, J., Olofsson, S., Bally, M. & Rydell, G. E. Parvovirus B19 VLP recognizes globoside in supported lipid bilayers. *Virology* **456**–**457**, 364–369 (2014).
29. Santos, L. *et al.* Characterization of sheep erythrocyte glycosphingolipids recognized by human anti-Forssman antibodies. *Glycobiology* (2020) doi:10.1093/glycob/cwaa032.
30. Buck, C. B. *et al.* The ancient evolutionary history of polyomaviruses. *PLoS Pathog.* **12**, 1005574 (2016).
31. Fried, H., Cahan, L. D. & Paulson, J. C. Polyoma virus recognizes specific sialyloligosaccharide receptors on host cells. *Virology* **109**, 188–192 (1981).
32. Tsai, B. *et al.* Gangliosides are receptors for murine polyoma virus and SV40. *EMBO J.* **22**, 4346–4355 (2003).
33. Smith, A. E., Lilie, H. & Helenius, A. Ganglioside-dependent cell attachment and endocytosis of murine polyomavirus-like particles. *FEBS Lett.* (2003) doi:10.1016/S0014-5793(03)01220-1.
34. Erickson, K. D., Garcea, R. L. & Tsai, B. Ganglioside GT1b is a putative host cell receptor for the Merkel cell polyomavirus. *J. Virol.* **83**, 10275–10279 (2009).
35. Gilbert, J. *et al.* Ganglioside GD1a Restores Infectibility to Mouse Cells Lacking Functional Receptors for Polyomavirus. *J. Virol.* **79**, 615–618 (2005).
36. Blaum, B. S. & Stehle, T. Sialic acids in nonenveloped virus infections. in *Advances in Carbohydrate Chemistry and Biochemistry* vol. 76 65–111 (Academic Press Inc., 2018).
37. Elphick, G. F. *et al.* The human polyomavirus, JCV, uses serotonin receptors to infect cells. *Science (80-.).* **306**, 1380–1383 (2004).
38. Assetta, B. *et al.* 5-HT2 receptors facilitate JC polyomavirus entry. *J. Virol.* **87**, 13490 (2013).
39. Schowalter, R. M., Pastrana, D. V. & Buck, C. B. Glycosaminoglycans and sialylated glycans sequentially facilitate merkel cell polyomavirus infectious entry. *PLoS Pathog.* **7**, 1002161 (2011).
40. Hurdiss, D. L., Frank, M., Snowden, J. S., Macdonald, A. & Ranson, N. A. The structure of an infectious human polyomavirus and its interactions with cellular receptors. *Structure* **26**, 839 (2018).
41. Geoghegan, E. M. *et al.* Infectious entry and neutralization of pathogenic JC polyomaviruses. *Cell Rep.* **21**, 1169–1179 (2017).
42. Assetta, B. *et al.* Genetic and functional dissection of the role of individual 5-HT2 receptors as

522 entry receptors for JC polyomavirus. *Cell Rep.* **27**, 1960-1966.e6 (2019).
523 43. Ströh, L. J. *et al.* Structural basis and evolution of glycan receptor specificities within the
524 polyomavirus family. *MBio* **11**, 1–21 (2020).
525 44. Palma, A. S., Feizi, T., Childs, R. A., Chai, W. & Liu, Y. The neoglycolipid (NGL)-based
526 oligosaccharide microarray system poised to decipher the meta-glycome. *Curr. Opin. Chem.*
527 *Biol.* **18**, 87–94 (2014).
528 45. Campanero-Rhodes, M. A. *et al.* N-glycolyl GM1 ganglioside as a receptor for simian virus 40.
529 *J. Virol.* **81**, 12846–12858 (2007).
530 46. Neu, U. *et al.* Structure-function analysis of the human JC polyomavirus establishes the LSTc
531 pentasaccharide as a functional receptor motif. *Cell Host Microbe* **8**, 309–319 (2010).
532 47. Neu, U. *et al.* A structure-guided mutation in the major capsid protein retargets BK
533 polyomavirus. *PLoS Pathog.* **9**, 1–13 (2013).
534 48. Khan, Z. M. *et al.* Crystallographic and glycan microarray analysis of human polyomavirus 9
535 VP1 identifies N-glycolyl neuraminic acid as a receptor candidate. *J. Virol.* **88**, 6100–6111
536 (2014).
537 49. Sorin, M. N. *et al.* Structural and functional analysis of natural capsid variants suggests sialic
538 acid-independent entry of BK polyomavirus. *CellReports* **42**, 112114 (2023).
539 50. Mayer, M. & Meyer, B. Group epitope mapping by saturation transfer difference NMR to identify
540 segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.* (2001)
541 doi:10.1021/ja0100120.
542 51. Blaum, B. S., Neu, U., Peters, T. & Stehle, T. Spin ballet for sweet encounters: Saturation-
543 transfer difference NMR and X-ray crystallography complement each other in the elucidation of
544 protein-glycan interactions. *Acta Crystallographica Section F: Structural Biology*
545 *Communications* vol. 74 451–462 (2018).
546 52. Grönberg, G., Nilsson, U., Bock, K. & Magnusson, G. Nuclear magnetic resonance and
547 conformational investigations of the pentasaccharide of the Forssman antigen and overlapping
548 di-, tri-, and tetra-saccharide sequences. *Carbohydr. Res.* **257**, 35–54 (1994).
549 53. Wu, A. M. & Sugii, S. Coding and classification of D-galactose, N-acetyl-D-galactosamine, and
550 β-D-Galp-[1→3(4)]-β-D-GlcpNAc, specificities of applied lectins. *Carbohydr. Res.* **213**, 127–
551 143 (1991).
552 54. Sanchez, J. F. *et al.* Biochemical and structural analysis of Helix pomatia agglutinin. A
553 hexameric lectin with a novel fold. **281**, 20171–20180 (2006).
554 55. Springer, T., Galfrè, G., Secher, D. S. & Milstein, C. Monoclonal xenogeneic antibodies to
555 murine cell surface antigens: identification of novel leukocyte differentiation antigens. *Eur. J.*
556 *Immunol.* **8**, 539–551 (1978).
557 56. Hauser, R., Fechner, G. & Brinkmann, B. A1 and A2 blood group substances: are there
558 structural differences? *Z. Rechtsmed.* **103**, 587–591 (1990).
559 57. Tanaka, N. & Leduc, E. H. A study of the cellular distribution of forssman antigen in various
560 species. *J. Immunol.* **77**, 198–212 (1956).
561 58. Koizumi, N., Hara, A., Uemura, K. & Taketomi, T. Glycosphingolipids in sheep liver, kidney,
562 and various blood cells. *Jpn. J. Exp. Med.* **58**, 21–31 (1988).
563 59. Slomiany, B. L. & Slomiany, A. Forssman glycolipid variants of dog gastric mucosa. Structure
564 of a branched ceramide octasaccharide. *Eur. J. Biochem.* **83**, 105–111 (1978).
565 60. Slomiany, B. L., Banas-Gruszka, Z., Zdebska, E. & Slomiany, A. Characterization of the
566 Forssman-active oligosaccharides from dog gastric mucus glycoprotein isolated with the use of
567 a monoclonal antibody. *J. Biol. Chem.* **257**, 9561–9565 (1982).
568 61. Ostrander, G. K., Levery, S. B., Hakomori, S. & Holmes, E. H. Isolation and characterization of
569 the major acidic glycosphingolipids from the liver of the English sole (Parophrys vetulus).
570 Presence of a novel ganglioside with a Forssman antigen determinant. *J. Biol. Chem.* **263**,
571 3103–3110 (1988).
572 62. Yamamoto, H. *et al.* Isolation and characterization of a novel Forssman active acidic
573 glycosphingolipid with branched isoglobo-, ganglio-, and neolacto-series hybrid sugar chains. *J.*
574 *Biochem.* **125**, 923–930 (1999).
575 63. Savage, A. V., Donoghue, C. M., D'Arcy, S. M., Koeleman, C. A. M. & van den Eijnden, D. H.
576 Structure determination of five sialylated trisaccharides with core types 1, 3 or 5 isolated from
577 bovine submaxillary mucin. *Eur. J. Biochem.* **192**, 427–432 (1990).
578 64. Hamelryck, T. W. *et al.* Carbohydrate binding, quaternary structure and a novel hydrophobic
579 binding site in two legume lectin oligomers from Dolichos biflorus. *J. Mol. Biol.* **286**, 1161–1177
580 (1999).
581 65. Lescar, J. *et al.* Structural basis for recognition of breast and colon cancer epitopes Tn antigen
582 and Forssman disaccharide by Helix pomatia lectin. *Glycobiology* **17**, 1077–1083 (2007).

583 66. Nagae, M. *et al.* Structural analysis of the human galectin-9 N-terminal carbohydrate
584 recognition domain reveals unexpected properties that differ from the mouse orthologue. *J.*
585 *Mol. Biol.* **375**, 119–135 (2008).
586 67. Kita, A. *et al.* Crystal structure of octocoral lectin SLL-2 complexed with Forssman antigen
587 tetrasaccharide. *Glycobiology* **27**, 696–700 (2017).
588 68. Bulai, T., Bratosin, D., Pons, A., Montreuil, J. & Zanetta, J. P. Diversity of the human
589 erythrocyte membrane sialic acids in relation with blood groups. *FEBS Lett.* **534**, 185–189
590 (2003).
591 69. Cid, E., Yamamoto, M. & Yamamoto, F. Mixed-Up Sugars: Glycosyltransferase Cross-
592 Reactivity in Cancerous Tissues and Their Therapeutic Targeting. *ChemBioChem* **23**,
593 e202100460 (2022).
594 70. Liu, Y. *et al.* Neoglycolipid-based oligosaccharide microarray system: Preparation of ngls and
595 their noncovalent immobilization on nitrocellulose-coated glass slides for microarray analyses.
596 *Methods Mol. Biol.* **808**, 117–136 (2012).
597 71. Liu, Y. *et al.* The minimum information required for a glycomics experiment (MIRAGE) project:
598 improving the standards for reporting glycan microarray-based data. *Glycobiology* **27**, 280–284
599 (2017).
600 72. Ritz, C., Baty, F., Streibig, J. C. & Gerhard, D. Dose-response analysis using R. *PLoS One* **10**,
601 e0146021 (2015).
602 73. Gildersleeve, J. C., Oyelaran, O., Simpson, J. T. & Allred, B. Improved procedure for direct
603 coupling of carbohydrates to proteins via reductive amination. *Bioconjug. Chem.* **19**, 1485–
604 1490 (2008).
605 74. Kabsch, W. *et al.* XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2010).
606 75. McCoy, A. J. *et al. Phaser* crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
607 76. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and
608 electrons: Recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* (2019)
609 doi:10.1107/S2059798319011471.
610 77. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures.
611 *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2011) doi:10.1107/S0907444911001314.
612 78. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr.*
613 *Sect. D Biol. Crystallogr.* (2004) doi:10.1107/S0907444904019158.
614
615
616
617 **Acknowledgements**

618 The glycan microarray studies were performed in the Carbohydrate Microarray Facility
619 at the Imperial College Glycosciences Laboratory, which is supported by Wellcome
620 Trust biomedical resource grants (099197/Z/12/Z, 108430/Z/15/Z, and 218304/Z/19/Z)
621 and in part by the March of Dimes Prematurity research center grant (22-FY18-82).
622 The sequence defined glycan microarrays contain many saccharides provided by
623 collaborators whom we thank, as well as members of the Glycosciences Laboratory
624 for their contribution in the establishment of the NGL-based microarray system. We
625 also thank the German Research Foundation for funding (FOR2327 – VIROCARB,
626 Project no. 269564371). Crystallographic data collection was performed at the
627 macromolecular beam lines X06SA and X06DA of the Swiss Light Source, access to

which was generously granted by the Paul Scherrer Institute in Villigen, Switzerland. We thank the Max-Planck Institute for Biology in Tübingen for the access to the NMR facility and Dr. Vincent Truffault for his assistance in NMR data acquisition. We also thank Prof. Barbara Mulloy (Imperial College London) for critical review of the manuscript. We thank the lab of Prof. Schulze-Osthoff at the University of Tübingen for the access to the flow cytometry appliances and especially Dr. Anja Schmitt for her help with the experiments.

Author information

Authors and affiliations

Interfaculty Institute for Biochemistry, University of Tübingen, Tübingen, Germany

Nils H. Rustmeier, Alexander Herrmann, Joshua Müller, Thilo Stehle

Glycosciences Laboratory, Faculty of Medicine, Imperial College London, London, UK

Lisete S. Silva, Antonio Di Maio, Ten Feizi, Yan Liu

Contributions

N.R. and T.S. conceptualized the study. A.H., J.M., and N.R. prepared and crystallized the protein. L.S.S and A.D.M. performed the glycan array screening experiments. A.D.M., Y.L. and T.F. performed the glycan array analyses and interpreted the glycan interaction data. N.R. performed and analyzed the NMR, SPR and flow cytometry experiments. A.H., J.M., and N.R. collected the crystallographic data and solved the structures. N.R. prepared the first draft of the manuscript and all contributed to the final revision. T.S. provided funding. All authors approved publication of the manuscript.

Ethics declarations

The authors declare no competing interests. Animal blood was acquired from vendors guaranteeing animal wellbeing. Human erythrocytes were derived from anonymous donors consenting research purposes.

Table 1 Carbohydrate antigens cited in this report.

Class	Name	Symbolic structure	Sequence
Globoseries glycosphingolipids	LacCer	β 4 β Cer	Galβ1-4Glcβ-Cer
	Gb3Cer	α 4 β 4 β Cer	Galα1-4Galβ1-4Glcβ-Cer
	Gb4Cer (P antigen)	β 3 α 4 β 4 β Cer	GalNAcβ1-3Galα1-4Galβ1-4Glcβ-Cer
	Forssman pentaosylceramide	α 3 β 3 α 4 β 4 β Cer	GalNAcα1-3GalNAcβ1-3Galα1-4Galβ1-4Glcβ-Cer
Histo blood group antigens	H antigen	α 2 β 3 R	Fucα1-2Galβ1-R
	A antigen	α 2 α 3 β 3 R	GalNAcα1-3(Fucα1-2)Galβ1-R
	B antigen	α 2 α 3 β 3 R	Galα1-3(Fucα1-2)Galβ1-R
O-Glycans	Tn antigen	α Ser/Thr	GalNAcα-Ser/Thr
	TF antigen	β 3 α Ser/Thr	Galβ1-3GalNAcα-Ser/Thr

Lac - lactose, Cer - ceramide, Gb3 - globotriose, Gb4 - globotetraose, R - glycan backbone linked to protein or lipid carriers

- Gal - GalNAc - Glc - GlcNAc - Fuc

657

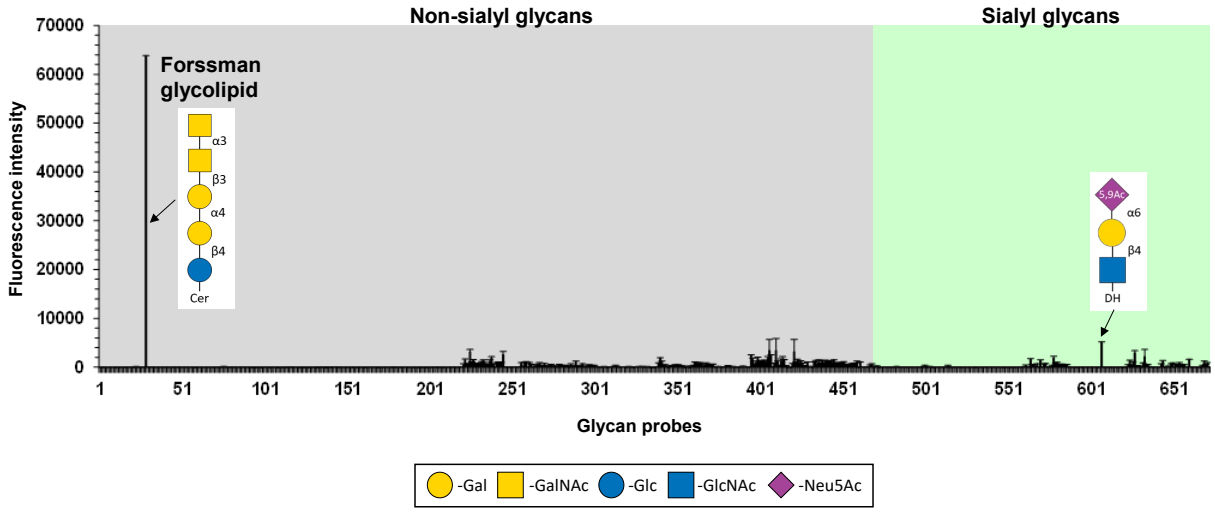


Figure 1 Glycan microarray screening analysis of ShPyV VP1 reveals selective engagement of Forssman glycolipid. The results are the means of fluorescence intensities of duplicate spots, printed at 5 fmol per spot. The error bars represent half of the difference between the two values. In the glycan array the 672 lipid-linked probes are grouped into non-sialylated and sialylated glycans as annotated by the colored panels. The list of glycan probes, their sequences and binding scores are in Table S1.

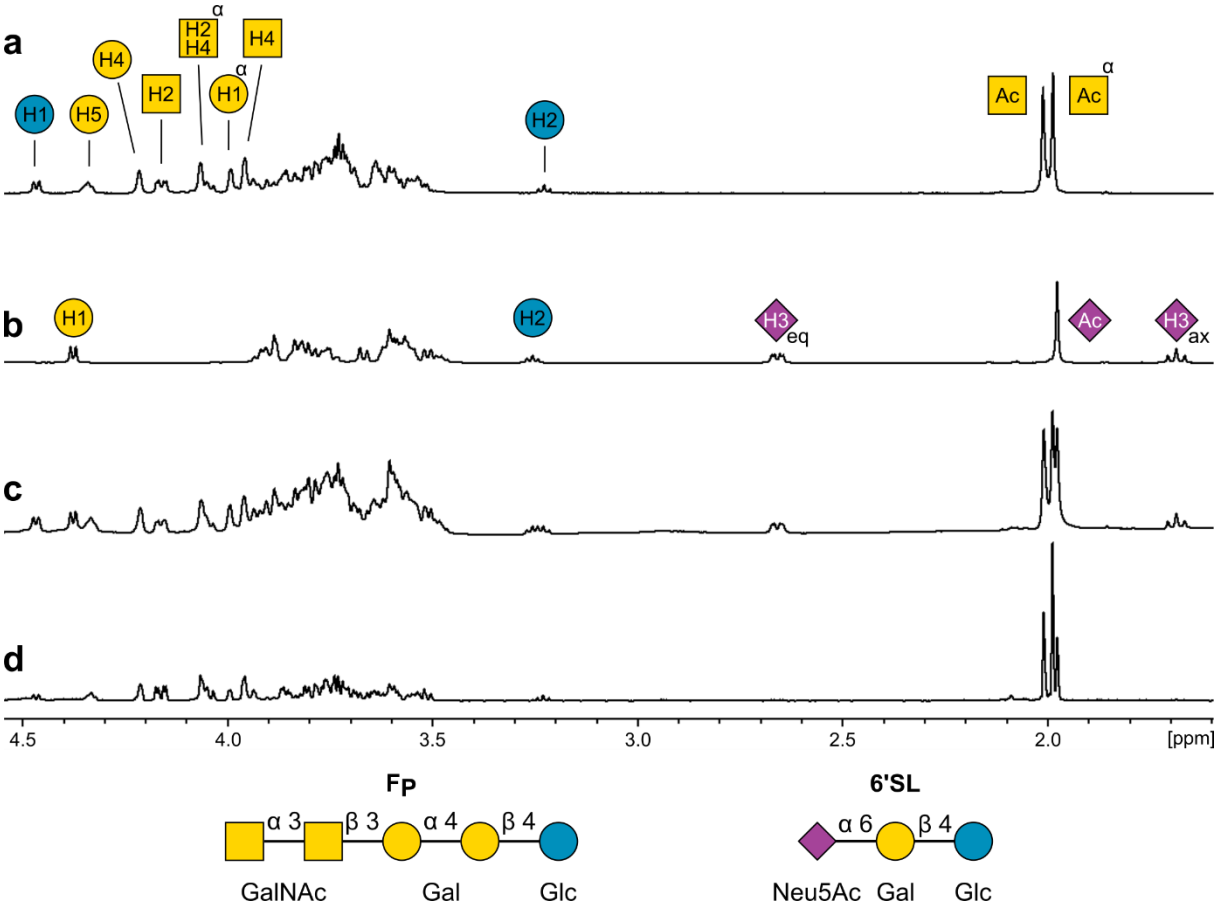
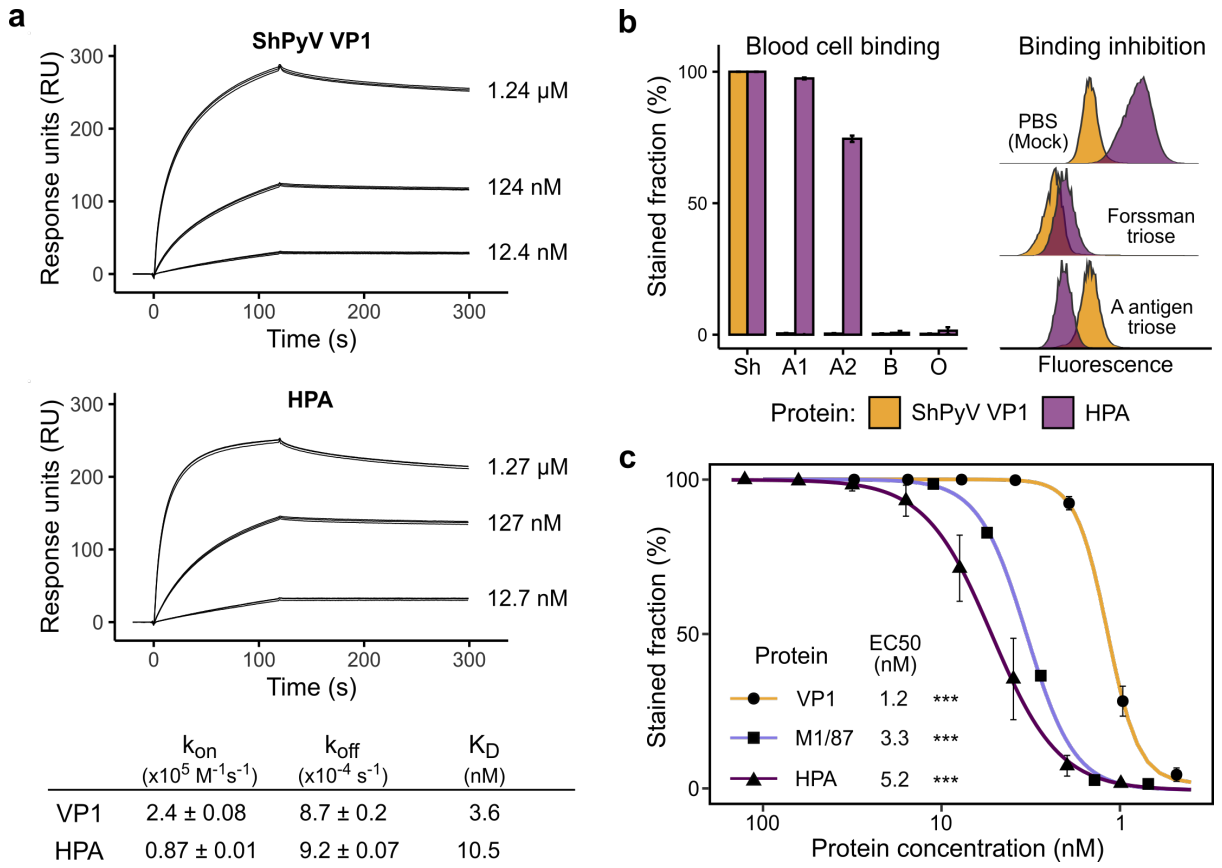
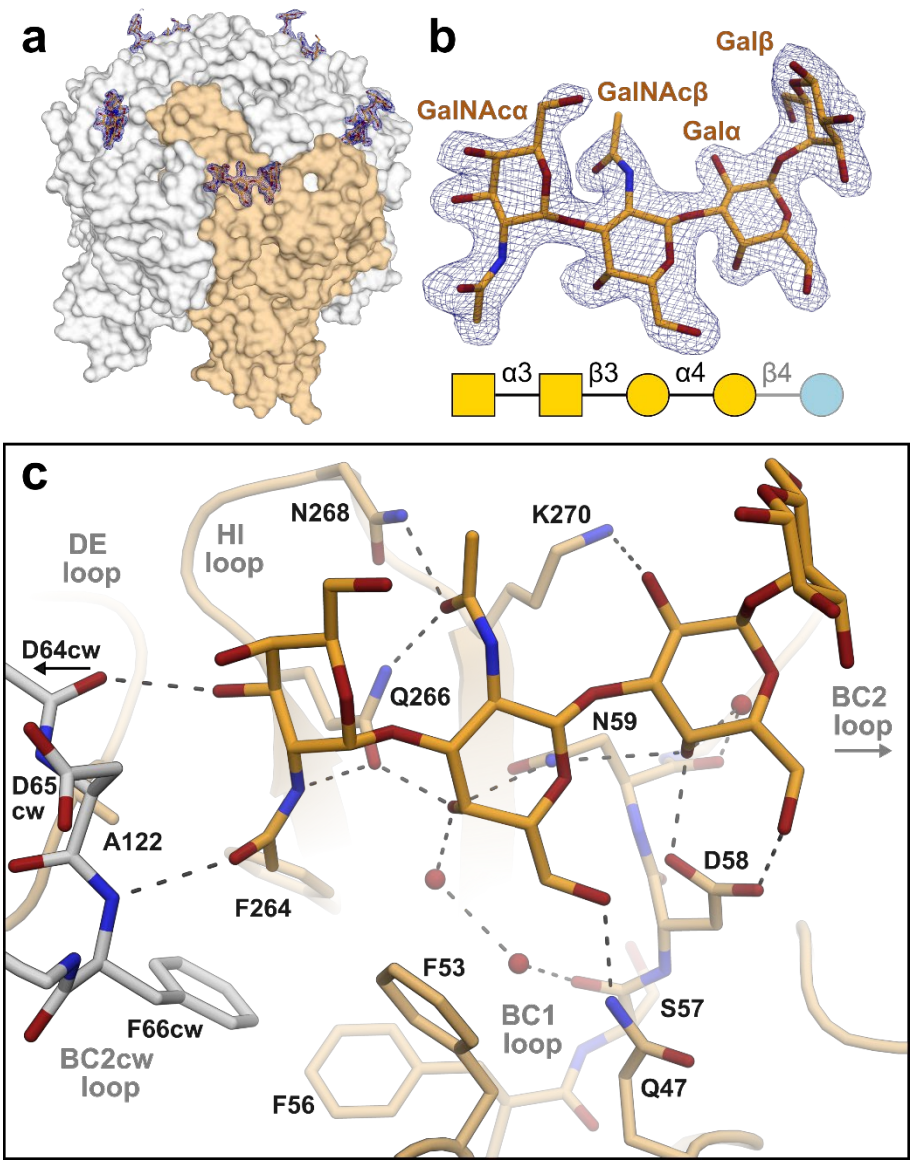


Figure 2 Assessment of the sheep polyomavirus VP1 – Forssman pentaose interaction in solution by STD NMR. ^{1}H spectra of **a** Forssman pentaose (F_P) and **b** 6'sialyllactose (6'SL) in their isolated forms. Proton resonances are labeled according to the carbon positions in the monosaccharides. **c** 1H spectrum of the test sample comprising 20 µM ShPyV VP1, 1 mM F_P, and 1 mM 6'SL collected under protein suppression. **d** The STD NMR spectrum recorded from the test sample.



	k_{on} ($x10^5$ $M^{-1}s^{-1}$)	k_{off} ($x10^{-4}$ s^{-1})	K_D (nM)
VP1	2.4 ± 0.08	8.7 ± 0.2	3.6
HPA	0.87 ± 0.01	9.2 ± 0.07	10.5

Figure 3 Flow cytometry and surface plasmon resonance analyses of Forssman antigen engagement. a Determination of the binding kinetics towards a Forssman-decorated surface using ShPyV VP1 and HPA. Response curves are shown as triplicates for each protein concentration. Kinetic parameters and the dissociation constant are provided in the table below. **b** Absolute cell attachment of ShPyV VP1 and *Helix pomatia* agglutinin (HPA) to Forssman antigen-positive red blood cells (RBCs) from sheep (Sh) and human erythrocytes of the blood groups A1, A2, B, and O (left). Effects of Forssman and blood group A antigen trioses on ShPyV VP1 and HPA attachment to sheep RBCs. Results are shown as concatenated and normalized histograms (right). **c** EC50 determination using ShPyV VP1, HPA, and anti-Forssman IgM M1/87 to sheep blood cells. Significance is indicated with *** for $p < 0.001$.



687

688 **Figure 4 Forssman pentaose binds to the external surface of ShPyV VP1 with**
689 **numerous contacts**. **a** The crystal structure of a single VP1 pentamer is shown with
690 five molecules of Forssman pentaose (F_P) associated to its outer surface. The bias-
691 reduced Fo-Fc omit electron density maps of F_P are displayed as blue meshes around
692 the oligosaccharides, with a contour level of 2.5 σ and a radius of 1.6 Å. The individual
693 F_P binding sites are each constituted by adjoining protein chains, highlighted by the
694 orange coloring of one VP1 monomer. **b** Magnified view of a bound F_P molecule
695 (above) with its symbol structure (below). The glucose of F_P, which is not visible in the
696 crystal structure, is grayed out in the symbol representation. **c** Molecular interactions
697 between VP1 and F_P are displayed with the hydrogen bonds indicated as dashed lines.

698 VP1 coloring corresponds to subfigure **a**. Water molecules are shown as spheres.
699 Throughout the figure, oxygen and nitrogen atoms are colored in red and blue,
700 respectively.
701

702 **Table 2** Intermolecular contacts between ShPyV VP1 and Forssman pentaose.

Monosaccharide		VP1 contacts	
		Van der Waals	
αGalNAc	Methyl	F53, F56, F66cw, A122, F264	
		Hydrogen-bonds	Distance (Å)
αGalNAc	O3	*D64cw*	2.74
	Amide N	Q266	2.70
	Acetyl O	*F66cw*	2.90
βGalNAc	O4	H_2O-H_2O-*S57*	2.95-2.90-2.63
	O4	N59	3.06
	O4	Q266	2.55
	O6	Q47	2.95
	Acetyl O	Q266	2.96
	Acetyl O	N268	2.78
αGal	O2	K270	3.31
	O4	D58	2.67
	O4	N59	3.45
	O4	H_2O-*N59*	2.61-2.83
	O6	D58	2.54

Residues using their backbone atoms to establish contacts are printed in italic. Distance values are derived from a single binding site and may alter between chains.

703