Glycobiology, 2016, vol. 26, no. 11, 1221–1233
doi: 10.1093/glycob/cww042
Advance Access Publication Date: 1 May 2016
Original Article

OXFORD

Glycan Recognition

Novel aspects of sialoglycan recognition by the Siglec-like domains of streptococcal SRR glycoproteins

Barbara A Bensing[1,2], Zahra Khedri[3], Lingquan Deng[3,4], Hai Yu[5], Akraporn Prakobphol[6], Susan J Fisher[6], Xi Chen[5], Tina M Iverson[7], Ajit Varki[3], and Paul M Sullam[2]

[2]Department of Medicine, The San Francisco Veterans Affairs Medical Center, and the University of California, San Francisco, San Francisco, CA 94121, USA, [3]The Glycobiology Research and Training Center, and the Department of Cellular and Molecular Medicine, University of California, San Diego, San Diego, CA 92093, USA, [5]Department of Chemistry, University of California, Davis, Davis, CA 95616, USA, [6]Department of Obstetrics, Gynecology and Reproductive Sciences, The University of California, San Francisco, San Francisco, CA 94143, USA, and [7]Department of Pharmacology, Vanderbilt University, Nashville, TN 27232, USA

[1]To whom correspondence should be addressed: Department of Medicine, VA Medical Center (111W2), 4150 Clement Street, San Francisco, CA 94121, USA. Tel: +1-415-221-4810; Fax: +1-415-750-6959; e-mail: barbara.bensing@ucsf.edu

[4]Present address: Department of Pathology, School of Medicine, Johns Hopkins University, Baltimore, MD 21287, USA.

Received 1 February 2016; Revised 23 March 2016; Accepted 25 March 2016

Abstract

Serine-rich repeat glycoproteins are adhesins expressed by commensal and pathogenic Gram-positive bacteria. A subset of these adhesins, expressed by oral streptococci, binds sialylated glycans decorating human salivary mucin MG2/MUC7, and platelet glycoprotein GPIb. Specific sialoglycan targets were previously identified for the ligand-binding regions (BRs) of GspB and Hsa, two serine-rich repeat glycoproteins expressed by *Streptococcus gordonii*. While GspB selectively binds sialyl-T antigen, Hsa displays broader specificity. Here we examine the binding properties of four additional BRs from *Streptococcus sanguinis* or *Streptococcus mitis* and characterize the molecular determinants of ligand selectivity and affinity. Each BR has two domains that are essential for sialoglycan binding by GspB. One domain is structurally similar to the glycan-binding module of mammalian Siglecs (sialic acid-binding immunoglobulin-like lectins), including an arginine residue that is critical for glycan recognition, and that resides within a novel, conserved YTRY motif. Despite low sequence similarity to GspB, one of the BRs selectively binds sialyl-T antigen. Although the other three BRs are highly similar to Hsa, each displayed a unique ligand repertoire, including differential recognition of sialyl Lewis antigens and sulfated glycans. These differences in glycan selectivity were closely associated with differential binding to salivary and platelet glycoproteins. Specificity of sialoglycan adherence is likely an evolving trait that may influence the propensity of streptococci expressing Siglec-like adhesins to cause infective endocarditis.

Key words: endocarditis, *MUC7*, platelet GPIb, sialyl-T antigen, Siglec

Introduction

The serine-rich repeat (SRR) glycoproteins are a family of adhesins expressed by both commensal and pathogenic Gram-positive bacteria. These surface structures have a common domain organization, which includes an atypical N-terminal signal peptide, a short serine-rich region, a ligand-binding domain, a long serine-rich region and a C-terminal cell wall anchoring domain (Figure 1). The serine-rich domains undergo extensive cytoplasmic glycosylation, and the glycosylated adhesin is then transferred to the bacterial cell surface via a dedicated transporter, the accessory Sec system (reviewed in Zhou and Wu 2009; Bensing et al. 2014). Despite the conserved domain organization of SRR glycoproteins, the ligand-binding regions (BRs) are highly divergent in sequence and in types of ligands to which they adhere (Wu et al. 1998; Takahashi et al. 2002; Bensing et al. 2004; Siboo et al. 2005; Samen et al. 2007; Shivshankar et al. 2009; Seo et al. 2012). This divergence is reflected in the high-resolution crystal structures that have been determined for five SRR glycoprotein BRs to date (Figure 1).

The variability of BR amino acid sequences is enormous, with some showing high conservation from strain to strain, and others showing extreme divergence. For example, the PsrP homologs expressed by different strains of *Streptococcus pneumoniae* vary by only a few amino acids. In contrast, the BRs of the SRR glycoproteins from oral streptococcal species such as *Streptococcus gordonii*, *Streptococcus sangunis*, *Streptococcus oralis*, and *Streptococcus mitis* diverge quite rapidly. The BRs range from ~200 amino acids (as seen in Hsa from *S. gordonii* DL1, and SrpA from *S. sanguinis* SK36) to >700 amino acids (as in MonX from *S. mitis* B6; Denapaite et al. 2010). Moreover, some of these BRs include modules that have no counterpart in other adhesins (Pyburn et al. 2011).

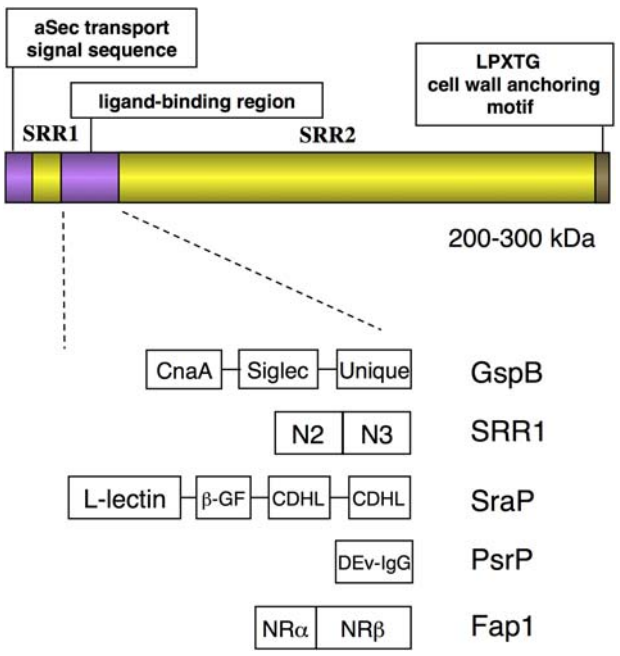

Fig. 1. Diagram of SRR glycoproteins and structural variability of the BRs. The upper portion of the figure indicates the conserved domain organization for the family of adhesins and the lower portion indicates the structural diversity of the BRs. The *S. gordonii* GspB domains are described in the text. The *Streptococcus agalactiae* SRR1 has two MSCRAMM folds, designated N2 and N3 (Seo et al. 2013). The *Staphylococcus aureus* SraP has a legume lectin-like module (L-lectin) that adjoins a ubiquitin-like β-grasp fold and a pair of cadherin-like modules (Yang et al. 2014). The *S. pneumoniae* PsrP has a DE-variant Ig fold (Schulte et al. 2014). The *Streptococcus parasanguinis* Fap1 has a pair of novel domains designated non-repeat α and non-repeat β (Ramboarina et al. 2010). This figure is available in black and white in print and in color at *Glycobiology* online.

A subset of the SRR glycoproteins expressed by oral streptococci is known to bind sialoglycans, and numerous reports have indicated that they are selective for α2–3-linked sialic acids on the human salivary mucin glycoprotein 2 (MG2, encoded by the *MUC7* gene and hereafter referred to as MG2/MUC7), and the human platelet membrane glycoprotein GPIb (the receptor for von Willebrand factor) (Takahashi et al. 1997, 2004; Kerrigan et al. 2002; Plummer et al. 2005; Takamatsu et al. 2005, 2006; Plummer and Douglas 2006). Whereas binding to MG2/MUC7 may be important for oral colonization, adherence to platelets can contribute to the pathogenesis of endocardial infections. Indeed, deletion of the SRR glycoprotein gene from *S. gordonii* strains M99 and DL1 (*gspB* and *hsa*, respectively), or a single amino acid substitution in GspB that abolishes sialoglycan binding, significantly impairs both the attachment of *S. gordonii* to platelets and the development of endocarditis in animal models of infection (Takahashi et al. 2006; Xiong et al. 2008). Thus, although the adhesins may have evolved to facilitate attachment to MG2/MUC7 on the salivary pellicle, serendipitous binding to the same or very similar glycan structures on GPIb can render the oral streptococci more virulent.

Previous studies have identified preferred sialoglycan ligands for several of the SRR glycoproteins (Prakobphol et al. 1999; Takamatsu et al. 2005; Deng et al. 2014). GspB and three closely related homologues demonstrate selective binding to Neu5Acα2–3Galβ1–3GalNAc (sialyl-T antigen, or sTa). Two other SRR adhesins (Hsa and SrpA) show much broader specificity. While Hsa avidly binds many trisaccharides carrying a terminal α2–3-linked sialic acid, a preferred sialoglycan ligand for SrpA has not yet been clearly defined. The number of unique BR sequences related to GspB, Hsa and SrpA continues to grow as additional bacterial genomes are sequenced. However, the extent to which the divergent BR sequences correspond to differences in sialoglycan selectivity and affinity is unknown.

We recently determined the high-resolution crystal structure of the $GspB_{BR}$ (Pyburn et al. 2011), which revealed three distinct domains, including a V-set Ig fold resembling that of mammalian sialic acid-binding immunoglobulin-like lectins (Siglecs). The Siglec domain includes residues that are important for binding to sTa and to platelet GPIb, and is flanked by the CnaA and Unique domains (Figure 1). In this report, we examine the contributions of these latter two domains of the $GspB_{BR}$ to sialoglycan binding. We also assess the binding properties of four additional Siglec-like BRs, and assess some of the molecular determinants of ligand selectivity and affinity, showing evidence for rapid diversification of recognition specificity.

Results

The combined Siglec and Unique domains are necessary and sufficient for sialoglycan binding

The BR of GspB consists of three distinct domains: the CnaA, Siglec and Unique (Pyburn et al. 2011). Three residues within the Siglec domain (Y443, R484 and Y485) are critically important for binding to sTa and human platelets (Pyburn et al. 2011), thus demonstrating that this domain is clearly required for sialoglycan binding. However, the contributions of the CnaA and Unique domains were not previously addressed. We therefore deleted each domain from $GspB_{BR}$, and then examined the effect on ligand binding and selectivity.

Whereas deletion of the Unique domain nearly abolished binding to sTa, deletion of the CnaA domain had little effect on this interaction (Figure 2A). In particular, deletion of the CnaA domain did not alter binding selectivity, when assessed on a custom array of glycans that includes 42 sialylated compounds along with seven non-sialylated controls (Table I and Figure 2B). This selectivity includes a preference for α- vs. β-linked sTa (array compounds **7** and **19**, respectively) that was observed with $GspB_{BR}$. Moreover, the combined Siglec and Unique domains could discriminate between the Neu5Ac and Neu5Gc forms of sialic acid, as indicated by the comparatively low binding to array compounds **8** and **20**, which are Neu5Gc versions of sTa. Thus, the Siglec and Unique domains are both necessary and sufficient for binding to sTa. Moreover, the narrow selectivity of GspB compared with Hsa and SrpA is not due to the CnaA domain, but instead resides within these two other domains.

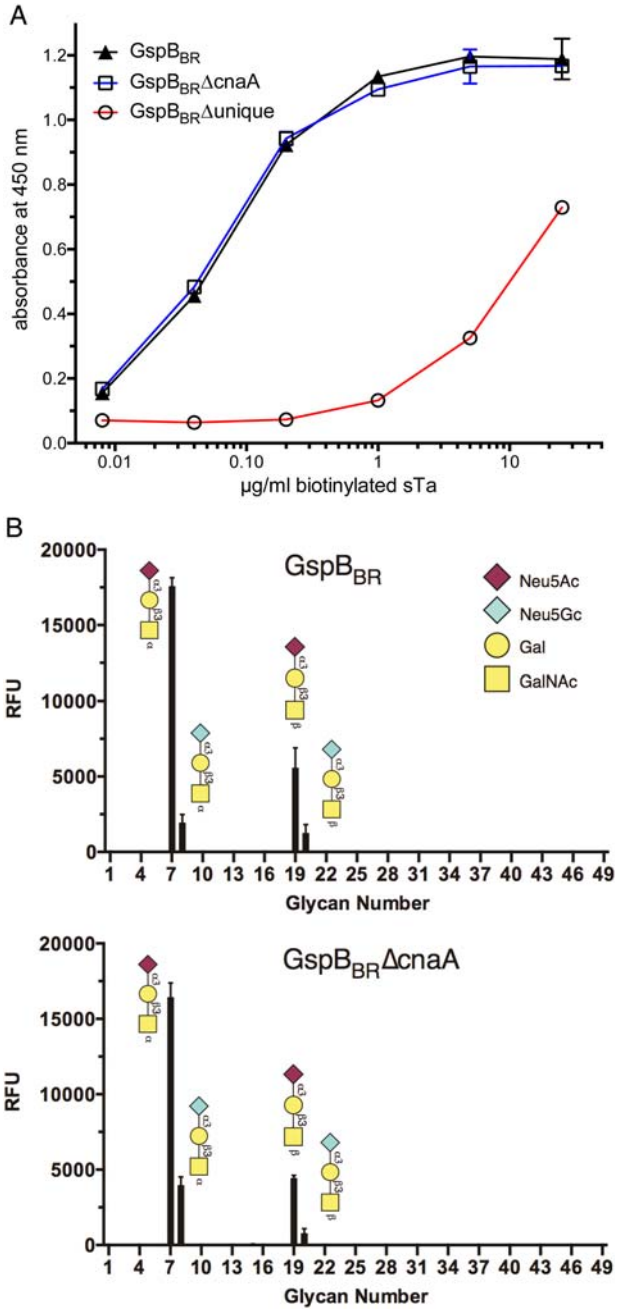

Fig. 2. Binding by GspB ΔcnaA and Δunique domain variants. (**A**) GST-BRs (500 nM) were immobilized in 96-well plates, and biotinylated sTa was supplied at the indicated concentrations. Binding is reported as the mean ± standard deviation, with $n = 3$. (**B**) Binding of GST-$GspB_{BR}$ (upper) and the ΔcnaA domain variant (lower) to an array of immobilized glycans, which includes compounds present on salivary and platelet glycoproteins ($n = 4$). This figure is available in black and white in print and in color at *Glycobiology* online.

Table I. Composition of glycans in the custom array

ID	Structure[a]
# 01	Neu5Acα2–6GalNAcαR1
# 02	Neu5Gcα2–6GalNAcαR1
# 03	Neu5Acα2–3Galβ1–4GlcNAcβR1
# 04	Neu5Gcα2–3Galβ1–4GlcNAcβR1
# 05	Neu5Acα2–3Galβ1–3GlcNAcβR1
# 06	Neu5Gcα2–3Galβ1–3GlcNAcβR1
# 07	Neu5Acα2–3Galβ1–3GalNAcαR1
# 08	Neu5Gcα2–3Galβ1–3GalNAcαR1
# 09	Neu5Acα2–6Galβ1–4GlcNAcβR1
# 10	Neu5Gcα2–6Galβ1–4GlcNAcβR1
# 11	Neu5Acα2–6Galβ1–4GlcβR1
# 12	Neu5Gcα2–6Galβ1–4GlcβR1
# 13	Neu5Acα2–3Galβ1–4GlcβR1
# 14	Neu5Gcα2–3Galβ1–4GlcβR1
# 15	Neu5Acα2–3GalβR1
# 16	Neu5Gcα2–3GalβR
# 17	Neu5Acα2–6GalβR1
# 18	Neu5Gcα2–6GalβR1
# 19	Neu5Acα2–3Galβ1–3GalNAcβR1
# 20	Neu5Gcα2–3Galβ1–3GalNAcβR1
# 21	Neu5Acα2–8Neu5Acα2–3Galβ1–4GlcβR1
# 22	Neu5Acα2–8Neu5Acα2–8Neu5Acα2–3Galβ1–4GlcβR1
# 23	Galβ1–4GlcβR1
# 24	Galβ1–4GlcNAcβR1
# 25	GalNAcαR1
# 26	Galβ1–3GalNAcβR1
# 27	Galβ1–3GalNAcαR1
# 28	Galβ1–3GlcNAcβR1
# 29	Galβ1–4GlcNAc6SβR1
# 30	Neu5Acα2–3Galβ1–4(Fucα1–3)GlcNAcβR1
# 31	Neu5Gcα2–3Galβ1–4(Fucα1–3)GlcNAcβR1
# 32	Neu5Acα2–3Galβ1–4(Fucα1–3)GlcNAc6SβR1
# 33	Neu5Gcα2–3Galβ1–4(Fucα1–3)GlcNAc6SβR1
# 34	Neu5Acα2–3Galβ1–3GlcNAcβ1–3Galβ1–4GlcβR1
# 35	Neu5Gcα2–3Galβ1–3GlcNAcβ1–3Galβ1–4GlcβR1
# 36	Neu5Acα2–3Galβ1–4GlcNAc6SβR1
# 37	Neu5Gcα2–3Galβ1–4GlcNAc6SβR1
# 38	Neu5Acα2–8Neu5Acα2–3Galβ1–4GlcβR2
# 39	Neu5Acα2–8Neu5Acα2–8Neu5Acα2–3Galβ1–4GlcβR2
# 40	Neu5Acα2–6(Neu5Acα2–3)Galβ1–4GlcβR1
# 41	Neu5Acα2–6(Neu5Gcα2–3)Galβ1–4GlcβR1
# 42	Neu5Acα2–6(Kdnα2–3)Galβ1–4GlcβR1
# 43	Neu5Gcα2–8Neu5Acα2–3Galβ1–4GlcβR1
# 44	Kdnα2–8Neu5Acα2–3Galβ1–4GlcβR1
# 45	Neu5Acα2–8Neu5Gcα2–3Galβ1–4GlcβR1
# 46	Neu5Acα2–8Neu5Gcα2–6Galβ1–4GlcβR1
# 47	Kdnα2–8Neu5Gcα2–3Galβ1–4GlcβR1
# 48	Neu5Gcα2–8Neu5Gcα2–3Galβ1–4GlcβR1
# 49	Neu5Acα2–8Neu5Acα2–6Galβ1–4GlcβR1

[a]R1 = $O(CH_2)_3NH_2$; R2 = $O(CH_2)_3NHCOCH_2(OCH_2CH_2)_6NH_2$.

A Siglec-like BR from *Streptococcus mitis* is selective for sTa, despite low sequence similarity to GspB

Although Siglec-like BRs have only been characterized in strains of *S. gordonii* and *S. sanguinis*, we recently identified genes encoding an SRR glycoprotein and an accessory Sec system in *Streptococcus mitis* SF100, an endocarditis-associated strain (Table II and unpublished results). The primary amino acid sequence of the $SF100_{BR}$ is only distantly related to the Siglec-like BRs, with <30% similarity to either GspB or Hsa, and with multiple gaps in the sequence alignments. However, the predicted secondary structure of 221 residues from the central portion of the putative $SF100_{BR}$ could be modeled onto the structure of the Siglec and Unique domains of the GspB PDB entry 3QC5 with >98% confidence, using the Phyre 2 program (Kelley et al. 2015) (Figure 3A). The $SF100_{BR}$ also includes two modules that do not resemble any reported structures. Of note, alignment of the amino acid sequences revealed that two of the three residues important for binding by GspB (R484 and Y485) were conserved in the Siglec-like domain of the $SF100_{BR}$ (Figure 3B). We therefore hypothesized

Table II. BR source information

BR	Codons[a]	Accession number	Species	Strain	Source
$GspB_{BR}$	233–617	AAL13053	*S. gordonii*	M99	IE patient (blood)
Hsa_{BR}	219–454	ABV1039	*S. gordonii*	DL1	
$NCTC10712_{BR}$	216–452	WP_045635027	*S. mitis*	NCTC10712	oral cavity
$SK678_{BR}$	226–466	EGC27373	*S. sanguinis*	SK678	oral cavity
$SK1_{BR}$[b]	242–657	EGF07837	*S. sanguinis*	SK1	human heart
$SF100_{BR}$	220–726	KU519294	*S. mitis*	SF100	IE patient (blood)

[a]The region of the SRR glycoprotein fused to GST for binding studies. The BR domain was presumed to reside between the two serine-rich domains.

[b]The $SK1_{BR}$, including the duplication of the Siglec and Unique modules, is identical to that of SK1058 and 97% similar (96%identity) to that of SK1087, two *S. sanguinis* strains isolated from human blood.

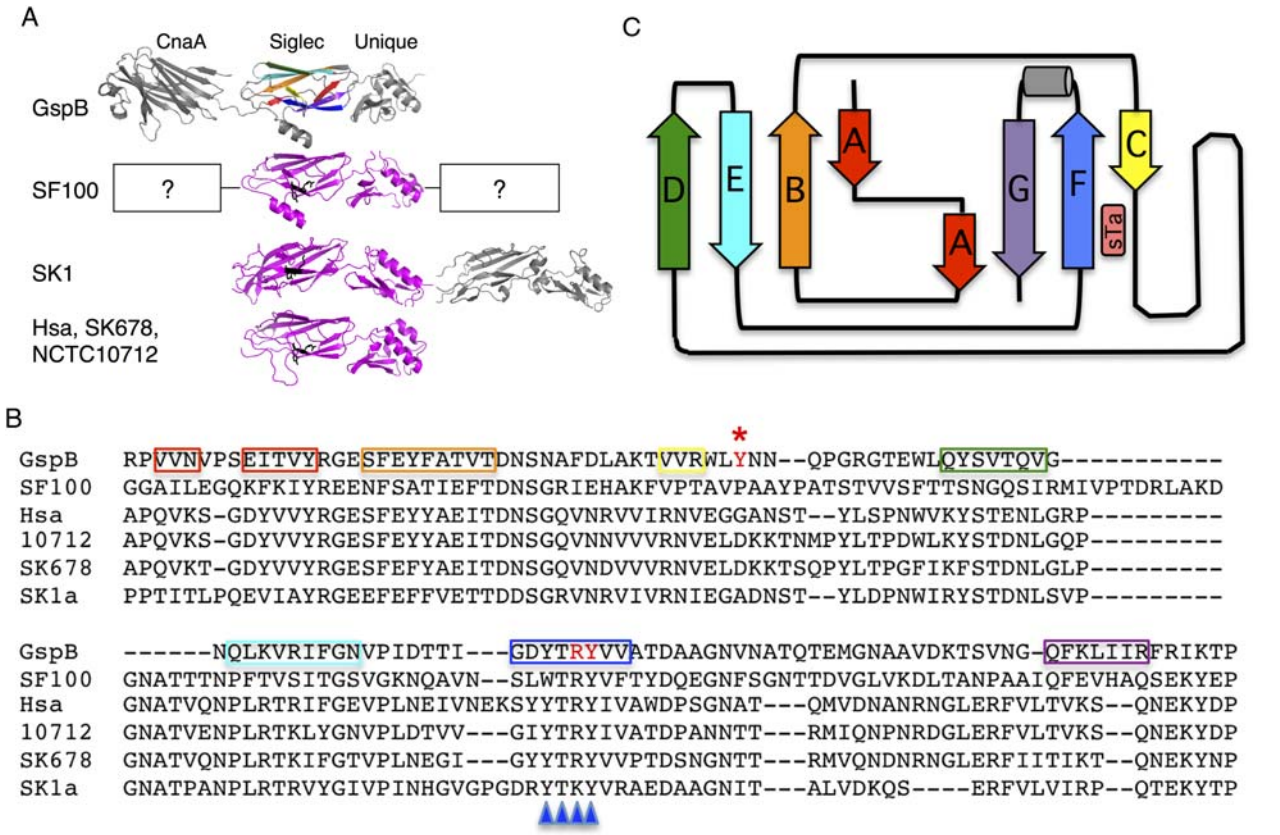

Fig. 3. Comparison of additional Siglec-like BR sequences. (**A**) BR domain composition. The β strands A through G of the GspB Siglec domain are indicated in color. A, red; B, orange; C, yellow; D, green; E, cyan; F, blue; G, violet. The predicted structures of the SF100, SK1a and SK678 Siglec and Unique domains are shown in magenta, with the F-strand arginine and tyrosine residues indicated as black sticks. The $SF100_{BR}$ includes two domains of unknown structure, indicated with question marks. The $SK1_{BR}$ has two Siglec and two Unique domains; the Siglec domain of the second pair (SK1b, shown in grey) lacks the conserved YTRY motif. (**B**) Alignment of the Siglec domain sequences. GspB residues that are essential for sTa binding (Y443, R484 and Y485) are indicated in red, and Y443 is denoted by a red asterisk. The β-strand residues are boxed and colored as above. The YTRY motif is indicated with blue triangles. (**C**) Ribbon diagram of the V-set Ig fold of GspB, with strands colored as indicated above (adapted from Pyburn et al. 2011).

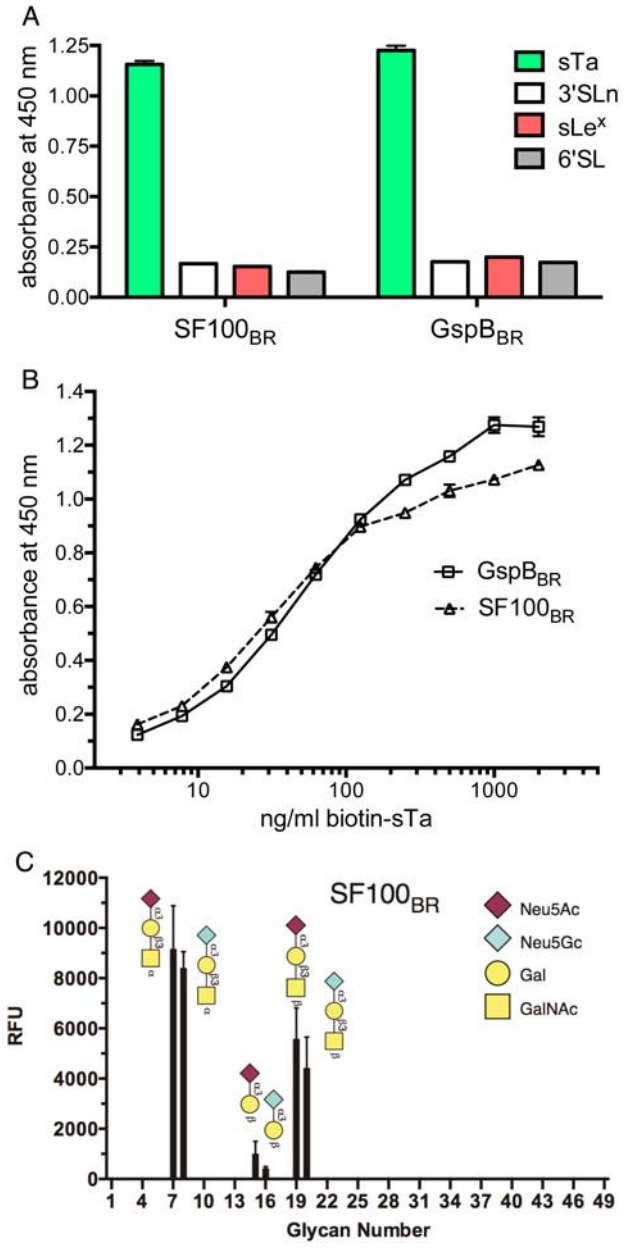

Fig. 4. Ligand binding by the *S. mitis* $SF100_{BR}$. (**A**) Binding of selected biotinylated glycans (10 μg/mL) to immobilized GST-BRs. Binding is reported as the mean ± standard deviation, with $n = 4$. (**B**) Binding of biotinylated sTa to immobilized GST-BRs. Binding is reported as the mean ± standard deviation, with $n = 2$. (**C**) Binding of GST-$SF100_{BR}$ to the array of immobilized glycans ($n = 4$). This figure is available in black and white in print and in color at *Glycobiology* online.

that the $SF100_{BR}$ would bind sialoglycans, but might display a distinctly different ligand repertoire, as compared with GspB.

To address this possibility, we first assessed by ELISA the binding of $SF100_{BR}$ to four different sialoglycans; sTa, 3′sialyllactosamine (3′SLn; Neu5Acα2–3Galβ1–4GlcNAc), 6′sialyllactose (6′SL; Neu5Acα2–6Galβ1–4Glc) and the sialyl Lewis antigen X (sLe^X; Neu5Acα2–3Galβ1–4[Fucα1–3]GlcNAc). Of the four glycans, only sTa was readily bound by the $SF100_{BR}$ (Figure 4A). When binding was assessed using a wide range of glycan concentrations, the binding affinities of GspB and SF100 BRs for sTa appeared to be comparable (Figure 4B).

We next examined binding to our array of sialylated glycans (Figure 4C). When tested with this greatly expanded set of ligands, the $SF100_{BR}$ again showed selective binding to sTa, with minimal binding to other compounds. However, some key differences between the SF100 and GspB ligand repertoires were noted. Unlike $GspB_{BR}$, the $SF100_{BR}$ did not demonstrate a strong preference for sTa linked in the α- vs. β-configuration (glycan **7** versus **19**, respectively). Additionally, the $SF100_{BR}$ could bind Neu5Gc compounds (glycans **8** and **20**) nearly as well as the Neu5Ac counterparts. Aside from these subtle differences, and despite the low similarity in the primary amino acid sequences, the results clearly demonstrate that the two BRs have similar binding spectra, with high selective binding for sTa. In conjunction with our previous studies, these results also indicate that selective sTa binding may be a common property of the SRR glycoproteins of oral streptococci.

Identification and characterization of a novel sialoglycan-binding motif in the Siglec domain

Alignment of the $GspB_{BR}$ and $SF100_{BR}$ with 19 other putative Siglec-like BR sequences revealed a semi-conserved consensus YTRY motif that coincides with the critically important R484 and Y485 residues in the GspB Siglec domain (Figure 3B and data not shown). These residues reside in the F strand of the V-set Ig fold (Figure 3C). The conserved arginine is particularly striking, in that the F-strand in mammalian Siglecs contains an arginine residue that is critical for sialoglycan binding, and has been shown to directly contact Neu5Ac or Neu5Gc in co-crystal structures (May et al. 1998; Alphey et al. 2003; Attrill et al. 2006; Zhuravleva et al. 2008). To determine whether the F-strand arginine has an analogous role in Siglec-like BRs beyond $GspB_{BR}$, we replaced this residue with glutamate in the $SF100_{BR}$, and then assessed binding to sTa. The $SF100_{BR}$ R490E variant showed a severe reduction in sTa binding (Figure 5A), confirming a key role for the conserved arginine residue in sialoglycan binding. The results also strongly suggest that, despite the highly divergent sequence, the $SF100_{BR}$ likely contains a Siglec fold (as predicted by Phyre2).

We next assessed the contribution of the YTRY motif to sialoglycan binding by the Hsa_{BR}. Hsa_{BR} was selected for this analysis because of its relatively high affinity for a variety of α2–3-linked sialoglycans, including disaccharides and trisaccharides (Deng et al. 2014). This allowed us to monitor for shifts in the ligand repertoire, as well as for effects on relative affinity. Surprisingly, a Y338F variant of the Hsa_{BR} showed slightly higher binding to three selected sialoglycans (Figure 5B), but did not noticeably impact the repertoire of ligands bound on the array (data not shown). Substitution of the conserved threonine and arginine residues (T339V and R340E, respectively) resulted in severely decreased binding to all sialoglycans (Figure 5A and B, and data not shown). A Y341F substitution resulted in a modest reduction in binding to Neu5Acα2–3Gal and 3′SLn, but a substantial decrease in binding to sTa (Figure 5B and C). These results indicate that T339 and R340 provide essential Neu5Ac or Gal contacts, and suggest that Y341 provides additional secondary contacts with the GalNAc at the reducing end of sTa.

We further determined the extent to which the loss of sialoglycan binding by these BRs impacted the ability to bind immobilized human platelets. The $SF100_{BR}$ was found to bind platelets, although with substantially lower affinity than was seen for Hsa_{BR} (Figure 6A). For both BRs, replacement of the F-strand arginine residue severely reduced platelet binding (Figure 6B). Thus, as was seen previously for GspB, the interaction with platelets is primarily driven by the ability to bind the sialoglycan ligands. Moreover, the arginine residue within the conserved YTRY motif is essential for this interaction.

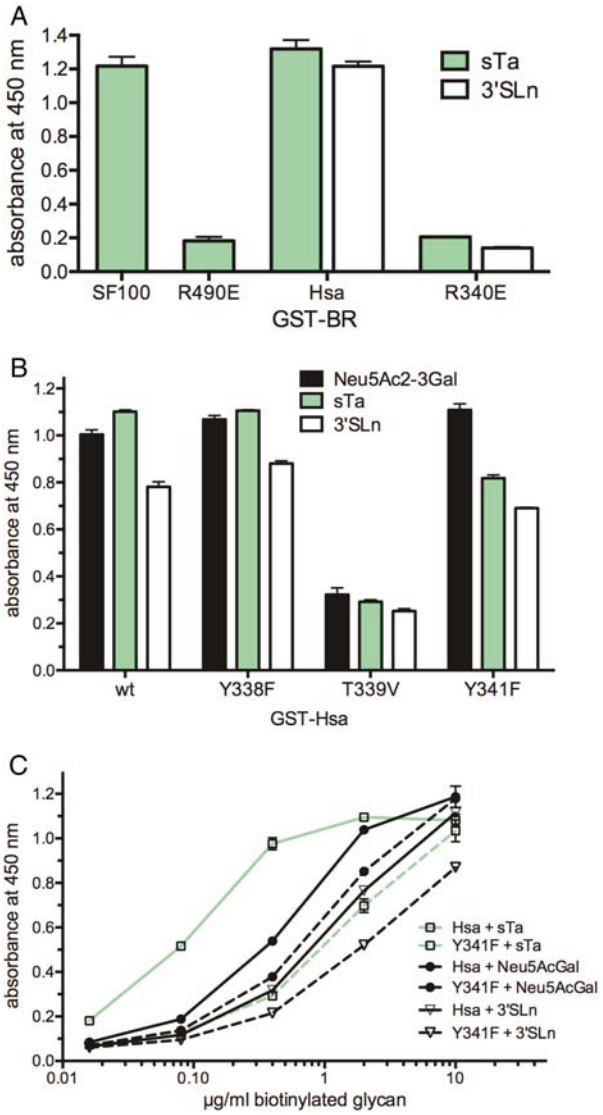

Fig. 5. Contribution of the F-strand motif residues to sialoglycan binding by the SF100 and Hsa BRs. (**A**) Binding of biotinylated sTa or 3′SLn (10 µg/mL) to immobilized GST-BRs. R490E and R340E are the F-strand arginine mutants of GST-SF100$_{BR}$ or GST-Hsa$_{BR}$, respectively. Binding is reported as the mean ± standard deviation, with $n = 4$. (**B**) Binding of selected biotinylated glycans (5 µg/mL) to immobilized wild-type or mutant GST-Hsa$_{BR}$. Binding is reported as the mean ± standard deviation, with $n = 3$. (**C**) Binding of biotinylated glycans to immobilized GST-Hsa$_{BR}$ wild type or the Y341F variant. Binding is reported as the mean ± standard deviation, with $n = 2$. This figure is available in black and white in print and in color at *Glycobiology* online.

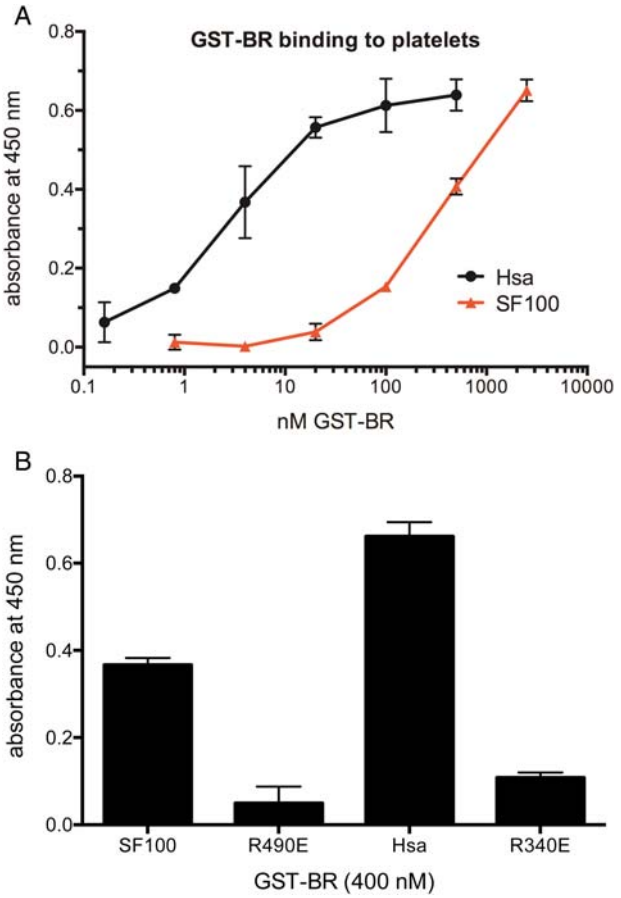

Fig. 6. The F-strand Arg is critical for platelet binding by Hsa and SF100 BRs. (**A**) Binding of GST-BRs to immobilized platelets. Binding is reported as the mean ± standard deviation, with $n = 4$. (**B**): Comparative binding of wild-type and mutant BRs (400 nM) to immobilized platelets. Binding is reported as the mean ± standard deviation, with $n = 6$. This figure is available in black and white in print and in color at *Glycobiology* online.

Three Hsa-like BRs show distinctly different ligand repertoires

Although most of the Siglec-like SRR glycoproteins characterized to date are selective for sTa, Hsa has high affinity for multiple glycans with terminal α2–3-linked sialic acids (Deng et al. 2014). To determine whether broad specificity in ligands is a common property of these adhesins, and to gain insight into additional features beyond the F-strand motif that contribute to ligand selectivity and affinity, we next examined the ligand-binding properties of the BRs from *S. mitis* NCTC10712, and *S. sanguinis* strains SK678 and SK1. The NCTC10712$_{BR}$ is 94% similar to Hsa$_{BR}$ (80% identity), and the SK678$_{BR}$ is 89% similar (72% identity). Both BRs include the F-strand YTRY motif, and could be modeled onto the GspB$_{BR}$ structure with 100% confidence. The SK1$_{BR}$ is unusual in that it includes two non-identical Siglec/Unique domain pairs (indicated as "a" and "b" in Figure 3A) that can both be modeled onto the GspB$_{BR}$ structure. The "a" pair is 79% similar (52% identical) to the Hsa$_{BR}$, and includes a YTKY sequence in the F strand of the Siglec domain. The "b" pair is less similar to Hsa$_{BR}$, and has YTFK in the F-strand, in place of the YTRY motif.

In an initial screen of binding to four defined sialoglycans, the three Hsa-like BRs displayed ligand repertoires that were distinctly different from each other and from Hsa$_{BR}$ (Figure 7A). The SK1$_{BR}$ readily bound all of the α2–3-linked sialoglycans, including the fucosylated tetrasaccharide sLe^X. The NCTC10712$_{BR}$ also showed broad specificity, including binding to sLe^X, but showed less avid binding to sTa. The SK678$_{BR}$ showed a preference for 3′SLn, which has not been seen with any other Siglec-like BR.

When binding was examined over a range of glycan concentrations, the SK1$_{BR}$ bound six selected α2–3 sialoglycans with nearly equal affinity (Figure 7B). Unlike the Hsa$_{BR}$, the SK1$_{BR}$ bound the fucosylated compounds sLe^X and sLe^a (the sialyl Lewis antigen a; Neu5Acα2–3Galβ1–3[Fucα1–4]GlcNAc), as well as the non-fucosylated counterparts 3′SLn and sLe^C (the sialyl Lewis antigen C; Neu5Acα2–3Galβ1–3GlcNAc). For the NCTC10712$_{BR}$, a higher

affinity for compounds with β1–4 vs. β1–3 linkage between the second and third monosaccharides was apparent (i.e. higher binding to 3′SLn, when compared with sTa and sLe^C). This bias was also evident in the interaction with fucosylated Lewis antigens, in that the $NCTC10712_{BR}$ readily bound sLe^X, but did not detectably bind sLe^a. The $SK1_{BR}$ and $NCTC10712_{BR}$ both displayed a relatively high affinity for the Neu5Acα2–3Gal disaccharide, which could explain in part the broad specificity for longer glycans (i.e. any glycan with a terminal Neu5Acα2–3Gal).

When tested on our sialoglycan array (Figure 7C), the $SK1_{BR}$ again displayed broad specificity. Unlike Hsa_{BR}, however, this BR could discriminate between Neu5Gc and Neu5Ac compounds. Moreover, the $SK1_{BR}$ showed reduced, rather than enhanced, reactivity with sulfated glycans, such as glycans **36** and **32**, which are 6-sulfo-GlcNAc forms of glycans **3** (3′SLn) and **30** (sLe^X), respectively. The $NCTC10712_{BR}$ also showed broad specificity for Neu5Ac compounds, and relatively little binding to Neu5Gc compounds, but showed strongly enhanced binding to sulfated forms of 3′SLn and sLe^X (i.e. stronger signals with **36** and **32** vs. **3** and **30**, respectively). The results for $SK678_{BR}$ were surprising in that this BR was very highly selective for 3′SLn, and could discriminate between 3′SLn and 3′SL (compounds **3** and **13**, respectively). Although a 6-sulfo modification of 3′SLn had no apparent effect on binding by $SK678_{BR}$, it did enhance binding to sLe^X (compound **32** vs. **30**). Thus, in spite of the sequence similarity, the three Hsa-like BRs display quite different ligand-binding characteristics.

Differences in ligand repertoire reflect differences in salivary and platelet glycoprotein recognition

Siglec-like BRs have previously been examined for binding to purified platelet receptors and salivary glycoproteins. However, the

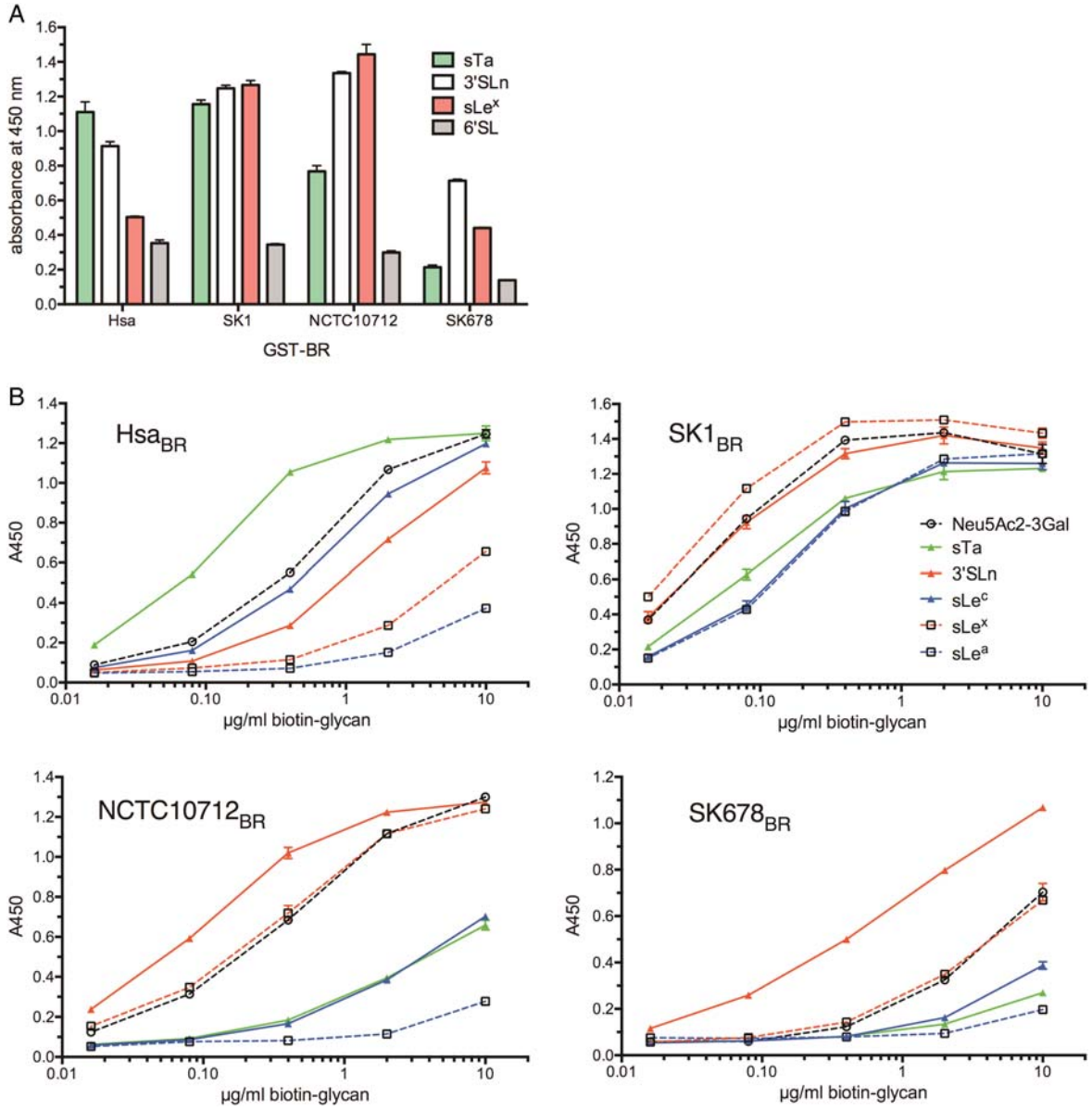

Fig. 7. Ligand binding by the Hsa-related SK1, NCTC10712 and SK678 BRs. (**A**) Binding of selected biotinylated glycans (10 μg/mL) to immobilized GST-BRs. Binding is reported as the mean ± standard deviation, with $n = 3$. (**B**) Binding of biotinylated glycans to immobilized GST-BRs. Binding is reported as the mean ± standard deviation, with $n = 2$. (**C**) Binding of GST-BRs to the array of immobilized glycans ($n = 4$). This figure is available in black and white in print and in color at *Glycobiology* online.

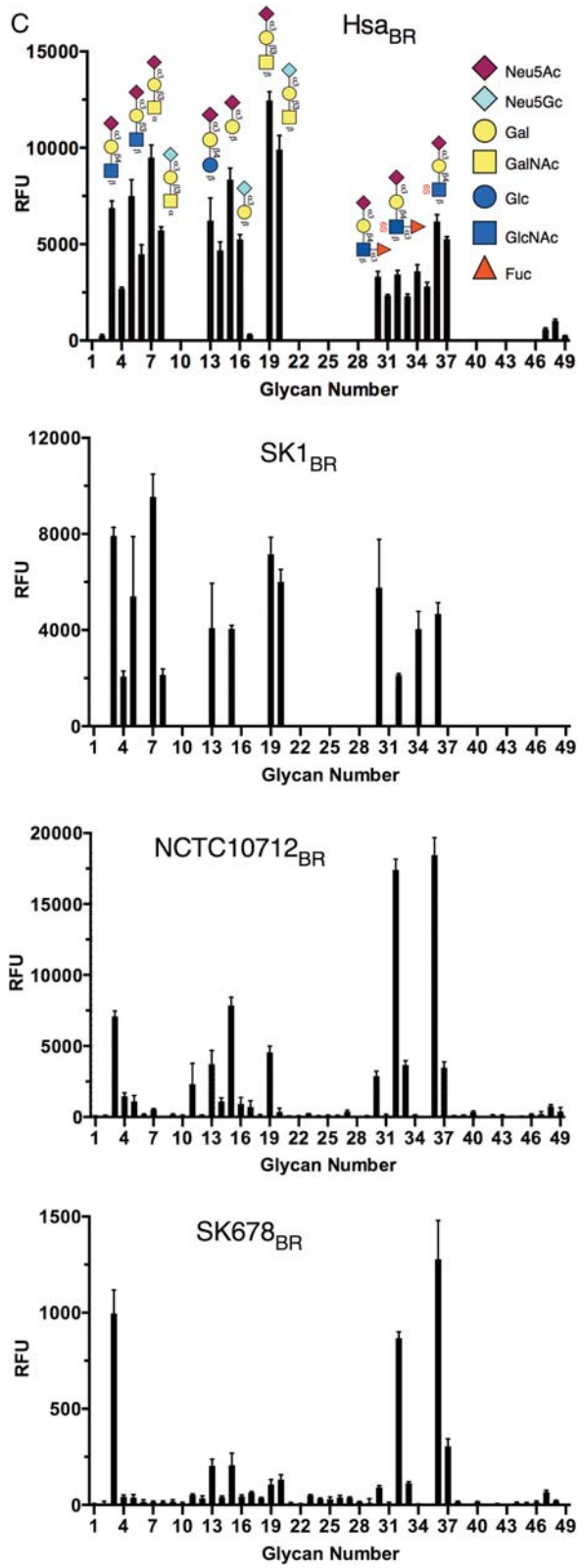

Fig. 7. *Continued*

salivary glycoproteins in particular are known to have an inherent heterogeneity of glycoforms. Specifically, the glycan composition of the salivary mucin MG2/MUC7 is known to vary with blood group and secretor status (Prakobphol et al. 1998; Karlsson and Thomsson 2009). To our knowledge, variability of glycan structures on platelet glycoproteins has not been reported. Given the likelihood that purification strategies may omit or enrich for some glycosylated sub-populations, we chose to examine binding to relatively unprocessed platelet and saliva samples. In addition, we probed these samples by far-western blotting, since this method can detect differently modified glycoprotein populations, as indicated by different electrophoretic mobilities (Takamatsu et al. 2006).

Initial testing of GspB$_{BR}$ and Hsa$_{BR}$ against sub-mandibular sublingual (SMSL) saliva from 15 donors revealed heterogeneous reactivity with MG2/MUC7 (i.e. strong reactivity with some donor samples, and little or no reactivity with others), whereas no variability in the strength of reactivity was seen with platelet GPIbα from different donors (data not shown). We next compared the reactivity of GspB$_{BR}$, Hsa$_{BR}$ and the four new Siglec-like BRs against a panel of SMSL saliva from five donors alongside one platelet sample (Figure 8). The results revealed that the sTa-selective SF100$_{BR}$ and GspB$_{BR}$ are subtly different in MG2/MUC7 reactivity (compare samples S3 and S5), but both BRs had extremely low reactivity with GPIbα, when compared with MG2/MUC7.

Among the Hsa-like BRs, Hsa$_{BR}$ reacted with faster-migrating forms of MG2/MUC7 (<150 kDa apparent MW), whereas SK1$_{BR}$, SK678$_{BR}$ and NCTC10712$_{BR}$ reacted with more slowly migrating glycoforms (>150 kDa). Differences in reactivity with the five MG2/MUC7 samples were also apparent. For example, the Hsa$_{BR}$ reacted well with samples S1 and S2, in addition to S4 and S5, whereas SK1$_{BR}$, SK678$_{BR}$ and NCTC10712$_{BR}$ showed much lower reactivity with samples S1 and S2. The Hsa$_{BR}$ reactivity with GPIbα was on par with the MG2/MUC7 reactivity. In contrast, the SK678 and NCTC10712 BRs showed very high GPIbα reactivity, whereas the SK1$_{BR}$ showed relatively low binding to GPIbα. The composite far-western blotting results suggest that the native glycan targets for these adhesins may be more complex, or may include different modifications, than those used in our ELISA and array studies.

We also compared the six BRs for binding to immobilized human platelets. The results paralleled those of the reactivity with GPIbα in far-western blots, in that the Hsa-like BRs showed higher levels of binding when compared with GspB$_{BR}$ and SF100$_{BR}$ (Figure 9). However, the SK678$_{BR}$ showed very high reactivity with GPIbα in the far-western blots, but only moderate binding to immobilized platelets (on par with the level of SK1$_{BR}$ binding). The combined results indicate that no two BRs are alike in reactivity with the natural human ligands, and may therefore recognize distinctly different features of these glycoproteins.

Platelet GPIb reactivity is correlated with the strength of binding to sulfated versus non-sulfated glycans

It was initially puzzling to observe the relatively weak reactivity of the SK1$_{BR}$ with immobilized platelets and GPIbα, despite strong reactivity with most α2–3 sialoglycans. A close inspection of the array data then indicated that the three BRs showing strong reactivity with GPIbα in far-western blots (Hsa$_{BR}$, SK678$_{BR}$ and NCTC10712$_{BR}$) can bind sulfated glycans as well as or better than the non-sulfated counterparts. Sulfation of sLe^X has been shown to enhance or inhibit binding by human sialoglycan-binding proteins (L-selectin and Siglecs), and could thus have important implications for the Siglec-like BRs beyond adherence to platelet and salivary glycoproteins. We therefore examined binding to the available biotinylated sulfated sialoglycans 6′S-sLe^X (6′-sulfo-Gal) and 6S-sLe^X (6-sulfo-GlcNAc) in an

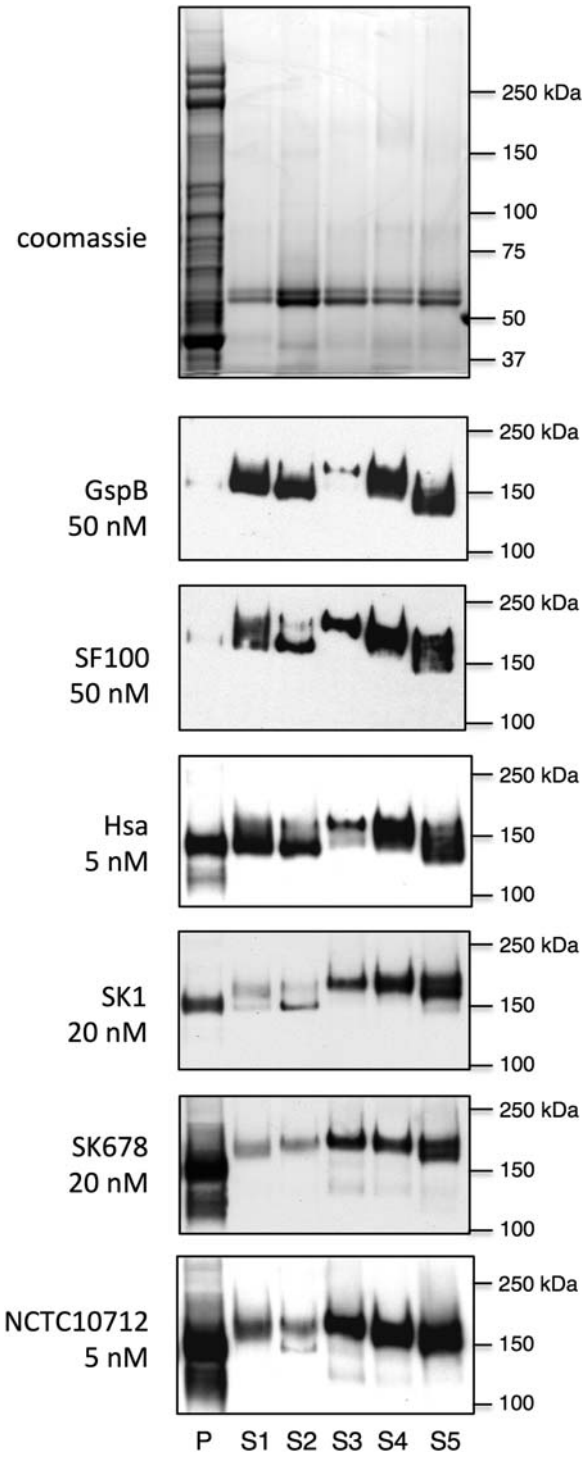

Fig. 8. Far-western blot of six GST-BRs against salivary and platelet glycoproteins. Lanes contain a lysate of fresh, washed platelets from a single donor (P), or saliva from five different donors (S1 through S5). Glycoproteins were separated by electrophoresis through a 3–8% polyacrylamide gradient, and then stained (top panel) or transferred to nitrocellulose and then probed with GST-BRs as indicated. No signals were detected outside of the cropped region, aside from a high MW band in the platelet sample. The 140–150 kDa BR-reactive bands were previously determined to contain predominantly GPIbα or MG2/MUC7 (Takamatsu et al. 2005, 2006). The precise amount of GPIbα and MG2/MUC7 within the samples is undetermined, but remained constant across the different BRs tested.

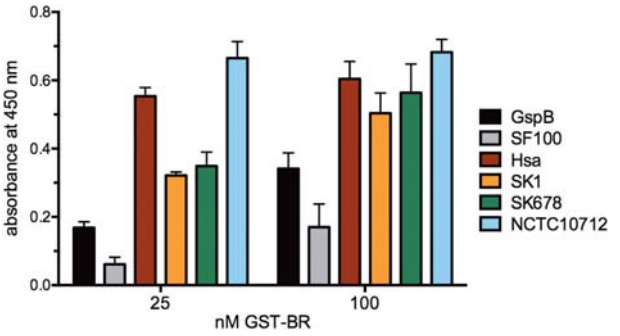

Fig. 9. Quantitative binding to immobilized human platelets. GST-BRs (25 or 100 nM) were applied to monolayers of fixed platelets. Binding is reported as the mean ± standard deviation, with $n = 4$. This figure is available in black and white in print and in color at *Glycobiology* online.

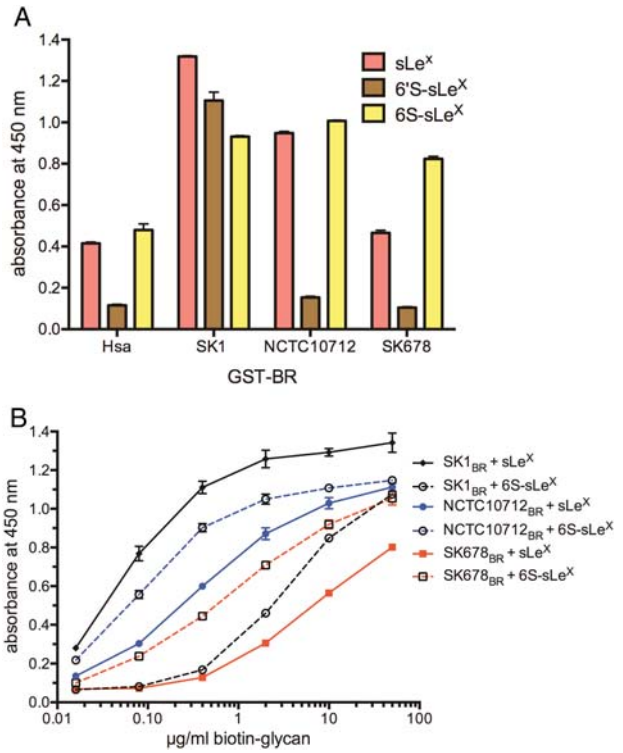

Fig. 10. Binding of the Hsa-like BRs to sulfated forms of sialyl Lewis X. (**A**) Binding of biotinylated sLe^X, 6′S-sLe^X or 6S-sLe^X (10 μg/mL) to immobilized GST-BRs. Binding is reported as the mean ± standard deviation, with $n = 3$. (**B**) Binding of biotinylated sLe^X or 6S-sLe^X to immobilized GST-BRs. Binding is reported as the mean ± standard deviation, with $n = 2$. This figure is available in black and white in print and in color at *Glycobiology* online.

ELISA assay. All of the Hsa-like BRs showed reduced binding to 6′S-sLe^X, as compared with non-sulfated sLe^X (Figure 10A). The SK1$_{BR}$ also showed reduced binding to 6S-sLe^X, whereas the Hsa$_{BR}$, SK678$_{BR}$ and NCTC10712$_{BR}$ did not. We next compared binding to sLe^X vs. 6S-sLe^X using a broad range of glycan concentrations (Figure 10B). The results corroborated those in the array, and clearly demonstrated that the binding of the SK678$_{BR}$ and NCTC10712$_{BR}$ to 6S-sLe^X is greater than binding to the non-sulfated tetrasaccharide, whereas SK1$_{BR}$ binding is strongly inhibited by the 6-sulfo-GlcNAc

modification. Since the relatively weak binding of the $SK1_{BR}$ to GPIbα and platelets is correlated with the inability to bind sulfated sialoglycans, the results further suggest that at least some of the glycans on GPIbα are likely to have the 6-sulfo-GlcNAc modification.

Discussion

Our findings reveal both conserved and diverse features of a family of sialoglycan-binding adhesins expressed by streptococci, and include the first examples of Siglec-like BRs expressed by *S. mitis*. In particular, our results demonstrate that the Siglec and Unique domains of $GspB_{BR}$ are necessary and sufficient for sialoglycan binding, and that these two domains are conserved in other streptococcal BRs. Moreover, the streptococcal adhesins display both structural and functional similarity to mammalian Siglecs, although they present some important and distinct differences. The similarities to mammalian Siglecs include a predicted V-set Ig fold, and the presence of a conserved arginine residue in the F-strand of the Siglec domain. However, we found that the F-strand of the streptococcal Siglec domain includes a novel YTRY motif containing residues essential for sialoglycan binding, as single amino acid substitutions in this region can abolish binding (Figure 5 and previous studies). The linear YTRY motif contrasts with the ligand-binding sites identified in mammalian Siglecs, which, in addition to the F-strand arginine, have a conserved aromatic and other contributing residues located in a gap of the neighboring G-strand (May et al. 1998; Alphey et al. 2003; Attrill et al. 2006; Zhuravleva et al. 2008). As with the mammalian Siglecs (Yamaji et al. 2002), residues from a loop adjacent to the C-strand in the streptococcal BRs may also contribute to ligand binding. Indeed, this region includes one of the three critical sialoglycan-binding residues of GspB (Y443), and appears to be hypervariable among the BRs examined here (Figure 3B). Finally, the essentiality of the adjacent Unique domain is reminiscent of the C2-set domain that adjoins the carbohydrate-binding module of mammalian Siglecs (reviewed in Angata 2006; Varki and Angata 2006). The precise role of the Unique domain is presently unclear, but it is unlikely to be directly involved in ligand binding. Our results suggest that the sialoglycans are orientated along the F-strand in a direction leading away from the Unique domain. That is, T339 and R340 of Hsa mediate binding to the terminal Neu5Ac-Gal, whereas Y341 appears to contact the substrate-proximal sugar (i.e. the GalNAc of sTa). This is in excellent agreement with the orientation of Neu5Gcα2–3Gal bound to $SrpA_{BR}$ observed in the co-crystal structure (Bensing et al. 2016). Since the Unique domain does not appear to interact directly with glycans, it is possible that this domain allosterically modulates the conformation of the Siglec domain. Proof of this possibility will require further analysis.

A surprising finding was the broad diversity of ligands bound by the streptococcal Siglec-like adhesins, and the variable impact of fucosylation and sulfation of sialoglycans on the interactions. Importantly, the ligand repertoires are not readily predicted from the primary amino acid sequences of the BRs. For example, the *S. mitis* $SF100_{BR}$ is selective for sTa, despite having a sequence quite unlike that of $GspB_{BR}$ and other sTa-specific Siglec-like BRs. In contrast, the three highly similar Hsa-like BRs are surprisingly diverse in their binding spectra. The *S. sanguinis* $SK678_{BR}$ is highly selective for 3′SLn, despite being very similar to the Hsa_{BR}, which displays broad specificity. The *S. mitis* $NCTC10712_{BR}$ shows broader specificity, but is distinct from the Hsa_{BR} in that it binds 3′SLn and sLe^X more avidly than sTa or sLe^C. The *S. sanguinis* $SK1_{BR}$ binds the widest variety of Neu5Acα2–3Gal sialoglycans seen to date, including the sialyl Lewis antigens, but does not readily bind sulfated or Neu5Gc compounds. Thus, the overall diversity of ligands is extensive, and at least comparable with that seen with the rapidly evolving CD33 family of mammalian Siglecs (Blixt et al. 2003; Angata 2006; Varki and Angata 2006).

The differences in ligand repertoire of the BRs are reflected in the reactivity with the salivary mucin MG2/MUC7 from different donors. This diversity has likely evolved in parallel with alterations in the human sialome (Varki and Angata 2006), but may be specifically related to modifications of the glycan moieties presented on MG2/MUC7, which are complex and incompletely characterized (Prakobphol et al. 1998; Karlsson and Thomsson 2009). The difference in ligand repertoire is also reflected in the different levels of binding to the platelet glycoprotein GPIbα. The interaction of bacteria with platelets is important in that it can render some commensal bacterial species accidental or opportunistic pathogens. Oral microbes can enter the bloodstream through lesions in the oral epithelium or during routine dental work, and then establish infections on a damaged or stressed endocardium. The interaction of microbes with platelet thrombi that have formed on injured endovascular surfaces has long been thought to be an important mechanism by which this process occurs (Durack and Beeson 1972; Durack 1975). In addition, we recently showed that the Siglec-like BRs can selectively target bacteria, including endocarditis-associated strains, to platelets in whole blood (Deng et al. 2014). This could enable the microbes to be passively carried to, and deposited on, damaged heart valves. Why species such as *S. gordonii*, which are minor components of the oral microbiota, may be disproportionately overrepresented in cases of infective endocarditis remains unclear, but may well be related to the presence of the sialoglycan-binding SRR adhesins.

An unresolved question is whether adherence to specific sialoglycans can impact the propensity of streptococci to cause endocarditis. Our results suggest that a high affinity for sTa or, more likely, a lack of binding to certain sulfated epitopes, may be important for this process. Certainly, the BRs from the limited set of endocarditis-associated strains tested here (M99, SF100 and SK1) show both high binding to sTa and a lack of binding to sulfated glycans. How the different mechanisms of binding to GPIb might affect pathogenesis is unclear, but it could be related to a modulation of platelet adherence to the damaged valve endothelium. GPIb has two roles in platelet adherence to endothelial surfaces. In addition to the well-known interaction with von Willebrand factor, the GPIb-IX-V complex has been characterized as a counter-receptor for P-selectin that is expressed on activated endothelial cells. Of note, $6S\text{-}sLe^X$ is a preferred ligand for selectins, including P-selectin. It is possible that the binding of microbes to GPIb via $6S\text{-}sLe^X$ could actually impede the further deposition of platelets onto infected thrombi, and the subsequent development of macroscopic vegetations. Thus, a more specific question is whether the binding of streptococci to sTa vs. other glycans on GPIb can have different effects on colonization of the endocardium. We are now in a position to answer these questions, by assessing various aspects of pathogenesis with different BRs expressed in an otherwise isogenic background.

Materials and methods

Reagents

Dulbecco's phosphate-buffered saline (DPBS), mutanolysin, lysozyme, horse radish peroxidase-conjugated antibodies, horse radish peroxidase-conjugated streptavidin, OPD and phosphate-citrate buffer were purchased from Sigma. Glutathione-sepharose and pGEX-3X were from GE Healthcare. Rabbit polyclonal anti-GST antibodies were from Life Technologies (for ELISA and far-western blotting) or from ThermoFisher Scientific (for array studies). Alexa Fluor 555-conjugated goat anti-rabbit IgG was from Molecular

Probes. Multivalent biotinylated glycans (polyacrylamide polymers of ~30 kDa containing a 5:20:100 molar ratio of biotin:glycan: acrylamide) were obtained from GlycoTech. A 10× casein solution (Roche Blocking Reagent) was diluted to 1× and used to block non-specific binding in some experiments.

Preparation of bacterial chromosomal DNA

Chromosomal DNA was extracted from streptococci using the Wizard Genomic DNA Purification Kit (Promega) as described by the manufacturer, except that the bacterial cells were washed with distilled water prior to lysis with a solution that contained 200 units mL^{-1} mutanolysin and 50 mg mL^{-1} lysozyme.

BR identification, cloning and expression

The SRR glycoprotein sequences of SK1, SK678 and NCTC10712 were identified through BLAST searches of the public databases, using Hsa_{BR} as the query sequence. The SF100 SRR glycoprotein was located by using the highly conserved GspB signal peptide in a BLAST search against the unpublished SF100 genome, and has been deposited in GenBank, under the accession number KU519294. Additional information regarding the BR domains used in these studies is indicated in Table II.

Cloning of the *gspB* and *hsa* BRs in pGEX-3X was described previously (Takamatsu et al. 2005). Additional BRs were identified and then cloned similarly. In brief, the corresponding DNA coding regions, along with 5′ BamHI and 3′ EcoRI linkers, were obtained as commercially synthesized products, or were amplified from chromosomal DNA by PCR. For the $SF100_{BR}$, codons 220–726 of the corresponding gene sequence were amplified from SF100 chromosomal DNA using the forward primer 5′-AAAAGGATCCCAGCTCGGG AGACAGTGAAAGAATC along with the reverse primer 3′-AAAAG AATTCGAGCTCACCGAGGCAGACTGGC (BamHI and EcoRI sites are underlined). The $SK1_{BR}$, $NCTC10712_{BR}$ and $SK678_{BR}$ coding regions, along with BamHI and EcoRI linkers, were synthesized (Life Technologies). The GspBΔunique ($GspB_{BR}$Δ523–617) was also synthesized (GenScript). GspBΔcnaA ($GspB_{BR}$ Δ252–398) was generated by two-stage PCR, using the partially overlapping primers 5′-ACTGAAAGTGCTGATACAGAAAGGCCAGTTGTTAATG (forward codons 248–251, plus 399–406) along with the reverse primer 5′-CCTTTCTGTATCAGCACTTTCAGTAACAGCTCGACG (codons 402–399, plus 251–244). Replacement of the arginine codons in the SF100 and Hsa BR gene sequences was accomplished by two-stage PCR reactions. Additional replacements in Hsa_{BR} were generated by synthesis of the entire variant BR coding region (Life Technologies).

After cloning in the pGEX-3X expression vector, the wild-type and mutant BR coding sequences were confirmed by DNA sequence analysis (Sequetech). Cultures of *E. coli* strain BL21 carrying the pGEX expression plasmids were grown in LB with 50 μg/mL carbenicillin until an OD_{600} of ~0.9, and the expression of GST fusion proteins was induced by the addition of IPTG to a final concentration of 1 mM. Cultures were incubated for 4 h at 24°C. Cells were harvested by centrifugation and lysed by sonication, and the GST fusion proteins were purified using glutathione-sepharose according to the manufacturer's instructions. The eluted proteins were exchanged into DPBS and stored at −80°C.

Binding of biotinylated glycans to immobilized GST-BRs

Purified GST-BRs (500 nM in DPBS) were immobilized in 96-well plates, and the binding of biotinylated glycans was assessed as described previously (Deng et al. 2014), with minor modifications. In brief, multivalent biotinylated glycans were added to wells at the indicated concentrations in DPBS containing 1× Blocking Reagent. After 90 min at RT, wells were rinsed three times to remove the unbound glycans, and bound glycans were detected with streptavidin-conjugated horse radish peroxidase, along with a solution of 0.4 mg OPD per mL phosphate-citrate buffer. The absorbance at 450 nm was measured after ~20 min.

Binding of GST-BRs to immobilized glycans in a microarray

Glycan microarrays were fabricated using epoxide-derivatized slides, fitted into hybridization cassettes and divided into eight subarrays as previously described (Padler-Karavani et al. 2012; Deng et al. 2014). The subarrays were blocked with ovalbumin (1%, w/v) in PBS pH 7.4 for 1 h at RT, with gentle shaking. The blocking solution was removed, and diluted GST-BRs at concentrations ranging from 10 to 100 nM were added. After incubation for 2 h at RT, the slides were washed extensively to remove any non-specifically bound proteins. The subarrays were then incubated for 1 h at RT with anti-GST antibodies (1:2000 dilution in DPBS), washed, incubated for 1 h at RT with Alexa Fluor 555-conjugated anti-rabbit IgG (1:10,000 in DPBS), washed again, and then dried. The microarray slides were scanned using a Genepix 4000B microarray scanner (Molecular Devices), and data analysis was performed using the Genepix Pro 7.0 analysis software.

Binding of GST-BRs to platelet monolayers

To assess binding to platelets, fresh human platelets were washed, fixed and immobilized in 96-well plates as described (Bensing and Sullam 2002). All subsequent binding steps were carried out at room temperature. To reduce non-specific adherence, the wells were treated with 50 μL of 1× Blocking Reagent in DPBS for 1 h. The blocking solution was replaced with 50 μL of purified GST-BRs, ranging from 0.16 to 2500 nM in 1× blocking solution. The plates were incubated for 1 h with vigorous rocking, wells were rinsed three times with 100 μL DPBS, and 50 μL of a rabbit polyclonal anti-GST diluted 1:500 in 1× blocking solution was added to each well. After 1 h, wells were rinsed three times with 100 μL DPBS, and 50 μL of a peroxidase-conjugated anti-rabbit antibody (1:5000 dilution in DPBS) was added. After incubation for 1 h, wells were rinsed three times with 100 μL DPBS, and 200 μL of a solution of 0.4 mg mL^{-1} OPD was added. The absorbance at 450 nm was read after ~30 min, and the values of wells containing the GST-BRs were adjusted by subtracting the average absorbance value of wells containing a GST control.

Far-western blotting

Samples of SMSL human saliva were collected as described previously (Prakobphol et al. 1998). Washed platelets or SMSL saliva were combined with an equal volume of 2× SDS–PAGE sample buffer and DTT (100 mM final concentration). Samples were boiled for 10 min and proteins were separated by electrophoresis on 3–8% polyacrylamide gradient gels (Life Technologies), and then transferred to BioTraceNT (Pall Corporation). Membranes were incubated for 1 h at RT with 1× Blocking Reagent in DPBS. GST-BRs were then added to a final concentration of 5–50 nM as indicated, and the membranes were incubated for 90 min at RT with gentle rocking. After rinsing three times with DPBS, the membranes were incubated for 1 h at RT with anti-GST diluted 1:2000 into DPBS containing 1× Blocking Reagent.

Membranes were rinsed three times with DPBS, and then incubated for 1 h at RT with horse radish peroxidase-conjugated goat anti-rabbit antibodies diluted 1:20,000 in DPBS. Membranes were again rinsed three times with DPBS, and then developed with SuperSignal West Pico (Thermo Scientific).

Funding

This work was supported by the Department of Veterans Affairs, the Northern California Institute for Research and Education, the National Institutes of Health (AI41513 and R21CA199881 to P.M. S.; AI106987 to P.M.S./T.M.I.; R01GM32373 and U01CA199792 to A.V.; R01DE021041 to S.J.F.) and the American Heart Association (14GRNT20390021 to T.M.I.).

Acknowledgements

We thank Tasia Pyburn for design of the GspBΔunique expression plasmid.

Conflict of interest statement

None declared.

Abbreviations

3′SLn, 3′sialyllactosamine; 6′SL, 6′sialyllactose; BRs, binding regions; DPBS, Dulbecco's phosphate-buffered saline; DPBS, Dulbecco's phosphate-buffered saline; Fuc, fucose; Gal, galactose; GalNAc, *N*-acetyl galactosamine; GlcNAc, *N*-acetyl glucosamine; MG2/MUC7, human salivary mucin glycoprotein 2, which is encoded by the *MUC7* gene; Neu5Ac, *N*-acetyl neuraminic acid; Neu5Gc, *N*-glycolyl neuraminic acid; OPD, *ortho*-phenylenediamine.; sLe^a, sialyl Lewis antigen a; sLe^C, sialyl Lewis antigen C; sLe^X, sialyl Lewis antigen X; SMSL, sub-mandibular sub-lingual; SRR, serine-rich repeat; sTa, sialyl-T antigen.

References

Alphey MS, Attrill H, Crocker PR, van Aalten DM. 2003. High resolution crystal structures of Siglec-7. Insights into ligand specificity in the Siglec family. *J Biol Chem*. 278:3372–3377.

Angata T. 2006. Molecular diversity and evolution of the Siglec family of cell-surface lectins. *Mol Divers*. 10:555–566.

Attrill H, Takazawa H, Witt S, Kelm S, Isecke R, Brossmer R, Ando T, Ishida H, Kiso M, Crocker PR et al. 2006. The structure of siglec-7 in complex with sialosides: Leads for rational structure-based inhibitor design. *Biochem J*. 397:271–278.

Bensing BA, Lopez JA, Sullam PM. 2004. The Streptococcus gordonii surface proteins GspB and Hsa mediate binding to sialylated carbohydrate epitopes on the platelet membrane glycoprotein Ibalpha. *Infect Immun*. 72: 6528–6537.

Bensing BA, Loukachevitch LV, McCulloch KM, Yu H, Vann KR, Wawrzak Z, Anderson S, Chen X, Sullam PM, Iverson TM. 2016. Structural basis for Sialoglycan binding by the Streptococcus sanguinis SrpA adhesin. *J Biol Chem*. 291:7230–7240.

Bensing BA, Seepersaud R, Yen YT, Sullam PM. 2014. Selective transport by SecA2: An expanding family of customized motor proteins. *Biochim Biophys Acta*. 1843:1674–1686.

Bensing BA, Sullam PM. 2002. An accessory sec locus of Streptococcus gordonii is required for export of the surface protein GspB and for normal levels of binding to human platelets. *Mol Microbiol*. 44:1081–1094.

Blixt O, Collins BE, van den Nieuwenhof IM, Crocker PR, Paulson JC. 2003. Sialoside specificity of the siglec family assessed using novel multivalent probes: Identification of potent inhibitors of myelin-associated glycoprotein. *J Biol Chem*. 278:31007–31019.

Denapaite D, Bruckner R, Nuhn M, Reichmann P, Henrich B, Maurer P, Schahle Y, Selbmann P, Zimmermann W, Wambutt R et al. 2010. The genome of Streptococcus mitis B6 – What is a commensal? *PLoS ONE*. 5:e9426.

Deng L, Bensing BA, Thamadilok S, Yu H, Lau K, Chen X, Ruhl S, Sullam PM, Varki A. 2014. Oral streptococci utilize a Siglec-like domain of serine-rich repeat adhesins to preferentially target platelet sialoglycans in human blood. *PLoS Pathog*. 10:e1004540.

Durack DT. 1975. Experimental bacterial endocarditis. IV. Structure and evolution of very early lesions. *J Pathol*. 115:81–89.

Durack DT, Beeson PB. 1972. Experimental bacterial endocarditis. I. Colonization of a sterile vegetation. *Br J Exp Pathol*. 53:44–49.

Karlsson NG, Thomsson KA. 2009. Salivary MUC7 is a major carrier of blood group I type O-linked oligosaccharides serving as the scaffold for sialyl Lewis x. *Glycobiology*. 19:288–300.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 10: 845–858.

Kerrigan SW, Douglas I, Wray A, Heath J, Byrne MF, Fitzgerald D, Cox D. 2002. A role for glycoprotein Ib in Streptococcus sanguis-induced platelet aggregation. *Blood*. 100:509–516.

May AP, Robinson RC, Vinson M, Crocker PR, Jones EY. 1998. Crystal structure of the N-terminal domain of sialoadhesin in complex with 3′ sialyllactose at 1.85 A resolution. *Mol Cell*. 1:719–728.

Padler-Karavani V, Song X, Yu H, Hurtado-Ziola N, Huang S, Muthana S, Chokhawala HA, Cheng J, Verhagen A, Langereis MA et al. 2012. Cross-comparison of protein recognition of sialic acid diversity on two novel sialoglycan microarrays. *J Biol Chem*. 287:22593–22608.

Plummer C, Douglas CW. 2006. Relationship between the ability of oral streptococci to interact with platelet glycoprotein Ibalpha and with the salivary low-molecular-weight mucin, MG2. *FEMS Immunol Med Microbiol*. 48:390–399.

Plummer C, Wu H, Kerrigan SW, Meade G, Cox D, Ian Douglas CW. 2005. A serine-rich glycoprotein of Streptococcus sanguis mediates adhesion to platelets via GPIb. *Br J Haematol*. 129:101–109.

Prakobphol A, Tangemann K, Rosen SD, Hoover CI, Leffler H, Fisher SJ. 1999. Separate oligosaccharide determinants mediate interactions of the low-molecular-weight salivary mucin with neutrophils and bacteria. *Biochemistry*. 38:6817–6825.

Prakobphol A, Thomsson KA, Hansson GC, Rosen SD, Singer MS, Phillips NJ, Medzihradszky KF, Burlingame AL, Leffler H, Fisher SJ. 1998. Human low-molecular-weight salivary mucin expresses the sialyl lewisx determinant and has L-selectin ligand activity. *Biochemistry*. 37:4916–4927.

Pyburn TM, Bensing BA, Xiong YQ, Melancon BJ, Tomasiak TM, Ward NJ, Yankovskaya V, Oliver KM, Cecchini G, Sulikowski GA et al. 2011. A structural model for binding of the serine-rich repeat adhesin GspB to host carbohydrate receptors. *PLoS Pathog*. 7:e1002112.

Ramboarina S, Garnett JA, Zhou M, Li Y, Peng Z, Taylor JD, Lee WC, Bodey A, Murray JW, Alguel Y et al. 2010. Structural insights into serine-rich fimbriae from Gram-positive bacteria. *J Biol Chem*. 285: 32446–32457.

Samen U, Eikmanns BJ, Reinscheid DJ, Borges F. 2007. The surface protein Srr-1 of Streptococcus agalactiae binds human keratin 4 and promotes adherence to epithelial HEp-2 cells. *Infect Immun*. 75:5405–5414.

Schulte T, Lofling J, Mikaelsson C, Kikhney A, Hentrich K, Diamante A, Ebel C, Normark S, Svergun D, Henriques-Normark B et al. 2014. The basic keratin 10-binding domain of the virulence-associated pneumococcal serine-rich protein PsrP adopts a novel MSCRAMM fold. *Open Biol*. 4: 130090.

Seo HS, Minasov G, Seepersaud R, Doran KS, Dubrovska I, Shuvalova L, Anderson WF, Iverson TM, Sullam PM. 2013. Characterization of fibrinogen binding by glycoproteins Srr1 and Srr2 of Streptococcus agalactiae. *J Biol Chem*. 288:35982–35996.

Seo HS, Mu R, Kim BJ, Doran KS, Sullam PM. 2012. Binding of glycoprotein Srr1 of Streptococcus agalactiae to fibrinogen promotes attachment to brain endothelium and the development of meningitis. *PLoS Pathog*. 8: e1002947.

Shivshankar P, Sanchez C, Rose LF, Orihuela CJ. 2009. The Streptococcus pneumoniae adhesin PsrP binds to Keratin 10 on lung cells. *Mol Microbiol*. 73:663–679.

Siboo IR, Chambers HF, Sullam PM. 2005. Role of SraP, a serine-rich surface protein of Staphylococcus aureus, in binding to human platelets. *Infect Immun*. 73:2273–2280.

Takahashi Y, Konishi K, Cisar JO, Yoshikawa M. 2002. Identification and characterization of hsa, the gene encoding the sialic acid-binding adhesin of Streptococcus gordonii DL1. *Infect Immun*. 70:1209–1218.

Takahashi Y, Sandberg AL, Ruhl S, Muller J, Cisar JO. 1997. A specific cell surface antigen of Streptococcus gordonii is associated with bacterial hemagglutination and adhesion to alpha2-3-linked sialic acid-containing receptors. *Infect Immun*. 65:5042–5051.

Takahashi Y, Takashima E, Shimazu K, Yagishita H, Aoba T, Konishi K. 2006. Contribution of sialic acid-binding adhesin to pathogenesis of experimental endocarditis caused by Streptococcus gordonii DL1. *Infect Immun*. 74:740–743.

Takahashi Y, Yajima A, Cisar JO, Konishi K. 2004. Functional analysis of the *Streptococcus gordonii* DL1 sialic acid-binding adhesin and its essential role in bacterial binding to platelets. *Infect Immun*. 72:3876–3882.

Takamatsu D, Bensing BA, Cheng H, Jarvis GA, Siboo IR, Lopez JA, Griffiss JM, Sullam PM. 2005. Binding of the Streptococcus gordonii surface glycoproteins GspB and Hsa to specific carbohydrate structures on platelet membrane glycoprotein Ibalpha. *Mol Microbiol*. 58:380–392.

Takamatsu D, Bensing BA, Prakobphol A, Fisher SJ, Sullam PM. 2006. Binding of the streptococcal surface glycoproteins GspB and Hsa to human salivary proteins. *Infect Immun*. 74:1933–1940.

Varki A, Angata T. 2006. Siglecs–the major subfamily of I-type lectins. *Glycobiology*. 16:1R–27R.

Wu H, Mintz KP, Ladha M, Fives-Taylor PM. 1998. Isolation and characterization of Fap1, a fimbriae-associated adhesin of Streptococcus parasanguis FW213. *Mol Microbiol*. 28:487–500.

Xiong YQ, Bensing BA, Bayer AS, Chambers HF, Sullam PM. 2008. Role of the serine-rich surface glycoprotein GspB of Streptococcus gordonii in the pathogenesis of infective endocarditis. *Microb Pathog*. 45: 297–301.

Yamaji T, Teranishi T, Alphey MS, Crocker PR, Hashimoto Y. 2002. A small region of the natural killer cell receptor, Siglec-7, is responsible for its preferred binding to alpha 2,8-disialyl and branched alpha 2,6-sialyl residues. A comparison with Siglec-9. *J Biol Chem*. 277:6324–6332.

Yang YH, Jiang YL, Zhang J, Wang L, Bai XH, Zhang SJ, Ren YM, Li N, Zhang YH, Zhang Z et al. 2014. Structural insights into SraP-mediated Staphylococcus aureus adhesion to host cells. *PLoS Pathog*. 10: e1004169.

Zhou M, Wu H. 2009. Glycosylation and biogenesis of a family of serine-rich bacterial adhesins. *Microbiology*. 155:317–327.

Zhuravleva MA, Trandem K, Sun PD. 2008. Structural implications of Siglec-5-mediated sialoglycan recognition. *J Mol Biol*. 375:437–447.