

# UC Davis

## UC Davis Previously Published Works

### Title

Tandem sialoglycan-binding modules in a *Streptococcus sanguinis* serine-rich repeat adhesin create target dependent avidity effects.

### Permalink

<https://escholarship.org/uc/item/7th330r3>

### Journal

The Journal of biological chemistry, 295(43)

### ISSN

0021-9258

### Authors

Stubbs, Haley E  
Bensing, Barbara A  
Yamakawa, Izumi  
et al.

### Publication Date

2020-10-01

### DOI

10.1074/jbc.ra120.014177

Peer reviewed

# Tandem sialoglycan-binding modules in a *Streptococcus sanguinis* serine-rich repeat adhesin create target dependent avidity effects

Received for publication, May 2, 2020, and in revised form, July 29, 2020. Published, Papers in Press, August 20, 2020, DOI 10.1074/jbc.RA120.014177

Haley E. Stubbs<sup>1</sup>, Barbara A. Bensing<sup>2,3</sup>, Izumi Yamakawa<sup>4</sup>, Pankaj Sharma<sup>4</sup>, Hai Yu<sup>5</sup>, Xi Chen<sup>5</sup>, Paul M. Sullam<sup>2</sup>, and T. M. Iverson<sup>4,6,7,\*</sup>

From the <sup>1</sup>Graduate Program in Chemical and Physical Biology, Vanderbilt University, Nashville, Tennessee, USA, the <sup>2</sup>Department of Medicine, Veterans Affairs Medical Center, San Francisco, California, USA, the <sup>3</sup>Department of Medicine, University of California, San Francisco, California, USA, the <sup>4</sup>Department of Pharmacology, Vanderbilt University, Nashville, Tennessee, USA, the <sup>5</sup>Department of Chemistry, University of California, Davis, California, USA, the <sup>6</sup>Department of Biochemistry, Vanderbilt University, Nashville, Tennessee, USA, and the <sup>7</sup>Center for Structural Biology, Vanderbilt University, Nashville, Tennessee, USA

Edited by Karen G. Fleming

Sialic acid-binding immunoglobulin-like lectins (Siglec)-like domains of streptococcal serine-rich repeat (SRR) adhesins recognize sialylated glycans on human salivary, platelet, and plasma glycoproteins via a YTRY sequence motif. The SRR adhesin from *Streptococcus sanguinis* strain SK1 has tandem sialoglycan-binding domains and has previously been shown to bind sialoglycans with high affinity. However, both domains contain substitutions within the canonical YTRY motif, making it unclear how they interact with host receptors. To identify how the *S. sanguinis* strain SK1 SRR adhesin affects interactions with sialylated glycans and glycoproteins, we determined high-resolution crystal structures of the binding domains alone and with purified trisaccharides. These structural studies determined that the ligands still bind at the noncanonical binding motif, but with fewer hydrogen-bonding interactions to the protein than is observed in structures of other Siglec-like adhesins. Complementary biochemical studies identified that each of the two binding domains has a different selectivity profile. Interestingly, the binding of SK1 to platelets and plasma glycoproteins identified that the interaction to some host targets is dominated by the contribution of one binding domain, whereas the binding to other host receptors is mediated by both binding domains. These results provide insight into outstanding questions concerning the roles of tandem domains in targeting host receptors and suggest mechanisms for how pathogens can adapt to the availability of a range of related but nonidentical host receptors. They further suggest that the definition of the YTRY motif should be changed to  $\phi$ TRX, a more rigorous description of this sialic acid-recognition motif given recent findings.

The serine-rich repeat (SRR) adhesins are a family of bacterial cell-surface glycoproteins containing two sequence motifs where serine constitutes ~50% of the sequence (Fig. 1 and Fig. S1). These adhesins follow a modular architecture that initiates with an atypical signal peptide, a short N-terminal serine-rich region, a ligand-binding region (frequently termed “adhesin<sub>BR</sub>”

or “strain<sub>BR</sub>,” e.g. the binding region from *Streptococcus sanguinis* strain SK1 is termed SK1<sub>BR</sub>), a second serine-rich repeat region that varies in length between several hundred and several thousand amino acids with serine as every other residue, e.g. ...SVSASTSASTSASTSAS... , and a cell wall anchoring motif. Fiber diffraction studies suggest that these repeat regions form a spring-like linker that tethers the host binding region to the bacterium (1).

The SRR adhesins are expressed by a variety of Gram-positive commensal and pathogenic bacteria and are broadly distributed (2). A survey of NCBI GenBank<sup>TM</sup> identified over a thousand sequences that may belong to this family. Indeed, all sequenced strains of *Streptococcus gordonii* and *S. sanguinis* encode serine-rich repeat adhesins (3), and homologs have been found in strains of *Streptococcus oralis* and *Streptococcus mitis*, as well as other oral streptococci (2, 4–6). One known functional role of these SRR adhesins is to mediate attachment to protein or glycoprotein receptors, which allows for adherence to host tissues. Accordingly, SRR adhesins have been linked to a variety of infections, including endocarditis, meningitis, and pneumonia (7–12).

Despite the conserved functional organization (Fig. 1), the ligand-binding regions are highly diverse with species-specific trends in the binding region type. Some of the ligand-binding regions of SRR adhesins target glycan structures (13). For example, the SRR adhesins of *S. gordonii* and *S. sanguinis* bind O-linked sialoglycans displayed on mucin-like proteins including salivary glycoprotein MUC7 and platelet glycoprotein GPIb (2, 14, 15). Binding to MUC7 may facilitate oral colonization, whereas interaction with platelet GPIb can allow streptococci to establish endocardial infections (14, 16).

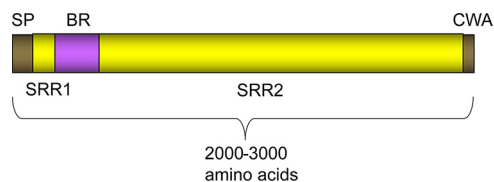
Interaction of the *S. gordonii* and *S. sanguinis* SRR adhesins with host sialoglycan structures relies upon a domain within the binding region related in fold to mammalian sialic acid-binding immunoglobulin-like lectins (Siglecs); indeed both are organized around a V-set Ig fold (6, 16–18). These “Siglec-like” SRR adhesins always contain a second domain immediately following the Siglec domain (16). Termed the Unique domain, this C-terminal region has no counterpart in mammalian Siglecs, and its function remains unknown.

This article contains supporting information.

✂ Author's Choice—Final version open access under the terms of the Creative Commons CC-BY license.

\* For correspondence: T. M. Iverson, tina.iverson@vanderbilt.edu.

## Structure of *S. sanguinis* SK1 adhesin

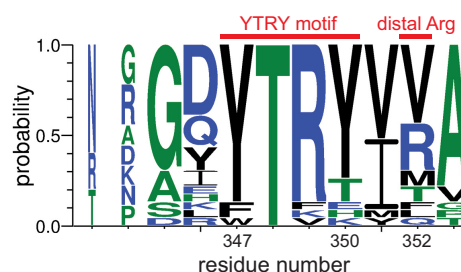


**Figure 1. General organization of SRR adhesin proteins.** SRR adhesins initiate with a ~90-amino acid N-terminal signal peptide (SP) that facilitates trafficking to a specialized glycoprotein transporter known as the accessory Sec system. The serine-rich repeat regions (SRR1 and SRR2) are extensively O-glycosylated in the bacterial cytoplasm prior to transport. The ligand-binding region (BR) varies depending upon the organism and contains structural modules that are highly diverse in sequence, fold, and function, which may provide binding specificity for different bacterial strains (6). The C-terminal cell wall anchor (CWA) includes an LPXTG sequence motif that covalently links the cell wall peptidoglycan.

Despite the conserved fold in the Siglec domain, the binding location for sialoglycans differs between the bacterial Siglec-like adhesins and mammalian Siglecs. The streptococcal Siglec-like adhesins hydrogen-bond with sialic acid via a semiconserved YTRY sequence motif on the F strand of the V-set Ig fold (Fig. 2) (6, 16, 17, 19). Of these, the first Tyr residue contributes only backbone interactions to the ligand. Here, the aromatic side chain faces away from the binding site and is involved in packing interactions that likely contribute to the correct presentation of central Thr-Arg. Thr-Arg makes multiple key side-chain hydrogen-bonding contacts to the sialic acid of host sialoglycans, and therefore the sequence of these central residues appears to be the most important for binding (16, 17). Prior mutagenesis of either the Thr or Arg in characterized Siglec-like adhesins substantially reduces binding to defined, synthetic sialoglycans and to platelets (6, 19, 20). Moreover, isogenic strains of streptococci containing mutations in the YTRY motif exhibit reduced virulence in an animal model (21). The final Tyr of the motif contributes a single hydrogen bond to the subterminal galactose of  $\alpha$ 2,3-sialoglycans and is therefore not involved in sialic acid recognition but may contribute to overall binding affinity of sialoglycans (16).

Recent structural and engineering studies revealed that sialoglycan binding and selectivity are also affected by three adjacent loops of high sequence diversity (18). Using nomenclature from the V-set Ig fold identified them as the CD loop, the EF loop, and the FG loop (16). Amino acid side chains in these loops directly hydrogen-bond with sialoglycan ligands (16, 17, 19), and the sequence diversity of these loops is proposed as a major determinant of sialoglycan selectivity in the Siglec-like adhesins. As a result, they have been termed “selectivity loops” (18).

The only structurally characterized Siglec-like binding region that differs somewhat in the topology of its binding pocket is found in SrpA from *S. sanguinis* strain SK36. In SrpA<sub>BR</sub>, the YTRY sequence is a noncanonical FTTRT but retains the central Thr-Arg important for ligand binding (17, 19). In addition, SrpA<sub>BR</sub> lacks an appreciable FG selectivity loop. Notably, SrpA<sub>BR</sub> contains a second Arg residue outside of the canonical binding sequence motif (Fig. 2) that cooperates with the noncanonical FTTRT sequence to promote sialoglycan binding (17, 19). This residue is not highly conserved in the Siglec-like binding regions (Fig. 2) and is located too far from the FTTRT motif to interact with a bound trisaccharide. However, structures of



**Figure 2. The sialoglycan-binding motif of Siglec-like SRR adhesins.** The conservation of residues is indicated by letter size with the larger letters representing a more strongly conserved residue. The positions of the YTRY motif (positions 347–350) and the distal Arg (position 352) are notated with a red line above the letters. The numbering reflects the residue positions within SK1<sub>Siglec1</sub>. The letters colored blue, green, and black indicate charged, nonpolar, and polar residues, respectively. The adhesins included in the alignment are WP\_125444382.1 from *S. gordonii* strain M99, WP\_046165954.1 from *S. gordonii* strain 72-40, WP\_080889728.1 from *S. gordonii* strain G9B, WP\_046165954.1 from *Streptococcus* sp. strain 1236FAA, WP\_009659981.1 from *Streptococcus* sp. strain AS14, WP\_002906900.1 from *S. sanguinis* strain SK115, WP\_125439128.1 from *S. sanguinis* strain SK150, WP\_125444035.1 from *S. sanguinis* strain SK678, WP\_081102781.1 from *S. gordonii* strain Chalis, WP\_045635027.1 from *S. gordonii* strain UB10712, WP\_080555651.1 from *S. sanguinis* strain SK1, WP\_011836739.1 from *S. sanguinis* strain SK36, WP\_080555852.1 from *S. sanguinis* strain SK408, WP\_000466180.1 from *S. sanguinis* strain SK140, WP\_046165954.1 from *S. sanguinis* strain PS478, WP\_087941957.1 from *S. sanguinis* strain SK1056, WP\_080557024.1 from *S. sanguinis* strain SK330, WP\_080560819.1 from *S. sanguinis* strain SK355, WP\_080555460.1 from *S. sanguinis* strain SK405, WP\_061600538.1 from *S. gordonii* strain SK49, and WP\_000466181.1 from *S. oralis* strain SF100. Although previously termed the YTRY motif, this binding sequence motif is formally defined as  $\phi$ TRX, where  $\phi$  represents Trp, Phe, or Thr, and X represents Tyr, Thr, Glu, His, or Lys.

SrpA<sub>BR</sub> show that the binding pocket is contiguous with this distal arginine both because of the presence of a Thr versus Tyr in the fourth position of the YTRY motif and because of the absence of the FG loop. These alterations extend the glycan-binding site, which may allow the accommodation of either two oriented trisaccharides or larger, branched sialoglycans. Indeed, a disialylated hexasaccharide has been modeled into this site with the distal arginine binding to the second sialic acid of this significantly larger and disialylated ligand (17).

The binding region from *S. sanguinis* strain SK1 (residues 252–660 and termed SK1<sub>BR</sub>) differs from structurally characterized Siglec-containing binding regions in two ways (6). First, it contains two copies of the Siglec and Unique domains in tandem (SK1<sub>Siglec1</sub>-SK1<sub>Unique1</sub>-SK1<sub>Siglec2</sub>-SK1<sub>Unique2</sub>) (Fig. S1) with sequence identity/similarity of 39%/56% between the two Siglec domains and 44%/50% between the two Unique domains. Tandem domains are rarely observed in sequences of the Siglec-like adhesins. Only eight other tandem domain Siglec-like binding regions are identifiable in GenBank<sup>TM</sup>. Seven of the eight SRR adhesins have  $\geq$ 94% identity and  $\geq$ 96% similarity to SK1<sub>BR</sub>; these adhesins are from various strains of *S. sanguinis* and one from *Streptococcus cristatus* (Fig. S1). The other adhesin containing a tandem domain is FapC from *S. oralis* subsp. *dentisani* strain F0392 (Fig. S1) (4), in which the FapC<sub>BR</sub>-binding region exhibits 25% identity and 39% similarity to SK1<sub>BR</sub>. This evolutionary relatedness is consistent with evidence that FapC<sub>BR</sub> may bind sialoglycans and may be important for oral colonization (4). The functional implications for tandem domains have not been explained in the literature.

The second way that SK1<sub>BR</sub> differs from other structurally characterized Siglec-like binding regions is that each of the putative Siglec domains of SK1<sub>BR</sub> contains a noncanonical YTRY sialic acid-binding sequence motif that lacks the central Thr-Arg deemed to be critical for binding in other Siglec-like adhesins (6). This motif in SK1<sub>Siglec1</sub> is YTKY, and the motif in SK1<sub>Siglec2</sub> is YTFK (6). Although it has been shown that SK1<sub>BR</sub> binds sialoglycans (6), it was unknown whether both Siglec domains could bind sialoglycans. If so, what is the relative contribution of each to binding? How do the noncanonical YTRY sequence motifs, particularly in SK1<sub>Siglec2</sub>, maintain the interaction with host receptors?

Here, we present crystal structures of unliganded and sialoglycan-bound SK1<sub>BR</sub>, which show that both SK1<sub>Siglec1</sub> and SK1<sub>Siglec2</sub> adopt V-set Ig folds and that both domains interact with sialoglycan ligands at the noncanonical YTRY motifs. We validate these interactions using binding studies of isolated SK1<sub>Siglec1+Unique1</sub> and SK1<sub>Siglec2+Unique2</sub> and demonstrate that each domain has a distinct selectivity profile for synthetic sialoglycans and glycoprotein ligands. The tandem domains allow increased binding to a host salivary glycoprotein, MUC7, via an avidity effect, possibly by binding simultaneously to two oriented, large glycans. In contrast, platelet binding mainly occurs via SK1<sub>Siglec2</sub>, and binding to the plasma glycoprotein PRG4 (also called lubricin) mainly occurs via SK1<sub>Siglec1</sub>. Taken together, these findings support a mechanism of host interaction in which the tandem domains of the *S. sanguinis* strain SK1 adhesin interact most strongly with a patch of oriented large glycans and indicate that each individual domain differently impacts binding to sialoglycoprotein targets. This adhesin architecture therefore allows for increased flexibility and breadth in the host receptors that are recognized.

## Results

### Structure of *S. sanguinis* SK1<sub>BR</sub>

To develop hypotheses for how SK1<sub>BR</sub> binds to sialoglycans, we began by determining its X-ray crystal structure using molecular replacement methods (Tables 1 and 2). The tandem repeats of unliganded SK1<sub>BR</sub> fold independently, and slight angles between the domains yield an elongated and overall arc shape (Fig. 3). The limits of the domains can be clearly distinguished as SK1<sub>Siglec1</sub> (residues 252–377), SK1<sub>Unique1</sub> (residues 378–453), SK1<sub>Siglec2</sub> (residues 454–573), and SK1<sub>Unique2</sub> (residues 574–660).

As is anticipated from the amino acid sequence conservation, the individual Siglec domains and Unique domains exhibit structural similarity, with RMS deviations in C $\alpha$  position of 1.058 Å between SK1<sub>Siglec1</sub> and SK1<sub>Siglec2</sub> and 0.639 Å between SK1<sub>Unique1</sub> and SK1<sub>Unique2</sub> (Fig. 4). To evaluate the basis for the higher overall RMS deviations in C $\alpha$  position of the Siglec domains, we overlaid SK1<sub>Siglec1</sub> and SK1<sub>Siglec2</sub> and identified whether this was a global difference or whether the structural difference was localized to specific regions. We identified disproportionately large structural deviations in the CD, EF, and FG selectivity loops that surround the putative sialoglycan-binding pockets (Fig. 4A), which contain different numbers of residues, with the FG loop of SK1<sub>Siglec2</sub> being so short that it is

**Table 1**

**Diffraction data collection statistics for the liganded and unliganded SRR-binding region from *S. sanguinis* strain SK1**

The values in parentheses are statistics for the highest resolution shell.

	None	sTa	3'sLn
<b>SBGrid Entry</b>	756	754	755
Resolution (Å)	2.00	1.55	2.10
Highest resolution shell (Å)	2.00–2.07	1.55–1.58	2.10–2.18
<b>Data collection</b>			
Beamline	APS 21-ID-F	APS 21-ID-F	APS 21-ID-F
Wavelength (Å)	0.97872	0.97946	0.97946
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2	P2 <sub>1</sub> 2 <sub>1</sub> 2	P2 <sub>1</sub> 2 <sub>1</sub> 2
Unit cell dimensions (Å)			
<i>a</i>	82.213	83.498	81.549
<i>b</i>	269.859	271.85	271.063
<i>c</i>	47.511	47.815	46.849
<i>R</i> <sub>sym</sub>	0.13 (0.85)	0.067 (0.615)	0.086 (0.514)
<i>R</i> <sub>pim</sub>	0.045 (0.285)	0.035 (0.402)	0.045 (0.311)
<i>I</i> / $\sigma$	25.78 (3.38)	26.6 (1.9)	16.26 (1.51)
Completeness (%)	100 (99.8)	93.1 (74.4)	92.34 (67.30)
Redundancy	10.0 (9.9)	4.3 (2.8)	4.3 (3.3)
CC <sub>1/2</sub>	0.910 (0.706)	0.997 (0.723)	0.999 (0.727)

effectively absent. The maximal displacement of these loops in overlays is 10.9 Å for the CD loop, 3.2 Å for the EF loop, and 8.4 Å for the FG loop. In the CD loop, this also manifests as a difference in secondary structure in which the CD loop of SK1<sub>Siglec1</sub> lacks secondary structure, whereas the CD loop of SK1<sub>Siglec2</sub> folds into an  $\alpha$ -helix. In addition, the CD loop of SK1<sub>Siglec2</sub> is displaced from the binding pocket when compared with the CD loop of SK1<sub>Siglec1</sub>. Together, these differences in loop length and structure result in a larger and more open binding site as compared with SK1<sub>Siglec1</sub>. This large binding pocket of SK1<sub>Siglec2</sub> is somewhat reminiscent of the binding pocket in the SrpA adhesin from *S. sanguinis*, which also lacks the FG loop (15–17).

### Structures of *S. sanguinis* SK1<sub>BR</sub> bound to sialoglycans

Prior glycan array analysis identified that SK1<sub>BR</sub> can bind to a broad range of defined, synthetic sialoglycan ligands. It was not clear, however, whether SK1<sub>Siglec1</sub> and SK1<sub>Siglec2</sub> are both important for these interactions or how the noncanonical YTRY motifs support glycan binding. We therefore determined cocrystal structures of SK1<sub>BR</sub> soaked with either 10 mM sialyl T-antigen (sTa), a core 1 glycan that can be conjugated to Ser or Thr residues of glycoproteins, or 10 mM 3'-sialyl-N-acetyllactosamine (3'sLn), a trisaccharide that can be a component of larger, branched glycans. For both sTa and 3'sLn, we observed the appearance of unambiguous electron density adjacent to the noncanonical YTRY motif of both SK1<sub>Siglec1</sub> and SK1<sub>Siglec2</sub>. We were able to model sTa and 3'sLn with confidence into this electron density, and the hydrogen bond networks observed are consistent with specific binding (Figs. 5, A–D, and 6, A–D). Thus, SK1<sub>Siglec1</sub> and SK1<sub>Siglec2</sub> are both capable of binding sialoglycan ligands and can do so simultaneously. Local conformational changes observed upon binding were slight. When each domain is individually aligned to the corresponding unliganded domain, the RMS deviations in the C $\alpha$  positions is <0.3 Å (Table S1). The biological significance of conformational changes of this magnitude cannot be determined.

As compared with adhesins with canonical YTRY motifs (16, 18, 19), the sialic acid of each trisaccharide interacts with the

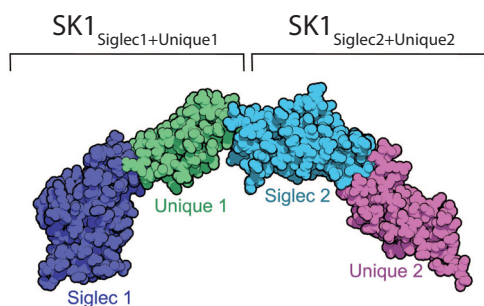
## Structure of *S. sanguinis* SK1 adhesin

**Table 2**

Refinement statistics for the liganded and unliganded SRR-binding region from *S. sanguinis* strain SK1

The Ramachandran statistics were obtained using the MolProbity output of Phenix (40). ASU, asymmetric unit.

	None	sTa	3sLn
<b>PDB entry</b>	6VS7	6VT2	6VU6
<b>Model content (per ASU)</b>			
Protein molecules	2	2	2
Glycans	4	4	4
Water molecules	945	1633	317
Ions	14	13	8
Other solvent	13	1	0
<b>Refinement</b>			
$R_{\text{cryst}}$	0.211	0.176	0.225
$R_{\text{free}}$	0.240	0.193	0.248
RMS deviation			
Bond lengths (Å)	0.015	0.022	0.018
Bond angles (°)	1.25	1.65	1.78
Ramachandran (%)			
Favored	97.05	98.4	97.3
Allowed	2.83	1.6	2.7
Outliers	0.12	0	0
Mean $B$ factors			
Protein (Å <sup>2</sup> )	34.90	15.52	33.57
Glycans (Å <sup>2</sup> )		20.93	49.26

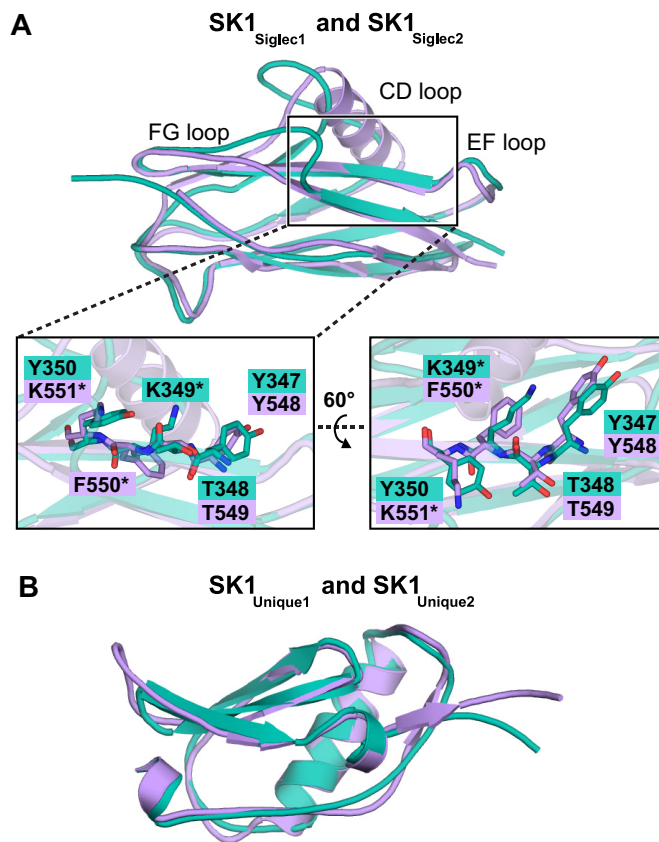


**Figure 3. Structure of the binding region of unliganded SK1.** SK1<sub>BR</sub> has four domains in the order SK1<sub>Siglec1</sub>, SK1<sub>Unique1</sub>, SK1<sub>Siglec2</sub>, and SK1<sub>Unique2</sub> depicted from left to right and colored by domain. The domain repeats, named SK1<sub>Siglec1+Unique1</sub> and SK1<sub>Siglec2+Unique2</sub>, are homologous but not identical.

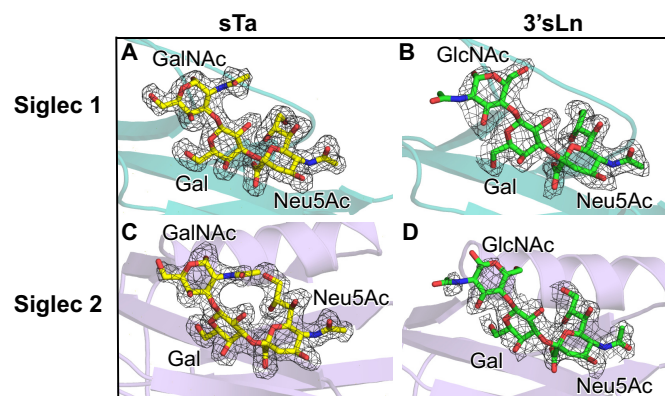
two noncanonical motifs of SK1 via fewer hydrogen-bonding interactions (Fig. 6, A–E). The contacts that are similar between SK1 and structurally characterized adhesins with YTRY motifs include interactions with the backbone of the N-terminal Tyr and side chain of the Thr residues in both noncanonical YTRY motifs (Fig. 6). Hydrogen-bonding to sialic acid additionally occurs via a backbone carbonyl in the EF loop (SK1<sup>G344</sup> and SK1<sup>G545</sup>) (Fig. 6, A–D). Although not previously reported for the SRR adhesins, this interaction is conserved across other adhesins (GspB<sup>I479</sup>, SrpA<sup>R342</sup>, and Hsa<sup>K335</sup>) (Fig. 6E) (6, 17, 18).

SK1<sub>Siglec1</sub> hydrogen-bonds to the sialic acid and galactose of both sTa and 3'sLn via the YTKY sequence (Fig. 6, A and B). Here, the Arg → Lys substitution eliminates one side-chain hydrogen bond (Fig. 6A and Figs. S2A and S3A), but the overall interaction remains similar to that supported by a canonical YTRY motif, like Hsa from *S. gordonii* strain Challis or GspB from *S. gordonii* strain M99 (Figs. S2B and 6E) (16).

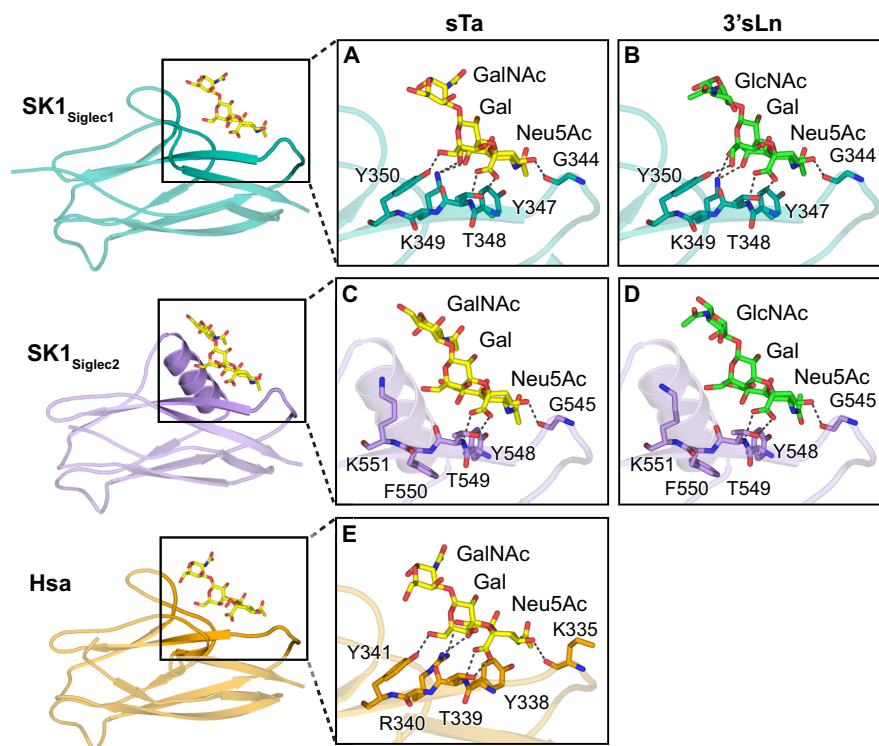
In contrast, the Phe and Lys residues in the SK1<sub>Siglec2</sub> YTFK sequence do not form hydrogen-bond contacts with the ligand.



**Figure 4. Structural comparison of the tandem Siglec and Unique domains.** A, and B, cartoon diagrams of unliganded SK1<sub>BR</sub>. SK1<sub>Siglec1</sub> and SK1<sub>Unique1</sub> are colored in teal, and the SK1<sub>Siglec2</sub> and SK1<sub>Unique2</sub> are colored in lavender. A, 98 eight atoms were aligned, and 26 were rejected after 5 cycles to output an RMS deviation of 1.058 Å. The noncanonical binding motifs of SK1<sub>Siglec1</sub> and SK1<sub>Siglec2</sub> are shown as sticks, and residues that deviate from the canonical YTRY motif definition are noted with an asterisk. B, 72 were aligned, and 3 were rejected after 2 cycles to output an RMS deviation of 0.639 Å.



**Figure 5. Ligand electron density.** A–B, SK1<sub>Siglec1</sub> is shown in teal cartoon. C–D, SK1<sub>Siglec2</sub> is shown in lavender cartoon. A, C, sTa and B, D, 3'sLn are shown as yellow and green sticks, respectively. Oxygen atoms are colored in red, and nitrogen atoms are in blue. Ligands were manually placed in Coot after refinement of the protein and prior to solvent placement. The ligands were then refined with rigid-body and real space refinement in Coot prior to solvent placement and final structure refinements.  $|F_o| - |F_c|$  electron density maps were calculated from coordinates that had been refined in Phenix (40) for three rounds after the removal of the sialoglycans from the model. Maps are contoured at  $3\sigma$  and are shown in dark gray mesh.



**Figure 6. SK1 interactions with ligands.** The structures of the Siglec domains of A–B, SK1<sup>Siglec1+Unique1</sup>, C–D, SK1<sup>Siglec2+Unique2</sup>, and E, Hsa (PDB entry 6EFD) (18) are shown in cartoon in teal, lavender, and orange, respectively. The adhesin residues that hydrogen bond with the ligands are shown as sticks. Hydrogen bonds between the adhesins and ligands are shown as dark gray dashed lines. The ligands sTa and 3'sLn are shown as yellow and green sticks, respectively. Oxygen and nitrogen atoms are colored red and blue, respectively.

In addition, the YTFK sequence of SK1<sup>Siglec2</sup> does not hydrogen-bond to the galactose or the variable third sugar, GalNAc/GlcNAc, of sTa/3'sLn (Fig. 6, C and D). Compensating for this, side chains in the helical CD loop make additional hydrogen-bond contacts (Figs. S2, C and D, and S3, B and C). Both observations indicate that sialic acid recognition by SK1<sup>Siglec2</sup> differs from characterized ligand interactions in the GspB, Hsa, and SrpA adhesins (16, 18, 19).

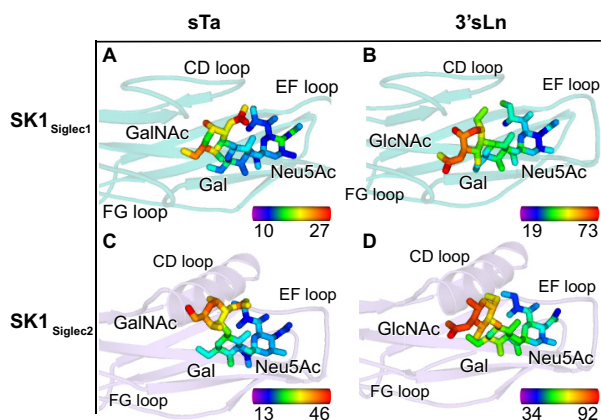
This reduced number of binding contacts is reflected in temperature factor analysis of the ligand. Crystallographic temperature factors can give a rough estimate of inherent mobility Fig. 7 and Fig. S4. It is important to note that the CD, EF, and FG loops, the YTRY motifs, and the ligands do not participate in crystal contacts in any of the three structures; as a result, the temperature factors are not influenced by crystal packing interactions. Here, the low temperature factors of sTa and SK1<sup>Siglec1</sup> suggest that the ligand has little mobility when bound, which may be interpreted as strong binding (Fig. 7A). In contrast, the higher temperature factors of 3'sLn bound to SK1<sup>Siglec1</sup> or of either sTa or 3'sLn bound to SK1<sup>Siglec2</sup> suggest higher mobility, which suggests that these could be lower-affinity ligands (Fig. 7, B–D). Nevertheless, in all cases, the temperature factor is the lowest at the sialic acid and increases over the length of the trisaccharide, consistent with lower ligand mobility at the sialic acid and increased mobility at the reducing end sugar.

We then performed the converse analysis, assessing the temperature factors of each Siglec domain near the ligand (Fig. 8).

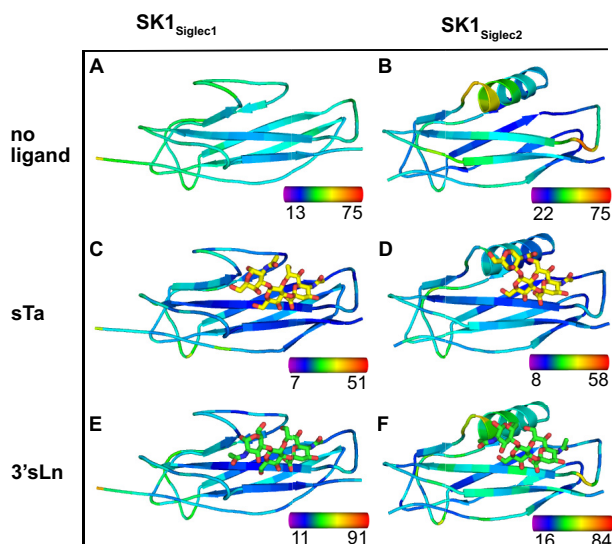
In the unliganded state, we observe that the selectivity loops have elevated temperature factor values as compared with the V-set Ig fold that forms the core of the Siglec domain, suggesting that they have increased mobility (Fig. 8, A and B). Upon ligand binding, the temperature factors of all loops become more similar to the temperature factors of the Siglec domain core in both SK1<sup>Siglec1</sup> and SK1<sup>Siglec2</sup> (Fig. 8, compare A and B with C–F). Moreover, the temperature factors of the YTFK motif in SK1<sup>Siglec1</sup> decreased upon ligand binding (Fig. 8, compare A, C, and E). The temperature factors of the YTFK motif in SK1<sup>Siglec2</sup> decrease upon sTa binding but do not decrease upon 3'sLn binding (compare Fig. 8, compare B, D, and F). This stabilization is more pronounced in the sTa-bound SK1<sup>Siglec2</sup> than in the 3'sLn-bound SK1<sup>Siglec2</sup>, possibly because of the additional hydrogen bond between sTa and the CD helix. This extra hydrogen bond further links the YTRY region and the CD helix, decreasing the mobility of both regions (Fig. S2, C and D). Together, this analysis suggests that the binding of ligands stabilizes the positions of the selectivity loops in both domains. This is consistent with prior molecular dynamics simulations of the Hsa adhesin from *S. gordonii* strain Challis, which suggested that these selectivity loops can adjust to optimize the interaction to ligands (18).

Intriguingly, the interactions between sialic acid of sTa or 3'sLn and the YTRY motif contain structural parallels to staphylococcal superantigen-like protein SSL5 bound to sialyl Lewis<sup>x</sup> (Fig. 9) (22). Prior comparisons of SSL5 with a range of evolutionarily unrelated sialic acid-binding proteins suggested a

## Structure of *S. sanguinis* SK1 adhesin

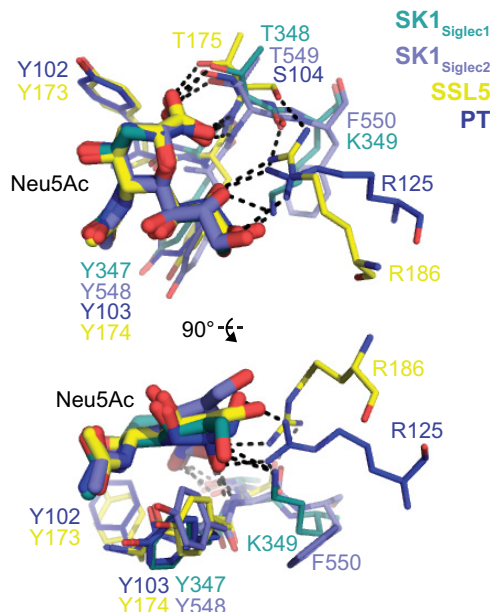


**Figure 7. Temperature factor analysis of Siglec domains and bound ligands.** A–D, SK1<sub>Siglec1</sub> is shown in teal cartoon, and SK1<sub>Siglec2</sub> is shown in lavender cartoon. The sTa liganded structures are shown in A and C, and the 3'sLn liganded structures are shown in B and D. Both sTa and 3'sLn are shown as sticks and colored by temperature factor, where blue represents a low temperature factor, and red represents a high temperature factor as depicted by the scale in the bottom right corner of each panel. The scale values are in Å<sup>2</sup>.



**Figure 8. Siglec domain colored by temperature factor.** A–B, SK1 unliganded, (C–D) sTa-bound, and E–F 3'sLn-bound are shown in cartoon and colored by temperature factor, where blue represents a low temperature factor, and red represents a high temperature factor. The color bars in the bottom right corner of each panel indicate the ranges of B factors in Å<sup>2</sup>. sTa and 3'sLn are shown as sticks. The oxygen and nitrogen atoms are colored red and blue, respectively.

common sialic acid recognition motif that contains two features: 1) a YY(T/S) motif on an edge strand of a  $\beta$  sheet and 2) an arginine distant in sequence but spatially adjacent to the YY (T/S) motif (22). Despite the somewhat different sequence elements in the sialic acid-binding motif of SK1<sub>BR</sub> and other SRR adhesins, the hydrogen-bonding pattern to the sialic acid is similar to that of SSL5 (Fig. 9). The arginine, or lysine in the case of SK1<sub>Siglec1</sub>, associated with sialic acid binding is provided from within the YTRY motif of SK1<sub>Siglec1</sub> (Lys<sup>349</sup>) and other Siglec-like adhesins, whereas it is outside of the YY(T/S) motif in SSL5 (Arg<sup>186</sup>) and PT (Arg<sup>125</sup>) and other sialic acid-binding proteins. The interactions between the



**Figure 9. Comparison of bacterial sialic acid-binding pockets.** Shown here is an overlay of the two Siglec domains of SK1<sub>BR</sub> with staphylococcal superantigen-like protein 5 (SSL5; PDB entry 2R61) and pertussis toxin (PT; PDB entry 1PTO) (48), two proteins in which the sialic acid recognition motif has previously been identified (22).

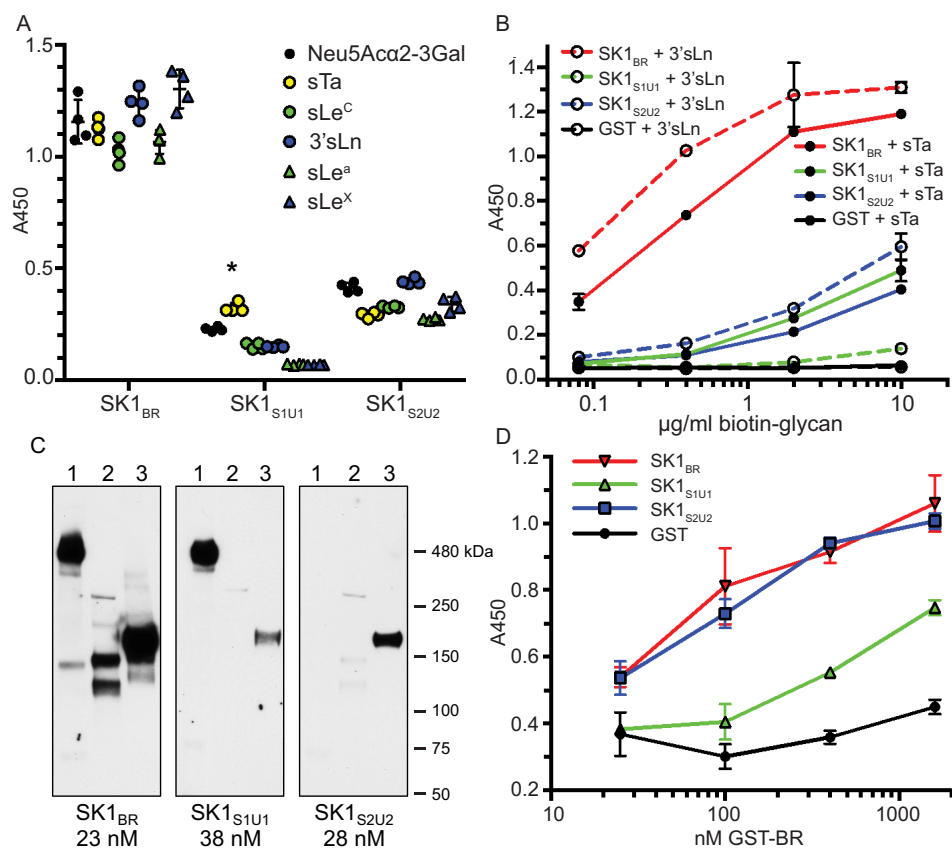
binding pocket and sialic acid may be a product of convergent evolution (22).

### SK1<sub>Siglec1</sub> and SK1<sub>Siglec2</sub> have unique selectivity profiles and exhibit synergistic binding

Based upon our observation that both SK1<sub>Siglec1</sub> and SK1<sub>Siglec2</sub> bind sialoglycans in a cocrystal structure, we tested the relevance of these interactions in binding to defined, synthetic glycans. To do this, we developed recombinant GST S-transferase (GST)-tagged proteins (Fig. S5) containing either the first Siglec and Unique domains, SK1<sub>Siglec1</sub>+Unique1 (SK1<sup>252–455</sup>), or the second Siglec and Unique domains, SK1<sub>Siglec2</sub>+Unique2 (SK1<sup>449–660</sup>) and compared the binding of these isolated binding modules to that of full-length SK1<sub>BR</sub>.

We began by evaluating how each repeat bound to a small library of tri- and tetrasaccharide sialoglycans at a single concentration of ligand (Fig. 10A). Both SK1<sub>Siglec1</sub>+Unique1 and SK1<sub>Siglec2</sub>+Unique2 bound to at least some of the tested sialoglycans, albeit less strongly than did full-length SK1<sub>BR</sub>. Consistent with the crystallographic temperature factor analysis (Fig. 7), SK1<sub>Siglec1</sub>+Unique1 showed a statistically significant preference for sTa. In contrast, SK1<sub>Siglec2</sub>+Unique2 appears to be more broadly selective. The latter finding is consistent with the observation that the SK1<sub>Siglec2</sub> domain does not make hydrogen-bonding contacts to the second and third sugars of trisaccharide in the crystal structures (Fig. 6, C and D).

We then performed a more detailed dose-dependent binding to sTa and 3'sLn (Fig. 10B). This analysis indicates that both SK1<sub>Siglec1</sub>+Unique1 and SK1<sub>Siglec2</sub>+Unique2 bind to sTa ~100-fold less strongly than the full-length SK1<sub>BR</sub>, which contains the two subdomains in tandem. This suggests the possibility of binding synergy when in the presence of high concentrations of



**Figure 10. Binding of SK1<sub>BR</sub> and split variants to glycans and glycoproteins.** A, biotin-glycan binding to immobilized GST-tagged SK1<sub>BR</sub> and the split binding modules ( $n = 4$  technical replicates). The asterisk indicates binding that was significantly greater than the level of binding to all other glycans in the set of six ( $p < 0.05$  using a two-way analysis of variance with Tukey's correction for multiple comparisons). B, binding of biotinylated sTa or 3'sLn to immobilized GST-tagged SK1<sub>BR</sub> and split constructs ( $n = 3$  technical replicates). C, binding of GST-tagged SK1<sub>BR</sub> and split constructs to glycoproteins in human plasma (lane 1), platelet lysate (lane 2), or submandibular sublingual saliva (lane 3). D, binding of GST-tagged SK1<sub>BR</sub> and SK1<sub>BR</sub> deletion constructs to immobilized human platelets ( $n = 3$  technical replicates). In A, B, and D, mean values  $\pm$  standard deviation are indicated. In cases where error bars are not evident, the deviations were smaller than the size of the symbol used for the data point. Background values for GST alone were not subtracted but are shown in B and D.

carbohydrate. Physiologically, small portions of large and complex branched glycans could have a similar appearance as high local concentrations in a binding assay. Such high levels of a particular glycan are hypothesized for glycoproteins that contain oriented glycan patches (23).

#### Synergistic properties of SK1<sub>BR</sub> affect binding to host receptors

We next evaluated whether binding to the synthetic glycans explains how SK1<sub>BR</sub> interacts with host glycoproteins. Previously identified ligands for the Siglec-like SRR adhesins are consistent with their biological roles in oral commensalism and the pathogenesis of infective endocarditis. These include salivary MUC7, platelet GPIIb $\alpha$ , and several O-glycosylated plasma glycoproteins. We therefore evaluated the interactions of GST-tagged SK1<sub>BR</sub>, SK1<sub>Siglec1+Unique1</sub>, and SK1<sub>Siglec2+Unique2</sub> with human salivary, platelet, and plasma glycoprotein targets via Far Western blotting. Isolated binding regions SK1<sub>Siglec1+Unique1</sub> and SK1<sub>Siglec2+Unique2</sub> both bound modestly to MUC7 compared with binding by the tandem domains of SK1<sub>BR</sub> (Fig. 10C). SK1<sub>Siglec1+Unique1</sub> readily bound a 460-kDa plasma protein, whereas SK1<sub>Siglec2+Unique2</sub> did not. Here, the tandem domains of SK1<sub>BR</sub> did not increase the binding over what was observed for SK1<sub>Siglec1+Unique1</sub>. Neither

SK1<sub>Siglec1+Unique1</sub> nor SK1<sub>Siglec2+Unique2</sub> bound appreciably to GPIIb $\alpha$  in the platelet lysate.

We also assessed binding to fixed, immobilized platelets (Fig. 10D). SK1<sub>Siglec1+Unique1</sub> bound weakly, whereas SK1<sub>Siglec2+Unique2</sub> bound more strongly. There was not a cooperative effect of linking these domains, because the binding of SK1<sub>BR</sub> to platelets could be fully explained by the binding of SK1<sub>Siglec2+Unique2</sub>. These results suggest that there is a high-affinity ligand for SK1<sub>BR</sub> on intact platelets that is due primarily to binding by SK1<sub>Siglec2+Unique2</sub>. The combined results indicate that SK1<sub>BR</sub> can bind multiple simple and complex sialoglycan ligands on biological targets via a combination of interactions.

#### Discussion

All previously determined structures of the binding regions of Siglec-like adhesins have a single Siglec domain and a single Unique domain. This prior work has shown that sialic acid-binding affinity largely stems from binding to a YTRY motif with the selectivity tuned via adjacent loop regions of the Siglec domain (18). Our data are consistent with each repeat of SK1<sub>BR</sub> following the same principles for binding and selectivity. Specifically, the sialoglycan binds via specific interactions between sialic acid and the noncanonical YTRY motif. Given this finding



## Structure of *S. sanguinis* SK1 adhesin

in conjunction with studies of SrpA, the YTRY motif can be more formally defined as  $\phi$ TRX, where  $\phi$  represents Trp, Phe, or Tyr, and X represents Tyr, Thr, Glu, His, or Lys (Fig. 2) (17, 19). This  $\phi$ TRX motif represents a sequence modification of the YY(S/T) motif found in other sialic acid-binding proteins, with  $\phi$ T of the  $\phi$ TRX corresponding to the YT of YY(S/T). As a result, both sequence motifs interact with sialic acid via a similar pattern of hydrogen bonds. For the Siglec-like adhesins, the selectivity loops may control the identity of the preferred sialoglycan. In SK1<sub>Siglec1</sub> and SK1<sub>Siglec2</sub>, these selectivity loops differentially impact the size of the binding pocket and the orientation of the ligand, resulting in unique selectivity profiles.

Although the  $\phi$ TRX motif and selectivity loops of SK1<sub>BR</sub> support our current understanding of sialoglycan binding, two unique features highlight unanswered questions regarding the link between sialoglycan binding and adhesion. For instance, SK1<sub>Siglec1+Unique1</sub> exhibited selectivity for sTa but had a lower affinity for glycoproteins than SK1<sub>Siglec2+Unique2</sub>. In contrast, SK1<sub>Siglec2+Unique2</sub> binds well to host components but poorly to purified trisaccharides. It is possible that the trisaccharides tested do not include the full biological ligand. SrpA similarly has a larger and more open binding pocket that could possibly accommodate a larger ligand (Fig. 11C). Hypotheses for the biological ligand can be developed by considering parallels to the SrpA adhesin from *S. sanguinis* strain SK36. Like SK1<sub>Siglec2+Unique2</sub>, SrpA binds poorly to purified tri- and tetrasaccharides *in vitro*, but SrpA binds robustly to human platelets. Lacking an FG loop, SrpA has a significantly larger binding site than many other SRR adhesins. This increased size of the binding site opens to a distal arginine residue (Fig. 11, C and F). For this reason, it has been proposed that physiological targets of SrpA may include a disialylated hexasaccharide or a patch of multiple, oriented glycans (17). Either possibility for the ligand could promote cooperativity that would be expected to increase adhesion to host targets. Like SrpA, SK1<sub>Siglec2</sub> similarly has a large sialoglycan-binding site because of the small size of the FG loop (Figs. 11B and 12A). We propose that SK1<sub>Siglec2</sub> may therefore also bind a “core 2” sialoglycan or patch of oriented glycans (Fig. 12, b and c). In contrast, the distal arginine residue in SK1<sub>Siglec1</sub> is occluded from the sialoglycan-binding site by a larger FG loop, likely prohibiting interactions with a sialoglycan longer than a trisaccharide (Figs. 11A and 12A); however, binding of multiple glycans is still a possibility. Alternatively, a large binding site alone may be sufficient for some adhesins to bind larger saccharides. Hsa also has a large and open binding site but with no distal arginine (Fig. 11, D and F). This is consistent with the ability of Hsa to bind both trisaccharides and larger, branched sialoglycans (2, 6, 24).

SK1<sub>BR</sub> also demonstrates how individual binding domains *versus* tandem linkage can differentially contribute to the affinity for host targets. Conceptually, tandem linking of these two binding regions would be expected to confer cooperative binding capability in binding to some biological targets (25). Or to put it another way, if a target protein contains glycan modifications that are correctly spaced and oriented, this adhesin could bind more strongly to host receptors via multivalent binding (Fig. 12, B and C) (26). Synergy was indeed observed between SK1<sub>Siglec1+Unique1</sub> and SK1<sub>Siglec2+Unique2</sub> when binding to platelet lysate or to salivary

glycoproteins (Fig. 10C). This suggests that both binding regions contribute to adherence for certain host targets and could be explained by binding to either a large branched glycan structure or a patch of clustered glycans (Fig. 12, B and C).

On the other hand, binding of SK1<sub>Siglec1+Unique1</sub> to human plasma is roughly equivalent to that of SK1<sub>BR</sub>, suggesting that SK1<sub>Siglec1+Unique1</sub> has a high-affinity sialoglycan target in human plasma, but SK1<sub>Siglec2+Unique2</sub> does not. SK1<sub>Siglec1+Unique1</sub> may be responsible for adherence to human plasma, consistent with an SK1<sub>Siglec1</sub>-dependent binding model (Figs. 10C and 12D). SK1<sub>Siglec2+Unique2</sub> seems to be solely responsible for adherence to immobilized platelets, suggesting an SK1<sub>Siglec2</sub>-dependent binding mode (Figs. 10D and 12E). The use of tandem repeats and multivalent binding capabilities could confer two distinct evolutionary advantages. First, tandem repeats can be separately mutated for a faster evolution. For example, if an adhesin contains two binding regions following a gene duplication event (27), the domains could then be individually mutated, with each domain conferring different selectivity for host receptors. Divergent evolution of the two binding domains could effectively double the evolutionary rate, leading to faster adaptation. Tandem linkage of binding domain modules could also allow an individual binding domain to evolve through an intermediate with lower affinity and broader specificity. This could allow for the evolution of larger changes in selectivity (28–30).

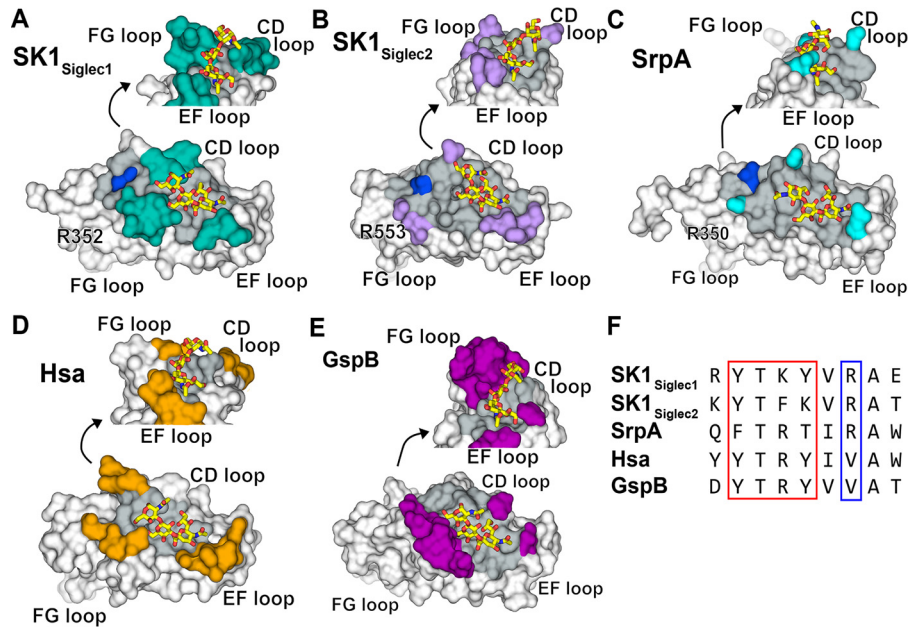
Second, the combined action of two binding domains could allow binding of this adhesin to a broader range of targets. Linking an sTa-specific domain with a domain of another or broader selectivity could allow adherence to either platelets or to other host targets. Increasing the range of ligands an adhesin can bind could increase tropism and allow bacteria to migrate from one tissue to another. Avidity and affinity optimization of protein scaffolds for recognition of on- and off-target biomolecules can increase the specificity of cellular targeting (31). This heteromultivalent binding could be important for increasing specificity for a target tissue (26, 32, 33).

Taken together, the findings reported here get us closer to addressing unanswered questions in the field. The finding that noncanonical motifs of SRR adhesins interact robustly with sialoglycan ligands reveals them as important for host interaction. More importantly, these results provide insight into how adhesive proteins adapt to various biological niches with different host receptors and provide evidence for adhesion to patches of oriented glycans (24). This feeds into ongoing work that seeks to develop a predictive model for streptococcal pathogenicity.

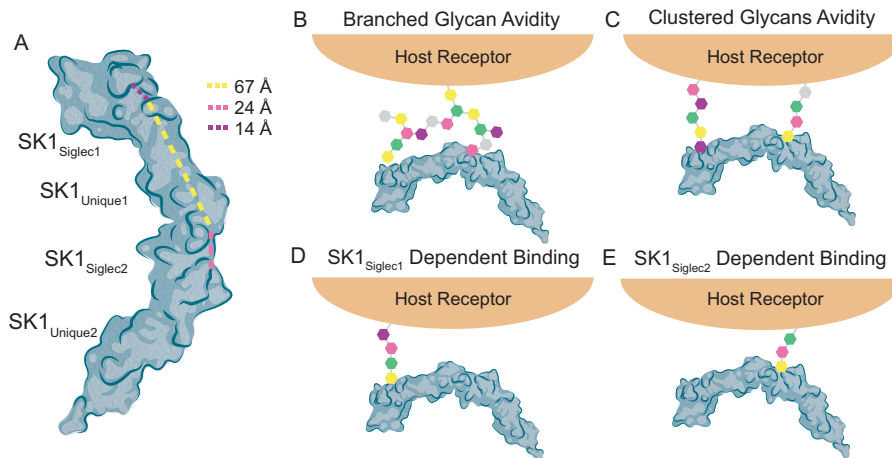
## Experimental procedures

### Expression and purification of SK1<sub>BR</sub>

DNA encoding residues 252–660 of the full-length SK1 adhesins, termed SK1<sub>BR</sub>, was cloned into the pBG101 vector (Vanderbilt), which encodes an N-terminal His<sub>6</sub>-GST affinity tag followed by a cleavage sequence for the 3C precision protease. SK1<sub>BR</sub> was expressed in *Escherichia coli* BL21(DE3) cells. The cells were grown in LB at 37 °C to an  $A_{600\text{ nm}}$  of ~0.6 and expression induced for 4 h with 0.5 mM isopropyl  $\beta$ -D-thiogalactopyranoside. The cells were harvested by centrifugation at  $9220 \times g$  for 15 min. The pellets were resuspended in 250 ml of



**Figure 11. SRR adhesins binding pocket size comparison.** A–E, the adhesins are shown in surface representations (16–18). The top image of the binding site is rotated 80° around the z axis and 70° around the x axis. The binding pocket of each is colored in gray, and the portions of the CD, EF, and FG loops that create the walls of the binding pocket are colored in teal, lavender, cyan, orange, and magenta for SK1<sub>Siglec1</sub>, SK1<sub>Siglec2</sub>, SrpA, Hsa, and GspB, respectively. The Arg distal to the YTRY motif is colored blue. Sialyl T antigen is shown as yellow sticks bound to each adhesin. F, multiple sequence alignment of the above adhesins is shown. The YTRY motif is outlined in red, and the distal Arg is outlined in blue.



**Figure 12. Model of target-specific effects of SK1.** A, the distance between the two binding sites is 67 Å, measured from the C<sub>α</sub> atoms of Tyr-347, the first residue in the YTRY motif of SK1<sub>Siglec1</sub>, and Arg-553, the distal arginine residue of SK1<sub>Siglec2</sub>. The length of the binding site in SK1<sub>Siglec1</sub> is 14 Å, measured from the C<sub>α</sub> atoms of Gly-344 and Lys-349. The length of the binding site in SK1<sub>Siglec2</sub> is 24 Å, measured from the C<sub>α</sub> atoms of Gly-545 to Arg-553. B–E, note that the glycans shown here are only meant to serve as a hypothetical glycan structure and are not meant to represent a specific, defined glycan target. B, multivalent binding of a branched glycan. Given the distance between the two binding sites, SK1<sub>BR</sub> could bind the same branched glycan with both SK1<sub>Siglec1</sub> and SK1<sub>Siglec2</sub>. C, multivalent binding of a patch of clustered glycans. Each binding site of SK1<sub>BR</sub> binds a separate glycan structure. D, SK1<sub>Siglec1</sub> dominant binding. E, SK1<sub>Siglec2</sub> dominant binding. Given the openness of the binding site and the presence of the distal Arg, SK1<sub>Siglec2</sub> may bind a short saccharide, two short saccharides, or a longer hexasaccharide.

lysing buffer (20 mM Tris-HCl, 150 mM NaCl, pH 7.6) supplemented with 2 μg/ml pepstatin, 2 μg/ml leupeptin, 1 μg/ml DNase, and 1 mg/ml lysozyme. The cells were lysed by sonication. The lysate was clarified by centrifugation at 38,465 × *g* for 1 h, and then the supernatant was filtered (0.45 μm) and loaded onto a 5-ml His Trap column. SK1<sub>BR</sub> was eluted with 75 mM imidazole elution buffer (20 mM Tris-HCl, 150 mM NaCl, 75 mM imidazole, pH 7.6). The N-terminal His<sub>6</sub>-GST tag was cleaved with 3C precision protease (2 mg/ml SK1<sub>BR</sub>, 20 mM Tris, 150 mM NaCl, 75 mM imidazole, pH 7.6). The imidazole was diluted, and

the protein was concentrated using a 30-kDa molecular-weight-cutoff centrifugal concentrator. Concentrated protein was passed through a His Trap column to remove the cleaved His<sub>6</sub>-GST tag. The protein was then purified by Superdex 200 size exclusion column (20 mM Tris-HCl, pH 7.6, 150 mM NaCl). Bradford assay was used to determine the final protein concentration.

#### Synthesis of *sTa* trisaccharide Neu5Acα2–3Galβ1–3GalNAc

Galβ1–3GalNAc (34) (30 mg, 0.078 mmol), Neu5Ac (37 mg, 0.117 mmol), and cytidine 5'-triphosphate (CTP) (66 mg, 0.117

## Structure of *S. sanguinis* SK1 adhesin

mmol) were dissolved in a solution containing Tris-HCl buffer (8.0 ml, 100 mM, pH 8.5) and MgCl<sub>2</sub> (20 mM). *Neisseria meningitidis* cytidine 5'-monophosphate-sialic acid (CMP-sialic acid) synthetase (35) (1.0 mg) and *Pasteurella multocida* sialyltransferase 1 M144D (36) (1.5 mg) were then added. The reaction was carried out by incubating the reaction mixture in an incubator shaker at 37 °C for 12 h. The reaction was monitored by TLC (EtOAc/MeOH/H<sub>2</sub>O/HOAc = 4:2:1:0.1, by volume) with *p*-anisaldehyde sugar staining and MS. When an optimal yield was achieved, the same volume (8 ml) of prechilled ethanol was added, and the mixture was incubated at 4 °C for 30 min. The sample was centrifuged, and the precipitates were removed. The supernatant was concentrated, passed through a BioGel P-2 gel filtration column, and eluted with water to obtain the partially purified product. Further purification was achieved using silica gel chromatography (EtOAc/MeOH/H<sub>2</sub>O = 4:2:1, by volume) and a final pass through of a P-2 gel filtration column to produce pure sTa (47 mg, 86%). NMR data were in agreement with those reported previously (37).

### Crystallization and structure determination

SK1<sub>BR</sub> (72 mg/ml in 150 mM NaCl, 20 mM Tris, pH 7.6) was crystallized using the hanging-drop vapor-diffusion method at 25 °C using a reservoir solution containing 20% (w/v) PEG 3350, 0.2 M MgSO<sub>4</sub>, 0.01 M SrCl<sub>2</sub>. The crystals were harvested 1 week later. To obtain the sialoglycan-bound SK1<sub>BR</sub> structures, crystals of the unliganded SK1<sub>BR</sub> were grown for 1 week, removed from the mother liquor, and placed in a new drop containing all of the crystallization components (20% (w/v) PEG 3350, 0.2 M MgSO<sub>4</sub>, 0.01 M SrCl<sub>2</sub>) and either 20 mM sTa (synthesized in-house) or 20 mM 3'sLn (Glycotech). The crystals were allowed to incubate with the glycan for 1 h at room temperature before harvesting.

The crystals were cryoprotected with 40% (1:1 ethylene glycol:glycerol) and 60% reservoir solution and then cryocooled by plunging in liquid nitrogen. X-ray diffraction data were collected using the Advanced Photon Source Beamline 21-ID-F and a Rayonix MX300 detector. The data were processed using HKL2000 (38). The data collection statistics are in Table 1.

The structure of unliganded SK1<sub>BR</sub> was determined by molecular replacement using the Phaser (39) subroutine in Phenix (40) and the coordinates of the binding region of the unliganded adhesin from *S. mitis* strain unliganded NCTC10712 (PDB entry 6EFF) (18) as the search model. All solvent molecules were removed from the coordinates prior to searches. Molecular replacement required two separate steps. The first step used the unliganded Siglec domain (residues 244–370) as the search model and identified four copies of the Siglec domain. The coordinates for the Siglec domain were then fixed, and a second step using the Unique domain as the search model (residues 371–445) identified three copies of the Unique domain. The final Unique domain was manually placed, and the connections between domains were made during refinement.

Structures of the sialoglycan-bound SK1<sub>BR</sub> were determined by rigid body refinement of the individual domains of the unliganded SK1<sub>BR</sub> with all solvent molecules removed. Unambiguous electron density for each sialoglycan was visible in the ini-

tial maps. The sialoglycans were manually placed into the difference electron density in Coot (41) immediately following structure determination and prior to solvent placement. The ligands were then individually adjusted in Coot (41) prior to refinement in Phenix (40). To avoid model bias in analysis of the ligand electron density, the ligands were deleted, and the final solvated model was refined in Phenix to produce  $|F_o| - |F_c|$  and  $2|F_o| - |F_c|$  electron density maps. The  $|F_o| - |F_c|$  maps obtained from this protocol are shown in Fig. 5 at  $3\sigma$ . Ligand placements were validated with MotiveValidator (42).

The models were improved using real space refinement in Coot (41) and reciprocal space refinement in Phenix (40). For unliganded SK1<sub>BR</sub>, 5% of the reflections (totaling 3466 reflections) were randomly selected to use as the free-R and were held separately for the duration of the refinement. For the liganded structures, the equivalent reflections were selected for use as the free-R in the reflection editor subroutine in Phenix (40). The final models of liganded and sialoglycan-bound structures of SK1<sub>BR</sub> each contain two copies of SK1<sub>BR</sub> in each asymmetric unit, with a single copy containing all residues of the purified protein, *i.e.* residues 252–660 of the full-length adhesin. For the sialoglycan-bound structures, one trisaccharide is bound to each Siglec domain, such that a single copy of SK1<sub>BR</sub> binds two glycans and there are four trisaccharides per asymmetric unit. Refinement statistics and information regarding the content of the models can be found in Table 2.

### Structural analysis

SK1<sub>Siglec1</sub> (residues 252–377) and SK1<sub>Siglec2</sub> (residues 455–573) were aligned using PyMOL (43) which rejected 26 atoms from the alignment and calculated an RMS deviation for the remaining C $\alpha$  positions of 1.058 Å. SK1<sub>Unique1</sub> (residues 378–454) and SK1<sub>Unique2</sub> (residues 574–660) were aligned using the same method and had a calculated RMS deviation of 0.639 Å. The maximum distances between the loops of SK1<sub>Siglec1</sub> and SK1<sub>Siglec2</sub> were measured from the C $\alpha$  atoms of SK1<sup>N229</sup> and SK1<sup>A497</sup> for the CD loop, SK1<sup>G342</sup> and SK1<sup>G543</sup> for the EF loop, and SK1<sup>D365</sup> and SK1<sup>K565</sup> for the FG loop. The sequence logo (Fig. 1) was generated using WebLogo 3.7.4 (44) from a Clustal Omega (45) multiple sequence alignment. Ligand–protein interactions were analyzed by PDBsumgenerate (46) and LigPlot (47).

### Adhesins containing individual single unique binding module

Individual binding modules of SK1<sub>BR</sub>, *i.e.* SK1<sub>Siglec1+Unique1</sub> (residues 252–455) and SK1<sub>Siglec2+Unique2</sub> (residues 449–660) were designed based upon the manual evaluation of the end of each folded domain in the crystal structure (Fig. S5). SK1<sub>Siglec1+Unique1</sub> and SK1<sub>Siglec2+Unique2</sub> were each expressed from pGEX-3X containing an N-terminal GST tag using the protocol detailed above for the full-length SK1<sub>BR</sub>.

### Far Western blotting

Human blood samples were collected under protocol 11-06207, approved by the University of California San Francisco Institutional Review Board, and these studies and protocols abide by the Declaration of Helsinki principles. The binding of GST-tagged SK1<sub>BR</sub>, GST-tagged SK1<sub>Siglec1+Unique1</sub>, and GST-

tagged SK1<sup>Siglec2+Unique2</sup> to plasma, platelet, and salivary glycoproteins was performed as described (6, 15).

### Binding to immobilized platelets

Human blood samples were collected under protocol 11-06207, approved by the University of California San Francisco Institutional Review Board, and these studies and protocols abide by the Declaration of Helsinki principles. Binding to formalin-fixed human platelet monolayers was performed as described (6). In brief, human platelets were freshly prepared, washed, and immobilized in 96-well plates. After blocking non-specific binding with 1× blocking reagent (Roche) in Dulbecco's PBS (DPBS), the blocking solution was replaced with 50 μl of GST-tagged SK1<sub>BR</sub>, and split binding module proteins were diluted to the indicated concentrations into 1× blocking solution, and the plates were incubated for 1 h at room temperature (~22°C) with vigorous rocking. The wells were washed three times with 100 μl DPBS, and the bound GST-tagged proteins were detected by using a rabbit anti-GST (Life Technologies) diluted 1:500 in 1× blocking solution, followed by a peroxidase-conjugated goat anti-rabbit (Sigma) diluted 1:5000 in DPBS, along with the chromogenic substrate *o*-phenylenediamine dihydrochloride (Sigma).

### Binding of biotin-glycans to immobilized binding regions

The binding of polyvalent biotinylated glycans (Glycotect) to GST-tagged SK1<sub>BR</sub> and split binding module proteins immobilized in 96-well plates was performed as described (6). In brief, the wells were coated with the GST-tagged proteins (500 nM in DPBS) by incubating overnight at 4°C. The wells were washed twice with DPBS, and biotinylated glycans that had been diluted to the indicated concentrations in 1× blocking reagent (Roche) in DPBS were added. The plates were incubated for 90 min at room temperature (~22°C) with vigorous rocking, and the unbound glycans were removed by washing three times with DPBS. The bound glycans were detected by using peroxidase-conjugated streptavidin (Sigma), followed by the chromogenic substrate *o*-phenylenediamine dihydrochloride (Sigma).

### Data availability

The coordinates and structure factors have been deposited in the RCSB Protein Data Bank with the accession codes 6VS7 (unliganded), 6VT2 (sTa-bound), and 6VU6 (3'sLn-bound). Raw X-ray diffraction data have been deposited with SBGrid with accession codes 754 (sTa-bound), 755 (unliganded), and 756 (3'sLn-bound). The raw data for the binding analyses in Fig. 10 are available upon request from paul.sullam@ucsf.edu. All other data are contained within the article.

**Acknowledgments**—We thank reviewer 1 for bringing to our attention the structural similarities between the Siglec-like adhesins and SSL5.

**Author contributions**—H. E. S., B. A. B., and P. S. data curation; H. E. S., I. Y., and P. S. software; H. E. S., B. A. B., I. Y., P. S., P. M. S., and T. M. I. formal analysis; H. E. S., I. Y., P. S., and T. M. I. valida-

tion; H. E. S., B. A. B., P. S., H. Y., P. M. S., and T. M. I. investigation; H. E. S., B. A. B., and P. S. visualization; H. E. S. and T. M. I. writing—original draft; H. E. S., B. A. B., X. C., P. M. S., and T. M. I. writing—review and editing; B. A. B., X. C., P. M. S., and T. M. I. conceptualization; B. A. B., I. Y., P. S., H. Y., P. M. S., and T. M. I. methodology; I. Y., P. S., X. C., and T. M. I. supervision; X. C., P. M. S., and T. M. I. funding acquisition; P. M. S. and T. M. I. resources; P. M. S. and T. M. I. project administration.

**Funding and additional information**—This work was funded by National Institutes of Health Grants AI41513 and U01CA221244 (to P. M. S.) and AI106987 (to T. M. I. and P. M. S.) and by American Heart Association Grants 19POST34450093 (to P. S.) and 17SDG33660424 (to B. A. B.). H. E. S. was supported by National Institutes of Health Training Grants 5T32EY007135-25 and 2T32GM008320-30. The Advanced Photon Source, a user facility operated for the U.S. DOE Office of Science, was supported under Contract DE-AC0206CH11357. Use of LS-CAT sector 21 was supported by Grant 085P1000817 from the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Conflict of interest**—The authors declare that they have no conflicts of interest with the contents of this article.

**Abbreviations**—The abbreviations used are: BR, binding region; GST, glutathione *S*-transferase; Siglec, sialic acid-binding immunoglobulin-like lectin; SRR, serine-rich repeat; sTa, sialyl T-antigen; 3'sLn, 3'-sialyl *N*-acetylglucosamine; RMS, root mean square; DPBS, Dulbecco's PBS.

### References

- Ramboarina, S., Garnett, J. A., Zhou, M., Li, Y., Peng, Z., Taylor, J. D., Lee, W. C., Bodey, A., Murray, J. W., Alguet, Y., Bergeron, J., Bardiaux, B., Sawyer, E., Isaacson, R., Tagliaferri, C., *et al.* (2010) Structural insights into serine-rich fimbriae from Gram-positive bacteria. *J. Biol. Chem.* **285**, 32446–32457 [CrossRef Medline](#)
- Deng, L., Bensing, B. A., Thamadolok, S., Yu, H., Lau, K., Chen, X., Ruhl, S., Sullam, P. M., and 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 [CrossRef Medline](#)
- Zheng, W., Tan, M. F., Old, L. A., Paterson, I. C., Jakubovics, N. S., and Choo, S. W. (2017) Distinct biological potential of *Streptococcus gordonii* and *Streptococcus sanguinis* revealed by comparative genome analysis. *Sci. Rep.* **7**, 2949 [CrossRef Medline](#)
- Ronis, A., Brockman, K., Singh, A. K., Gaytán, M. O., Wong, A., McGrath, S., Owen, C. D., Magrini, V., Wilson, R. K., van der Linden, M., and King, S. J. (2019) *Streptococcus oralis* subsp. *dentisani* produces monolateral serine-rich repeat protein fibrils, one of which contributes to saliva binding via sialic acid. *Infect. Immun.* **87**, e00406-19 [CrossRef Medline](#)
- Denapaite, D., Brückner, R., Nuhn, M., Reichmann, P., Henrich, B., Maurer, P., Schähle, Y., Selbmann, P., Zimmermann, W., Wambutt, R., and Hakenbeck, R. (2010) The genome of *Streptococcus mitis* B6: what is a commensal? *PLoS One* **5**, e9426 [CrossRef Medline](#)
- Bensing, B. A., Khedri, Z., Deng, L., Yu, H., Prakobphol, A., Fisher, S. J., Chen, X., Iverson, T. M., Varki, A., and Sullam, P. M. (2016) Novel aspects of sialoglycan recognition by the Siglec-like domains of streptococcal SRR glycoproteins. *Glycobiology* **26**, 1222–1234 [CrossRef Medline](#)
- Xiong, Y. Q., Bensing, B. A., Bayer, A. S., Chambers, H. F., and Sullam, P. M. (2008) Role of the serine-rich surface glycoprotein GspB of

## Structure of *S. sanguinis* SK1 adhesin

- Streptococcus gordonii* in the pathogenesis of infective endocarditis. *Microb. Pathog.* **45**, 297–301 [CrossRef Medline](#)
8. Seo, H. S., Xiong, Y. Q., and Sullam, P. M. (2013) Role of the serine-rich surface glycoprotein Srr1 of *Streptococcus agalactiae* in the pathogenesis of infective endocarditis. *PLoS One* **8**, e64204 [CrossRef Medline](#)
  9. van Sorge, N. M., Quach, D., Gurney, M. A., Sullam, P. M., Nizet, V., and Doran, K. S. (2009) The group B streptococcal serine-rich repeat 1 glycoprotein mediates penetration of the blood-brain barrier. *J. Infect. Dis.* **199**, 1479–1487 [CrossRef Medline](#)
  10. Takahashi, Y., Takahashi, E., Shimazu, K., Yagishita, H., Aoba, T., and Konishi, K. (2006) Contribution of sialic acid-binding adhesin to pathogenesis of experimental endocarditis caused by *Streptococcus gordonii* DL1. *Infect. Immun.* **74**, 740–743 [CrossRef Medline](#)
  11. Shivshankar, P., Sanchez, C., Rose, L. F., and Orihuela, C. J. (2009) The *Streptococcus pneumoniae* adhesin PsrP binds to Keratin 10 on lung cells. *Mol. Microbiol.* **73**, 663–679 [CrossRef Medline](#)
  12. Mistou, M. Y., Dramsi, S., Brega, S., Poyart, C., and Trieu-Cuot, P. (2009) Molecular dissection of the *secA2* locus of group B *Streptococcus* reveals that glycosylation of the Srr1 LPXTG protein is required for full virulence. *J. Bacteriol.* **191**, 4195–4206 [CrossRef Medline](#)
  13. Singh, A. K., Woodiga, S. A., Grau, M. A., and King, S. J. (2017) *Streptococcus oralis* neuraminidase modulates adherence to multiple carbohydrates on platelets. *Infect. Immun.* **85**, e00774-16 [CrossRef Medline](#)
  14. Takamatsu, D., Bensing, B. A., Cheng, H., Jarvis, G. A., Siboo, I. R., López, J. A., Griffiss, J. M., and Sullam, P. M. (2005) Binding of the *Streptococcus gordonii* surface glycoproteins GspB and Hsa to specific carbohydrate structures on platelet membrane glycoprotein Ib $\alpha$ . *Mol. Microbiol.* **58**, 380–392 [CrossRef Medline](#)
  15. Bensing, B. A., Li, Q., Park, D., Lebrilla, C. B., and Sullam, P. M. (2018) Streptococcal Siglec-like adhesins recognize different subsets of human plasma glycoproteins: implications for infective endocarditis. *Glycobiology* **28**, 601–611 [CrossRef Medline](#)
  16. Pyburn, T. M., Bensing, B. A., Xiong, Y. Q., Melancon, B. J., Tomasiak, T. M., Ward, N. J., Yankovskaya, V., Oliver, K. M., Cecchini, G., Sulikowski, G. A., Tyska, M. J., Sullam, P. M., and Iverson, T. M. (2011) A structural model for binding of the serine-rich repeat adhesin GspB to host carbohydrate receptors. *PLoS Pathog.* **7**, e1002112 [CrossRef Medline](#)
  17. Loukachevitch, L. V., Bensing, B. A., Yu, H., Zeng, J., Chen, X., Sullam, P. M., and Iverson, T. M. (2016) Structures of the *Streptococcus sanguinis* SrpA binding region with human sialoglycans suggest features of the physiological ligand. *Biochemistry* **55**, 5927–5937 [CrossRef Medline](#)
  18. Bensing, B. A., Loukachevitch, L. V., Agarwal, R., Yamakawa, I., Luong, K., Hadadianpour, A., Yu, H., Fialkowski, K. P., Castro, M. A., Wawrzak, Z., Chen, X., Baudry, J., Smith, J. C., Sullam, P. M., and Iverson, T. M. (2019) Selectivity and engineering of the sialoglycan-binding spectrum in Siglec-like adhesins. *bioRxiv* [CrossRef](#)
  19. Bensing, B. A., Loukachevitch, L. V., McCulloch, K. M., Yu, H., Vann, K. R., Wawrzak, Z., Anderson, S., Chen, X., Sullam, P. M., and Iverson, T. M. (2016) Structural basis for sialoglycan binding by the *Streptococcus sanguinis* SrpA adhesin. *J. Biol. Chem.* **291**, 7230–7240 [CrossRef Medline](#)
  20. Urano-Tashiro, Y., Takahashi, Y., Oguchi, R., and Konishi, K. (2016) Two arginine residues of *Streptococcus gordonii* sialic acid-binding adhesin Hsa are essential for interaction to host cell receptors. *PLoS One* **11**, e0154098 [CrossRef Medline](#)
  21. Bensing, B. A., Li, L., Yakovenko, O., Wong, M., Barnard, K. N., Iverson, T. M., Lebrilla, C. B., Parrish, C. R., Thomas, W. E., Xiong, Y., and Sullam, P. M. (2019) Recognition of specific sialoglycan structures by oral streptococci impacts the severity of endocardial infection. *PLoS Pathog.* **15**, e1007896 [CrossRef Medline](#)
  22. Baker, H. M., Basu, I., Chung, M. C., Caradoc-Davies, T., Fraser, J. D., and Baker, E. N. (2007) Crystal structures of the staphylococcal toxin SSL5 in complex with sialyl Lewis X reveal a conserved binding site that shares common features with viral and bacterial sialic acid binding proteins. *J. Mol. Biol.* **374**, 1298–1308 [CrossRef Medline](#)
  23. Cohen, M., and Varki, A. (2014) Modulation of glycan recognition by clustered saccharide patches. *Int. Rev. Cell Mol. Biol.* **308**, 75–125 [CrossRef Medline](#)
  24. Narimatsu, Y., Joshi, H. J., Nason, R., Van Coillie, J., Karlsson, R., Sun, L., Ye, Z., Chen, Y. H., Schjoldager, K. T., Steentoft, C., Furukawa, S., Bensing, B. A., Sullam, P. M., Thompson, A. J., Paulson, J. C., et al. (2019) An atlas of human glycosylation pathways enables display of the human glycome by gene engineered cells. *Mol. Cell* **75**, 394–407 [CrossRef Medline](#)
  25. Nagashima, H., Tezuka, T., Tsuchida, W., Maeda, H., Kohroki, J., and Masuho, Y. (2008) Tandemly repeated Fc domain augments binding avidities of antibodies for Fc $\gamma$  receptors, resulting in enhanced antibody-dependent cellular cytotoxicity. *Mol. Immunol.* **45**, 2752–2763 [CrossRef Medline](#)
  26. Moonens, K., Van den Broeck, I., De Kerpel, M., Deboeck, F., Raymaekers, H., Remaut, H., and De Greve, H. (2015) Structural and functional insight into the carbohydrate receptor binding of F4 fimbriae-producing enterotoxigenic *Escherichia coli*. *J. Biol. Chem.* **290**, 8409–8419 [CrossRef Medline](#)
  27. Apic, G., Gough, J., and Teichmann, S. A. (2001) Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol.* **310**, 311–325 [CrossRef Medline](#)
  28. Moonens, K., Gideonsson, P., Subedi, S., Bugaytsova, J., Roma $\delta$ , E., Mendez, M., Nordén, J., Fallah, M., Rakhimova, L., Shevtsova, A., Lahmann, M., Castaldo, G., Brännström, K., Coppens, F., Lo, A. W., et al. (2016) Structural insights into polymorphic ABO glycan binding by *Helicobacter pylori*. *Cell Host Microbe* **19**, 55–66 [CrossRef Medline](#)
  29. Chen, S. L., Hung, C. S., Pinkner, J. S., Walker, J. N., Cusumano, C. K., Li, Z., Bouckaert, J., Gordon, J. I., and Hultgren, S. J. (2009) Positive selection identifies an *in vivo* role for FimH during urinary tract infection in addition to mannose binding. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22439–22444 [CrossRef Medline](#)
  30. Aspholm-Hurtig, M., Dailide, G., Lahmann, M., Kalia, A., Ilver, D., Roche, N., Vikström, S., Sjöström, R., Lindén, S., Bäckström, A., Lundberg, C., Arnqvist, A., Mahdavi, J., Nilsson, U. J., Velapatiño, B., et al. (2004) Functional adaptation of BabA, the *H. pylori* ABO blood group antigen binding adhesin. *Science* **305**, 519–522 [CrossRef Medline](#)
  31. Cszimar, C. M., Petersburg, J. R., Perry, T. J., Rozumalski, L., Hackel, B. J., and Wagner, C. R. (2019) Multivalent ligand binding to cell membrane antigens: defining the interplay of affinity, valency, and expression density. *J. Am. Chem. Soc.* **141**, 251–261 [CrossRef Medline](#)
  32. Moonens, K., and Remaut, H. (2017) Evolution and structural dynamics of bacterial glycan binding adhesins. *Curr. Opin. Struct. Biol.* **44**, 48–58 [CrossRef Medline](#)
  33. Worstell, N. C., Singla, A., Saenkham, P., Galbadage, T., Sule, P., Lee, D., Mohr, A., Kwon, J. S., Cirillo, J. D., and Wu, H. J. (2018) Hetero-multivalency of *Pseudomonas aeruginosa* lectin LecA binding to model membranes. *Sci. Rep.* **8**, 8419 [CrossRef Medline](#)
  34. Yu, H., Yan, X., Autran, C. A., Li, Y., Etzold, S., Latasiewicz, J., Robertson, B. M., Li, J., Bode, L., and Chen, X. (2017) Enzymatic and chemoenzymatic syntheses of disialyl glycans and their necrotizing enterocolitis preventing effects. *J. Org. Chem.* **82**, 13152–13160 [CrossRef Medline](#)
  35. Yu, H., Yu, H., Karpel, R., and Chen, X. (2004) Chemoenzymatic synthesis of CMP-sialic acid derivatives by a one-pot two-enzyme system: comparison of substrate flexibility of three microbial CMP-sialic acid synthetases. *Bioorg. Med. Chem.* **12**, 6427–6435 [CrossRef Medline](#)
  36. Sugiarto, G., Lau, K., Qu, J., Li, Y., Lim, S., Mu, S., Ames, J. B., Fisher, A. J., and Chen, X. (2012) A sialyltransferase mutant with decreased donor hydrolysis and reduced sialidase activities for directly sialylating Lewis<sup>X</sup>. *ACS Chem. Biol.* **7**, 1232–1240 [CrossRef Medline](#)
  37. Lau, K., Yu, H., Thon, V., Khedri, Z., Leon, M. E., Tran, B. K., and Chen, X. (2011) Sequential two-step multienzyme synthesis of tumor-associated sialyl T-antigens and derivatives. *Org. Biomol. Chem.* **9**, 2784–2789 [CrossRef Medline](#)
  38. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 [Medline](#)
  39. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 [CrossRef Medline](#)
  40. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., et al. (2019) Macromolecular structure determination using X-rays, neutrons and electrons:

- recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 [CrossRef Medline](#)
41. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 [CrossRef Medline](#)
42. Varekova, R. S., Jaiswal, D., Sehnal, D., Ionescu, C. M., Geidl, S., Pravda, L., Horsky, V., Wimmerova, M., and Koca, J. (2014) MotiveValidator: interactive web-based validation of ligand and residue structure in biomolecular complexes. *Nucleic Acids Res.* **42**, W227–W233 [CrossRef Medline](#)
43. DeLano, W. L. (2012) *The PyMOL Molecular Graphics System*, Schrödinger, LLC, New York
44. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 [CrossRef Medline](#)
45. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 [CrossRef Medline](#)
46. Laskowski, R. A. (2009) PDBsum new things. *Nucleic Acids Res.* **37**, D355–D359 [CrossRef Medline](#)
47. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**, 127–134 [CrossRef Medline](#)
48. Stein, P. E., Boodhoo, A., Armstrong, G. D., Heerze, L. D., Cockle, S. A., Klein, M. H., and Read, R. J. (1994) Structure of a pertussis toxin-sugar complex as a model for receptor binding. *Nat. Struct. Biol.* **1**, 591–596 [CrossRef Medline](#)